UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ISOLATION AND FUNCTIONAL GENETIC ANALYSIS OF *EUCALYPTUS* WOOD FORMATION GENES

by

HONGHAI ZHOU

Submitted in partial fulfillment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences
Department of Genetics
University of Pretoria
Pretoria
August 2005

Under the supervision of Dr Alexander A. Myburg
and Prof. David K. Berger (co-supervisor)

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.

Honghai Zhou

August 2005

PREFACE

Eucalyptus is an important genus of forest tree species which are widely used as a resource for wood and fibre for the pulp and papermaking industry. However, the yield of wood and fibre of *Eucalyptus* and other woody species is not high enough to meet the world-wide demand. The pulp and papermaking industry is also a large source of pollution mainly due to the use of many toxic chemicals when wood fibre is processed. The amount of chemicals required for pulping can be reduced by altering wood properties such as lignin content and composition. Paper quality can also be improved by changing wood properties such as microfibril angles. However, our understanding of the wood formation process is far from complete. This is an obstacle for us to genetically engineer wood properties of *Eucalyptus* and other tree species. The functions of individual wood formation genes need to be studied first to understand this complex biological process. However, it is quite difficult to obtain transgenic *Eucalyptus* trees. Transformation efficiency is very low in *Eucalyptus* trees and requires a substantial investment in tissue culture facilities. Even if successful transformation and regeneration is achieved, it can take another 24 months to obtain a reliable transgene phenotype. Model systems, such as *Arabidopsis*, may be useful for functional analysis of wood formation genes identified in *Eucalyptus*. In this project we evaluated the use of *Arabidopsis* to model the functions of *Eucalyptus* wood formation genes, specifically these involved in lignin and cellulose biosynthesis in xylem cell walls.

This thesis consists of a literature review (Chapter One), a report of the establishment of the *Arabidopsis thaliana* model system (Chapter Two and Chapter Three) and the isolation of two novel *Eucalyptus* wood formation genes (Chapter Four). Chapter Two is

prepared partly as complementary materials and methods for Chapter Three. Chapter Three and Chapter Four were independently prepared for publication. Therefore, some sections of the thesis are redundant.

In Chapter One, an overview of wood formation mechanisms is provided. First, we discuss the fundamental process of wood formation in *Eucalyptus* and other woody species. Second, molecular biology methods that are used and can be used to study wood formation in *Eucalyptus* are introduced. Third, model systems for wood formation research are discussed. Some challenges to functional genetics analysis of wood formation genes are described at the end of the chapter.

Chapter Two and Chapter Three describe the establishment of *Arabidopsis thaliana* as a model system to analyze the function of wood formation genes previously isolated in *Eucalyptus*. To test the utilization of this model system we transformed a *Eucalyptus* lignin biosynthesis gene, *cinnamoyl CoA reductase* (*EUCCR*), into wild-type *Arabidopsis* (Col-0) (for overexpression) and into an *Arabidopsis* mutant line where the ortholog of *EUCCR* is mutated (for complementation). Six *Eucalyptus cellulose synthase* (*EgCesA*) genes have recently been isolated in our laboratory. One of these, *EgCesA1*, was transformed into an *Arabidopsis* mutant line where the ortholog of *EgCesA1* is mutated (for complementation). However, the complementation experiment for *EUCCR* failed. Three positive T1 lines were obtained in the complementation test for *EgCesA1*. The analysis of these lines is still in progress. Chapter Three describes the overexpression test for *EUCCR*. This chapter is prepared for publication in *Journal of Plant Physiology*. The detailed Materials and Methods for Chapter Three are discussed in Chapter Two. The complementation experiment and other results, such as leaf PCR and GUS assays, from the overexpression experiment are also described in Chapter Two.

Cellulose biosynthesis requires the delivery of activated glucose molecules to the cellulose synthase complex in the cell membrane. The proteins responsible for delivery if these activated sugar molecules, sucrose synthases, are important targets for the modification of cellulose biosynthesis in forest trees. The isolation of two novel *Eucalyptus* sucrose synthase genes is discussed in Chapter Four. This chapter discusses the reasons why *Eucalyptus* sucrose synthases are important for wood formation in *Eucalyptus*, the methods we used to isolate and characterize two novel *Eucalyptus* sucrose synthase genes, and the results to prove the functions of these genes.

The following are accepted abstracts from the results of this project:

1 . **Honghai Zhou, David K. Berger, Alexander A. Myburg 2004.** Investigation of *Arabidopsis thaliana* as a model plant for functional analysis of wood formation genes of *Eucalyptus* trees. South African Genetics Society 2004 Conference, Stellenbosh, South Africa, 4 – 7 April 2004. (Poster presentation)

2 . **Honghai Zhou, David K. Berger, Shawn D. Mansfield, Alexander A. Myburg 2005.** Overexpression of *Eucalyptus* cinnamoyl-CoA reductase in *Arabidopsis* alters lignin content and composition of *Arabidopsis* stems. IUFRO Tree Biotechnology 2005, Pretoria, South Africa, 7-11 November 2005. (Poster presentation) (Accepted)

3 . **Honghai Zhou, Alexander A. Myburg 2005.** Cloning, characterization and expression profiling of two *Eucalyptus* sucrose synthase genes. IUFRO Tree Biotechnology 2005, Pretoria, South Africa, 7-11 November 2005. (Poster presentation) (Accepted)

In addition, the following manuscripts are currently in preparation for publication:

1. Honghai Zhou, David K. Berger, Shawn D. Mansfield, Alexander A. Myburg

Overexpression of *Eucalyptus* cinnamoyl-CoA reductase in *Arabidopsis* alters lignin content and composition of *Arabidopsis* stems. Prepared for *Journal of Plant Physiology*

2. **Honghai Zhou, Frank Maleka, Alexander A. Myburg** Cloning, characterization and expression profiling of two novel *Eucalyptus* sucrose synthase genes. Prepared for *Functional Plant Biology*

ACKNOWLEDGEMENTS

I would like to give my acknowledgements to the following people, organizations and institutes for supporting me to complete this project:

- Dr A.A. Myburg for his excellent supervision, motivation and creative leadership of this project.
- Prof. Dave Berger for his excellent co-supervision.
- My colleagues in the (FMG) Forest Molecular Genetics Laboratory: Adrene Laubscher, Anita Steyn, Elna Cowley, John Kemp, Luke Solomon, Martin Ranik, Michelle Victor, Minique De Castro, Mmoledi Mphahlele Dr Solomon Fekybelu, and Dr Yoseph Beyene for their cooperation and valuable recommendations.
- Prof. Shawn D. Mansfield for his cooperation to perform the chemical analysis of the transgenic plants.
- Frank Maleka for his cooperation to sequence the genomic copies of two *Eucalyptus* sucrose synthase genes.
- Mondi Business Paper South Africa and Sappi Forest Products for funding contributions to this project.
- The Human Resources and Technology for Industry Programme (THRIP) and National Research Foundation (NRF) for financial support of the research.
- My parents for financial support and understanding.
- My girlfriend for support and understanding.
- Forestry and Agricultural Biotechnology Institute (FABI) and Department of Genetics in the University of Pretoria for providing facilities.
- All my Chinese friends in South Africa for their help.

TABLE OF CONTENTS

Chapter One

Literature Review

Functional genetics of wood formation in *Eucalyptus*:

From genes to wood

1.1 Introduction

Trees are indispensable for human existence. Forests are valuable for wildlife habitats, scenic vistas and recreational areas. They are a resource of wood and fibre, which is a raw material used for pulp, paper, furniture, construction, and a crucial source of renewable energy. It is well known that actively growing wood cells are important sinks for excess atmospheric CO_2. They therefore reduce one of the major sources of global warming. Wood and related products constitute a very important commodity in international trade markets. However, the growing human population of the world is making it increasingly difficult to meet the demand for wood products.

The best way to meet future demand for wood is to grow wood as a crop in the same way that we grow wheat and rice (Fenning and Gershenzon, 2002). To domesticate wild tree species and develop high-yielding, intensively managed plantation forests, the best available modern scientific methods need to be applied. A better understanding of complex wood formation mechanisms, in particular, the functions of wood formation genes, and the application of biotechnology (marker-assisted breeding, genetic engineering and tissue culture) to forests should help us develop superior trees in terms of the quality and quantity of the wood (Fenning and Gershenzon, 2002).

Eucalyptus is a large and fascinating woody plant genus that contains more than 700 species (Brooker and Slee, 1996). *Eucalyptus* trees can be found in a variety of sizes with some species being among the largest trees in the world. Interestingly, some members of this genus are mere shrubs. *Eucalyptus* tree species can be planted widely, because they are able to find nourishment in a complexity of soils and can survive in a variety of climates. For example, they grow in deserts, swamplands, valleys and alpine regions (Tibbits *et al.*, 1997). *Eucalyptus* pulp has been used as a popular source for

paper and fibreboard, especially for the paper industry that switched from long to short fibre (Santos, 1997). Generally, *Eucalyptus* trees used for pulping are grown from five to seven years, which is a shorter growth period than many other trees. The most popular *Eucalyptus* species used in the pulp and papermaking industry is *E. globulus* because of its wood properties which are beneficial for pulp production (Sanchez, 2002; Grattapaglia, 2004).

Three sections are included in this review. The first section discusses the fundamental process of wood formation in *Eucalyptus* and other woody species. This process includes the activities of the vascular cambium and the properties of the plant cell wall, which mainly consists of cellulose, lignin, hemi-cellulose, pectin and cell wall proteins. The latest research on wood formation is then discussed. In the section on functional genetics of wood formation, molecular biology methods and model systems used in our and other labs for functional analysis of wood formation genes are discussed. The last section of the chapter discusses the challenges and opportunities of wood formation studies in the genomic and post-genomic eras.

1.2 The wood formation process

Considering the importance of wood, it is interesting that our understanding of the cellular, molecular, and developmental processes that underlie wood formation (xylogenesis) is far from complete. Xylogenesis, or xylem development, is an example of cell differentiation in a complex form. This process is regulated by many factors, both exogenous (photoperiod and temperature) and endogenous (phytohormones) and the interaction between them (Fukuda, 1996). The involvement of gene families and the plasticity of the metabolism involved add additional complexity to the process of wood formation (Baucher *et al.*, 1998).

Mature wood cells are formed in a five-step process including cell division from cambial initials, expansion (elongation and radial enlargement), wall thickening (involving the deposition of lignin, cellulose, hemicellulose, pectins, and cell wall proteins), programmed cell death (PCD), and the formation of heartwood (HW) (Fukuda, 1996). Therefore, the genes and proteins which control this process need to be studied in order to allow scientists to design strategies for the modification of wood properties. The regulation of wood formation genes in response to developmental and environmental stimuli is one of the possible causes of variation in wood properties (Darley *et al.*, 2001).

1.2.1 The vascular cambium, xylem and phloem

Wood is a product of the vascular cambium, a secondary meristem derived from the procambium, which, in turn, develops from the apical meristem (Larson, 1994). The vascular cambium is comparable to the apical meristem in concept, but unlike the apical meristem, the cambium contains two morphologically distinct cell types: fusiform cambial initials and ray cambial initials, which give rise to the axial and horizontal cells, respectively, in the secondary xylem and phloem (Mellerowicz *et al.*, 2001). The central cells of the cambium have stem cell-like properties and are able to replace themselves with each division, which gives rise to either xylem or phloem mother cells (Larson, 1994). The vascular cambium plays a major role in the lateral growth of gymnosperms and woody angiosperms (Plomion *et al.*, 2001). Genes specifically expressed in the vascular cambium could be targets for the genetic engineering of wood properties. Also, hormones and other signals are thought to play important roles in the regulation of the cambial activity. For example, auxin transport from leaves to roots provides the major signal for cambial development (Sundberg *et al.*, 2000). The molecular and

physiological mechanisms of vascular development are now considered an important area of research (Plomion *et al*. 2001).

Xylem and phloem are derived from xylem and phloem mother cells, respectively, which in turn as stated before are products of cambial initials. Secondary xylem in trees mainly contains tracheary elements, parenchyma cells, and sclerenchyma fibres, which differentiate in predictable ratios, although the placement within the xylem is not precise (Esau, 1965). The xylem cells play crucial roles in water transport and mechanical support of the entire plant. Secondary phloem is comprised of sieve tube elements, fibres, parenchyma cells, and companion cells. Phloem is essential to the transport of photosynthate from photosynthetic tissues (source tissues) to developing tissues (sink tissues). The volume of xylem is normally bigger than that of phloem because xylem mother cells are produced more quickly than phloem mother cells (Plomion *et al*. 2001). The ratio of xylem to phloem development may be under hormonal control. Aloni (2001) reported that low concentrations of auxin induce phloem differentiation, while high levels promote xylem differentiation. Israelsson *et al.* (2003) reported that elevated levels of gibberellic acid (GA) were positively related to numbers of xylem fibres and xylem fibre length.

1.2.2 The plant cell wall

Plants have two types of cell walls, namely primary walls and secondary walls. All plant cells have primary walls, which are deposited at the beginning of cell differentiation and consist mainly of lignin and polysaccharides that can be classified as randomly arranged cellulose, hemicelluloses, and pectic substances. Hemicelluloses and pectin are often referred to as matrix polysaccharides, which are synthesized within Golgi cisternae and then transported to the wall by exocytosis. In some plant cells, secondary cell walls are

deposited inside the primary wall after the cell reaches its definitive size. The secondary wall can be divided into three different layers, S_1, S_2 and S_3 (Timell, 1986). Secondary walls represent composites of cellulose, hemicelluloses and lignin. The thickness of microfibrils and microfibril angles differ from S layer to S layer. The S2 layer is the thickest and most important in terms of mechanical support (Mellerowicz *et al*., 2001). It also determines the overall microfibril angle of wood fibres, which is an important determination of fibre strength. The microfibril angle in the S1 layer is almost transverse; the angle gradually changes clockwise from S1 to S2, then to S3 layer; finally, the microfibril angle in S3 layer is back to transverse (Mellerowicz *et al*., 2001).

1.2.2.1 Cellulose

Cellulose is an important component of plant cell walls; therefore it is important to plants and human beings. Plant cellulose is the most abundant bio-macromolecule in nature. It is not only an economically important commodity, but also the primary determinant of fibre strength. Cellulose accounts for 28% to 30% of dry weight in forage grasses (Theander and Westerlund, 1993) and 42% to 45% of wood (Smook, 1992). The mechanism of cellulose biosynthesis is not well understood, due to the large protein complexes involved, and the unknown identity of other molecules associated with this process.

Plant cellulose is synthesized at rosette-like structures which consist of up to 36 hexagonally arranged protein subunits (cellulose synthases) and which are thought to be embedded in the plasma membrane (Williamson *et al.*, 2001). Due to the glycosyl transferase activity of cellulose synthase, this protein catalyzes the formation of β-1, 4 bonds between the substrate molecules, UDP-glucose (Saxena *et al.*, 1995). The resulting cellulose molecules are therefore unbranched 1, 4-β-D-glucans. After it is

transported into the plant cell wall through the cell membrane, cellulose is assembled into microfibrils as the result of an association of inter- and intra- hydrogen bonds between the different glucan chains (Plomion *et al.*, 2001). Therefore, individual microfibrils are composed of up to 36 glucan chains. There are at least ten cellulose synthase genes (*CesA*) in the genome of the model plant, *Arabidopsis* (Richmond, 2000; Delmer *et al.*, 2001), and loss-of-function analyses have revealed the functions of at least six *CesA* genes (Richmond, 2000). Three of these were identified to be involved in primary cell wall development (*AtCesA1*, *AtCesA3* and *AtCesA6*), and the other three were shown to be associated with secondary cell wall development (*AtCesA4*, *AtCesA7* and *AtCesA8*) in *Arabidopsis* (Doblin *et al*., 2002). *CesA* genes have also been isolated from other plant species, such as cotton (Pear *et al.*, 1996), poplar (Joshi *et al.*, 2004) and barley (Burton *et al.*, 2004).

Other molecules and proteins are also involved in this complex biosynthetic process (Figure 1.1). A plasma membrane-bound endo-1,4-β-D-glucanase, KORRIGAN (KOR), which is important for normal wall assembly and cell elongation, may be required for cellulose synthesis (Nicol *et al*., 1998). Cellulose biosynthesis is probably initiated with the conjugation of the cellulose monomer, UDP-glucose, to a sterol sitosterol primer, which is the major plant sitosterol, forming sitosterol-β-glucoside (Peng *et al*., 2002). KOR may act to remove glucan units from these primers, which are subsequently incorporated into the growing glucan chain by cellulose synthases (Figure 1.1, Peng *et al*., 2002).

The protein sucrose synthase (SuSy; EC 2.4.1.13), a homotetrameric enzyme, is thought to play an important role in cellulose biosynthesis through the supply of UDP-glucose as shown in Figure 1.1 (Kleczkowski, 1994). Sucrose synthase catalyzes the reversible

conversion of sucrose and UDP to UDP-glucose (UDP-Glc) and fructose, but it functions to degrade sucrose under most conditions, particularly in nonphotosynthetic tissues (Kleczkowski, 1994). It is believed that oxygen and sugar availability strongly affect the transcription of sucrose synthase gene (Koch, 2004). In most of studies it was found that there are at least two non-allelic isoforms of SuSy that perform different physiological functions and SuSy proteins have been divided into four groups (A new group, a monocot group, a dicot SUS1 group, and a dicot SUSA group, Komatsu *et al.*, 2002; Baud *et al.*, 2004). Cellulose content has been increased 150% to 200% in endodermal cells of wheat roots under hypoxia, which up-regulates SuSy activity (Albrecht, 2000), probably due to the direct supply of substrate for cellulose synthesis.

1.2.2.2 Lignin

Lignin is a major structural component of the secondary plant cell wall. It is the second most abundant biopolymer in nature after cellulose and it is of fundamental importance for the development of higher plants and for resistance to biotic and abiotic stresses. It is also important to industry and agriculture (Campbell and Sederoff, 1996). Lignification emerged about 430 million years ago and the deposition of lignin in plant cell walls was part of the mechanism that allowed the development of upright plants of large size adapted to a terrestrial habitat (Boudet, 2000). The hydrophobic nature of lignin provides the ability to the vascular system to transport water and nutrient solutions. Lignin needs to be removed from cellulose during the pulping process by the use of hazardous and expensive chemicals that are serious sources of pollution (Boerjan *et al.*, 2003). In addition, lignin content is negatively associated with the quality and digestibility of fodder crops such as alfalfa, which in turn impacts the livestock industry (Moore and Jung, 2001).

Lignin is a complex heterogeneous phenolic polymer, which is derived mainly from the oxidative polymerization of three hydroxycinnamyl alcohol monomers (monolignols). The three monolignols, *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, only differ in their degree of methoxylation. Angiosperm lignins mainly consist of a mixture of G, S and traces of H units, derived from coniferyl, sinapyl and *p*-coumaryl alcohols, respectively. In gymnosperms, coniferyl alcohol and *p*-coumaryl alcohol are utilized to form lignin (Anterola and Lewis, 2002). Lignin composition varies between taxa and cell types, and is influenced by developmental processes and environmental stress conditions (Raes *et al.*, 2003). Lignin composition in the cell wall is an important economic trait. High S-unit levels facilitate the delignification process employed by the pulp and paper industry because of the high methoxylation state of this unit (Baucher *et al.*, 1998).

Many efforts have been made to understand monolignol biosynthesis, transport and polymerization. Numerous genes involved in monolignol biosynthesis (Figure 1.2) have been identified over the past decade. Phenylalanine ammonia-lyase (PAL), coumaroyl-CoA-3-hydroxylase (C3H), 4-coumarate-CoA ligase (4CL) and caffeoyl CoA *O*-methyltransferase (CCoAOMT) are part of the phenylpropanoid pathway (Figure 1.2), where phenylalanine is converted to cinnamoyl coenzyme A (CoA) (Boerjan *et al.*, 2003). Cinnamoyl CoA Reductase (CCR), coniferyl aldehyde-5-hydroxylase (Cald5H), caffeic acid *O*-methyltransferase (COMT) and coniferyl alcohol dehydrogenase (CAD) are the main enzymes utilized for the specific step dedicated to the monolignol biosynthetic branch (Figure 1.2, Anterola and Lewis, 2002). At the biochemistry and genetic levels, the entire pathway of lignin biosynthesis has not been characterized in a single tissue or a single species. The currently accepted model of monolignol synthesis envisages a metabolic grid leading to G and S units.

Dixon *et al.* (2001) suggested an alternative model in which metabolic channelling allows for independent pathways to G and S lignin, which is supported by the results of genetic manipulation experiments and enzyme specificity studies which are inconsistent with the classical model of monolignol biosynthesis. The precise mechanisms controlling the transport and polymerization of monolignols are still unknown, although it is thought that the polymerization procedure is probably catalyzed by peroxidases and laccases (Boudet *et al.*, 1995; Boerjan *et al.*, 2003).

1.2.2.3 Hemi-cellulose, pectin and cell wall proteins

The cellulose microfibrils are associated with mixtures of matrix polysaccharides (referred to herein as pectins and hemicelluloses), accounting for about 25% of the dry weight of wood. Unlike cellulose, hemi-celluloses occur as heteropolymers like glucomannan, galactoglucomannan, arabinogalactan, and glucuronoxylan, or as homopolymers such as galactan, arabinan, and β-1, 3-glucan (Plomion *et al*. 2001). In the primary walls of angiosperms, xyloglucan is the major component of hemicellulose, and it mainly consists of a 1, 4-β-D-glucan backbone carrying 1, 6-α-D-xylose moieties on three consecutive glucose residues. The second and third xylose residues are thought to carry D-galactose in β-1, 2-linkage, and usually L-fucose substitute the second of these galactose residues (Reiter, 2002). Xyloglucan content is important for wood development. Park (2004) over-expressed xyloglucanase, which is able to degrade xyloglucan in poplar, and the loosening of xyloglucan in the cell wall promoted poplar plant growth and enhanced cellulose accumulation.

Pectin, another matrix polysaccharide in the plant primary cell wall, plays a fundamental role in the regulation of cell wall extensibility. McCartney *et al.* (2003) reported that the modification of pectin is strongly correlated with the alteration of cell

wall growth in expanding tissues of *Arabidopsis*. Pectin is exported, like hemi-cellulose, from the Golgi apparatus as highly esterified galacturonan which is then deesterified by cell wall-bound pectin methylesterases (Goujon *et al.*, 2003). Pectin biosynthesis is also a complex procedure and at least 53 different enzymatic activities are needed to synthesize pectic polysaccharides (Scheible, 2004).

Plant cell walls consist primarily of carbohydrates and polyphenolics as described above, but proteins are also important components and make up approximately 10% of the cell wall mass (Borderies *et al.*, 2003). Over 400 putative cell wall proteins have been identified in *Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000). Some of these are cell wall enzymes thought to contribute to the regulation of growth and development (Darley *et al.*, 2001), others are structural proteins that function in cell wall architecture (Cassab, 1998), and the rest are defence proteins that function in the response to biotic and abiotic stresses (Qin *et al.*, 2003).

1.3 Functional genetics of wood formation

As described above, wood formation is a complex biological process coordinated by large number of genes. To understand the functions of these genes, the tools of functional genetics, which is used to determine biological functions of the identified genes, need to be used in *Eucalyptus* and other woody species. This section will discuss the utilization of functional genetics tools, including molecular biology methods with the aid of bioinformatics and model systems used to study the genetics of wood formation.

1.3.1 Bioinformatics

High throughput sequencing, transcriptomic and proteomic studies are generating massive amounts of data and the volume of data is growing exponentially. Storing, utilizing and making sense of such a wealth of information is becoming a major challenge in all post-genomic sciences. Bioinformatics, the field of "research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioural or health data, including those to acquire, store, organize, archive, analyze, or visualize such data" (Becker, 2005), has become extremely important to process these biological data. For instance, bioinformatic tools can be used for plant DNA sequence comparison and gene prediction according to the degree of sequence homology. Also, gene regulation, translation profile, and protein interactions in plants can be determined using correlated technologies, as discussed in the following section, combined with the use of particular bioinformatics software.

1.3.2 Functional analysis of wood formation genes

1.3.2.1 Gene discovery

The nature of an organism is determined by the genes that it expresses. To analyse the functions of genes, the first step is to discover the genes that we are interested in. Obviously, the ultimate approach for gene discovery in plants is to sequence the whole genome of a plant species. Relatively cheaper and easier methods include the analysis of expressed sequence tags (ESTs) (Adams *et al.*, 1991) and cDNA-Amplified Fragment Length Polymorphism (cDNA-AFLP) (Vos *et al.*, 1995). More traditional methods, such as cDNA library screening using a particular probe, or degenerate PCR, are quicker to isolate new genes, but they can only be used to discover a relatively small number of

genes at one time.

Several plant genome sequencing projects have been started and some are complete. Major landmarks in plant biology included the first genomic sequence of *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative, 2000), and the genome sequences of two rice subspecies (Yu, 2002; Goff *et al.*, 2002). Also, the poplar genome sequence is now available on the internet (genome.jgi-psf.org/poplar/poplar.home.html). A genome sequencing project of *Eucalyptus camaldulensis* is in progress in Japan and it is believed that the genome sequence will be available within 3-4 years (S. Tabata, personal communication). The *Arabidopsis* genome contains about 30000 putative genes and gene predictions reveal that there are 32000 to 50000 genes in the rice genome. *Arabidopsis* and rice have been widely used as model plants for dicots (Sibout *et al.*, 2002; in this project) and monocots (Kainuma, 2003), respectively. The functions of numerous genes in *Arabidopsis* and rice have been determined using molecular methods. Many scientists begin their work by the analysis of orthologous genes in either *Arabidopsis* or rice to determine the functions of genes in the plant they work on.

The analysis of ESTs (Adams *et al.*, 1991) from complementary DNA (cDNA) libraries represents another large-scale gene discovery method in modern biological research. This is the most common format for gene discovery in higher eukaryotes. Millions of ESTs from different organisms, tissues and developmental stages have been generated. So far, at least 400 000 *Eucalyptus* ESTs have been sequenced by private organizations world-wide (Myburg, personal communication). Furthermore, more than 80 000 ESTs have been sequenced for pine (http://www.pinetree.ccgb.umn.edu) and more than 130000 for poplar (http://www.poppel.fysbot.umn.se). These EST projects were specifically set up for the study of wood formation in pine (Johnson *et al.*, 2002) and

poplar (Sterky *et al.*, 1998). More and more ESTs from different tissues, such as xylem and cambium of different woody species may also be used to elucidate the genetic basis for the differences in wood quality between different species.

High-throughput cDNA-AFLP technology (Vos *et al.*, 1995) is a robust genome-wide gene discovery tool. cDNA-AFLP begins with digestion of cDNA by two restriction enzymes and ligation of specific adapters. Adapter-specific primers are used to amplify the ligated fragments. cDNA-AFLP is able to amplify a limited number of fragments due to the use of selective nucleotides at the 3' end of the PCR primers, which ensures preferential amplification of only a subset of fragments. The differentiation and quantification of the same transcript derived from different tissues, developmental stages or environmental conditions therefore can be performed on AFLP gels. The transcript-derived fragments (TDFs) then can be sequenced after excision from the gel (Breyne *et al.*, 2003). This PCR-based method may identify new and rarely expressed transcripts. Using this tool, some genes have been isolated from plant and other organisms (Durrant *et al.*, 2000; Qin *et al*., 2000; Sutcliffe *et al.*, 2000; van der Biezen *et al.*, 2000; Kornmann *et al.*, 2001). In our lab, cDNA-AFLP has been used to study wood formation in *Eucalyptus* (Ranik *et al.*, unpublished) and six cellulose synthase genes have been identified using the original results from cDNA-AFLP analysis (Ranik and Myburg, unpublished).

Bioinformatics tools give us much information about the putative functions of discovered genes, but the precise functions of the genes and, more important, predicted orphan genes (i.e. genes without annotated functions) cannot be determined by bioinformatics tools. Transcriptomic and proteomic methods, and traditional transgenic technologies will help define the role of each gene in wood formation.

1.3.2.2 Transcription profiling

It is believed that genes involved in the same biological process will be co-expressed, so the transcription profiles of genes across different tissues and conditions are preliminary indicators of gene function. For the analysis of transcription pattern of small numbers of genes, Northern blotting and real-time RT-PCR can be employed. The approaches used for transcription profile analysis of large amounts of genes are classified into two categories: closed systems which are gene sequence-based and open systems which do not need any prior sequence information.

Examples of closed systems include cDNA microarray and oligonucleotide arrays. Both types of array experiments consist of five steps: (1) array construction, (2) biological sample (RNA) preparation, (3) hybridization of the nucleic acid sample onto the array, (4) signal detection, and (5) data mining and analysis (Schena *et al.*, 1995). cDNA microarray analysis is a transcriptional profiling technology based on the intensity of fluorescence signal, which comes from the complementary hybridization of the labelled cDNAs and the probes on the array (Schena *et al.*, 1995). cDNA microarray analysis has been widely employed in woody species to investigate transcription profiles during wood formation (Hertzberg *et al.*, 2001; Israelsson *et al.*, 2003; Yang *et al.*, 2004). We are using cDNA microarray to study the transcription pattern of wood formation genes in *Eucalyptus*. The major difference between cDNA and oligonucleotide arrays is that in oligonucleotide array technology, probes are synthesized *in situ* using photolithography (e.g. Affymetrix arrays) (Fodor *et al.*, 1993). cDNA microarrays are more widely used than oligonucleotide arrays because they are cheaper and sequence data (required for oligo design) are not available for many species.

cDNA-AFLP and serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995)

are typical examples of open systems. cDNA-AFLP analysis can be used to reveal spatial and temporal expression patterns as explained above. SAGE, the complement of cDNA microarrays and oligonucleotide arrays, is a technique that allows rapid, quantitative analysis of complete transcriptomes (Velculescu *et al.*, 1995). In a SAGE experiment, 3' derived short (9-14 bp) cDNA fragments in one tissue is ligated and sequenced, and each gene is represented by the same cDNA tag. Therefore, the frequency of each cDNA tag represents the abundance of the corresponding transcript. Although SAGE analysis has been employed for the study of wood formation in pine (Lorenz and Dean; 2002) and some other plants (Fizames *et al.*, 2004; Gibbings *et al.*, 2003), it is not as widely used as cDNA-AFLP, probably due to the technical complexity and the cost of high-throughput sequencing required.

1.3.2.3 Protein function analysis

Transcription profiling reveals the comparable abundance of transcripts, but it does not always provide information about the abundance or structure of corresponding proteins. There is often not a close relationship between mRNA and protein abundance due to translational and post-translational modifications. Therefore, gene functions need to be studied further at the protein level. Traditional and newer genome-wide protein function analysis tools can be used to achieve this.

More traditional biochemistry methods, such as *in vitro* expression of the protein in *E. coli* cells, protein purification and enzyme assays, are the best ways to study the functions of single or small numbers of proteins. The first *Cinnamoyl-CoA reductase* (*CCR*) (a key lignin biosynthesis gene) gene was isolated from *Eucalyptus* by Lacombe *et al.* (1997). The function of this *Eucalyptus CCR* (*EUCCR*) gene was preliminarily analyzed by studying the expression of *EUCCR* cDNA in *E. coli*. The purified EUCCR

enzyme was then used to determine the K_m value for its natural substrates and for its co-factor. The data from the K_m value determination was in agreement with the predicted value of the native EUCCR enzyme. These biochemistry methods were also used to analyze the function of sucrose synthase, a key cellulose biosynthesis gene (Winter *et al.*, 1997; Zhang *et al.*, 1999).

Protein electrophoresis and mass spectrometry are the most widely used technologies to profile large numbers of proteins. The interaction of protein with other molecules represents another level of gene function. Protein electrophoresis and mass spectrometry are used together to study protein expression patterns and this is complemented by protein microarray analysis (Angenendt, 2005). Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) allows the parallel separation and analysis of proteins extracted from complex samples. The intensity of each protein revealed as an individual spot on the gel indicates the abundance of the protein in the sample. The amino acid sequences of the proteins on the gel are determined using modified mass spectrometry, matrix-assisted laser desorption ionization time-of-light (MALDI-TOF) (Yates, 2000). The corresponding protein can be searched in the public databases, only if similar proteins have previously been characterized. These technologies have been employed for wood formation studies in poplar (Lu *et al.*, 2003). The activity of a protein in the plant, tissue or even the single cell is complex and the protein is not insular. In these systems, it often needs to work co-ordinately and interact with other molecules such as other proteins, peptides, low molecular weight compounds, oligosaccharides or DNA (Poetz *et al.*, 2005). Protein microarrays allow researchers to study these interactions (Angenendt, 2005). Although the use of this method to study wood formation has not been reported, the potential of this technology for wood formation study is apparent.

1.3.2.4 Loss-of-function and gain-of-function analysis of gene function

The approaches described above are not sufficient to understand the function of all individual genes and interactions among them. Loss-of-function and gain-of-function tools are needed for functional genomics to reveal the direct linkage between gene and phenotype, and are based on the disruption or overexpression of native genes, respectively.

Many different non-transgenic and transgenic approaches have been developed for loss-of-function analysis. Large mutant populations have been constructed for *Arabidopsis* (http://www.arabidopsis.org) using non-transgenic tools, such as ethyl methane sulfonate (EMS) mutagenesis, which usually induces G to A transitions, and fast neutron bombardment, that causes small deletions (Weigel and Glazebrook, 2002). T-DNA insertion mutagenesis represents the major transgenic technology for functional genomics of *Arabidopsis* (http://www.arabidopsis.org) and rice (http://www.postech.ac.kr/life/pfg/tagging). Introduction of heterologous transposons into the *Arabidopsis* genome is also frequently used to identify gene functions (Weigel and Glazebrook, 2002). More recently, RNA interference (RNAi) tools have been proven to be robust to specifically knock down genes and identify functions with high throughput in plants combined with the convenient Gateway recombination cloning system (Helliwell *et al.*, 2002). Loss-of-function techniques result in the inactivation of the gene investigated. The subsequent phenotype, if any, provides us some information on gene function. For instance, the *Arabidopsis* cell wall mutants, *irregular xylem 1* (*irx1*) (Taylor *et al.*, 2000) and *irregular xylem 4* (*irx4*) (Jones *et al.*, 2001), have remarkably collapsed cell walls. These suggested the functions of the mutated genes, which will be discussed below, for vascular development.

A gain-of-function method causes a gene to be overexpressed in a tissue and/or developmental stage where the gene normally does not express or expresses weakly. Similarly, the change of phenotype provides the researchers with information on gene function. However, it needs to be determined if phenotypical changes are derived from the specific function of an overexpressed gene, or if they are less specific changes due to the abnormality of the situation where excessive gene products are produced. This is a complementary approach to loss-of-function methods to analyze gene function.

It is difficult to develop mutant populations for *Eucalyptus* due to the obstacles of obtaining transgenic plants, long generation times, difficulty of inbreeding and rarity of Mendelian traits in wild populations. Therefore the use of model systems, which will be discussed below, is crucial to understand wood formation in this plant species even after the whole genome sequence project of *Eucalyptus* is finished.

1.3.3 Model systems for wood formation

1.3.3.1 *Arabidopsis*

During the last decade, *Arabidopsis thaliana* has become universally recognized as a model plant system. Although it is a non-commercial member of the mustard family, it is widely studied because it develops, reproduces, and responds to environmental stimuli in much the same way as many crop plants, including forest trees (Nieminen *et al.*, 2004). What is more, *Arabidopsis* is easy and inexpensive to grow, and produces many seeds; this allows extensive genetic and molecular experiments (arabidopsis-p450.biotec.uiuc.edu). Also, *Arabidopsis* has a comparatively small, completely sequenced genome and this simplifies and facilitates biological analysis. Compared to other plants species, it lacks highly repeated and less-informative DNA

sequences that complicate genome analysis (Brown, 2002).

Although *Arabidopsis* is a herbaceous species, it is has previously been shown to undergo secondary growth in hypocotyls (Lev-Yadun, 1995; Chaffey *et al.*, 2002) and inflorescence stems (Ko *et al.*, 2004) due to the fact that vascular cambium of *Arabidopsis* takes on a similar pattern to that in woody species (Busse and Evert 1999 a,b). The *Arabidopsis* hypocotyl undergoes substantial secondary growth under appropriate growth conditions that delay senescence, i.e. repeated cutting in inflorescence stems or the use of short day light cycle (Chaffey *et al.*, 2002). Extensive amounts of secondary xylem fibres and vessel elements are produced, and a continuous cambial zone is formed. Their structural and ultrastructural characteristics are highly similar to those in angiosperm trees (Chaffey *et al.* 2002). Recently, Ko *et al.* (2004) reported that *Arabidopsis* body weight is able to induce secondary growth in inflorescence stems probably due to the activity of phytohormones.

How can we use *Arabidopsis* as a model plant to analyze the functions of wood formation genes involved in tree species such as *Eucalyptus*? The first step is mutation analysis (loss-of-function analysis) and/or overexpression analysis (gain-of-function analysis) of the orthologous gene in *Arabidopsis thaliana*. Numerous mutants, which are involved in cell wall formation, have been identified in *Arabidopsis* (http://www.arabidopsis.org). For instance, *irregular xylem 4* (*irx4*) is the representative mutation which occurs in the *Arabidopsis Cinnamoyl-CoA reductase1* (*AtCCR1*) gene (Jones *et al.*, 2001). This mutant has collapsed cell walls and decreased lignin content (Jones *et al.*, 2001). The ortholog of *AtCCR1* has been isolated from *Eucalyptus* (Lacombe *et al.*, 1997). Three secondary cell wall related cellulose synthase mutants of *Arabidopsis*, *irregular xylem 1* (*irx1*) (Taylor *et al.*, 2000), *irregular xylem 3* (*irx3*)

(Taylor *et al.*, 1999) and *irregular xylem 5* (*irx5*) (Taylor *et al.*, 2003) are caused by mutations in three *Arabidopsis* cellulose synthase genes, *Arabidopsis cellulose synthase 8* (*AtCesA8*), *Arabidopsis cellulose synthase 7* (*AtCesA7*) and *Arabidopsis cellulose synthase 4* (*AtCesA4*), respectively. All of these mutants have remarkably decreased cellulose content in their cell walls. The primary cell wall related *Arabidopsis* cellulose synthase mutant, *ixr1* (Scheible *et al.*, 2001), is derived from the mutations in *Arabidopsis cellulose synthase 3* (*AtCesA3*). The homologs of *AtCesA8*, *AtCesA7*, *AtCesA4* and *AtCesA3* have also been isolated from *Eucalyptus* (Ranik and Myburg, unpublished). The overexpression of *Arabidopsis* cell wall formation genes can also provide investigators with some information about gene functions. The secondary step is to express *Eucalyptus* wood formation genes in *Arabidopsis* using a simple transformation method, such as the floral dip method (Figure 1.3, Clough and Bent, 1998). The *Eucalyptus* CCR gene and cellulose synthase genes can be transformed into wild type and corresponding mutants for overexpression and complementation analysis. Here, complementation analysis means the transformation of homologous genes into corresponding mutants to analyze the gene function. The functions of other wood formation genes in *Eucalyptus* can be analyzed using the same methods as described above. The overexpression method can be used independently if corresponding mutants are not available after new wood formation genes are isolated from *Eucalyptus*. Some genes, such as transcription factor and cell signalling genes, may not function appropriately in *Arabidopsis*. The functions of these genes need to be investigated in other model systems.

1.3.3.2 Other systems

Other systems, besides *Arabidopsis*, are frequently used to study wood formation. A

'model' tree is needed to accelerate domestication and understand some woody plant-specific processes such as dormancy and ray cells formation (Taylor, 2002). Poplar can be used for these studies. Poplar trees, in addition to their value for wood products, are one of the few trees which can be easily transformed, regenerated, vegetatively propagated and grown to tree size in a short time (Brunner *et al.*, 2004). That facilitates the investigation of wood formation and the functions of correlated genes. Like *Arabidopsis*, the genome sequence of poplar (http://genome.jgi-psf.org/poplar/poplar.home.html) is also available as described above.

Zinnia elegans, a herbaceous species as well, is used as a model system for the study of tracheary element (TE) formation. TEs can be induced from individual leaf mesophyll cells of *Zinnia elegans* by appropriate manipulation of the culture medium and the formation of xylem cells can be easily analyzed in vitro (Fukuda and Komamine, 1980). Cotton fibre is especially used to study cellulose biosynthesis, because mature cotton fibre consists mainly of secondary cell walls and contains at least 90% crystalline cellulose by weight (Haigler *et al.*, 2001). Obviously, this property is an obstacle for the study of other wood formation process, such as lignin biosynthesis, in this system, because lignin consists of 25%-35% in forest trees (Plomion *et al.*, 2001). Also, the single cell-derived cotton fibre is initiated from the ovule epidermis without cambial activity (Basra and Malik, 1984). Therefore this system cannot be used for the study of cambial activity which is crucial for wood formation.

Actually, *Eucalyptus per ser* can also be used as a model system to study the wood formation process *in vivo*. Spokevicius *et al.* (personal communication) transformed constructs containing *Eucalyptus* wood formation gene or promoters into the cells of

cambial zone (CZ) of *Eucalyptus* tree by pipeting the *Agrobacterium* solution containing these constructs into CZ. Subsequently, the functions of wood formation gene or promoter can be analyzed in different tissues, such as xylem, cambium and phloem, due to the differentiation fate of cambial cells as described above.

1.4 Challenges

Our understanding of wood formation is far from complete, although much progress has been made over the past years. Some "black boxes" are left in our understanding of lignification. The factors which control the initiation and heterogeneous deposition of the various monomers are still unknown (Anterola and Lewis, 2002). Few genes involved in this complex process have been isolated so far. Current monolignol biosynthesis pathways are mainly based on substrate specificity studies with assumed substrates that are readily available during wood formation. Although numerous monolignol biosynthesis genes have been identified, whether all of these genes are important for lignification and how they work coordinately to provide substrates for the proteins in the next step *in vivo* are not completely clear (Boerjan *et al.*, 2003). More detailed experiments need to be designed to reveal the functions of these monolignol biosynthesis genes deeply in a single tissue and/or a single species. Lignin biosynthesis is controlled temporally and spatially. The isolation of transcription factors, such as MYB (Tamagnone *et al.*, 1998) and LIM (Kawaoka and Ebinuma, 2001), proved these kinds of control at transcriptional level, but the potential posttranscriptional, translational, and/or posttranslational control also need to be investigated. Similar control mechanisms should also exist for cellulose biosynthesis.

We do not know all the components of the cellulose synthase complex and we do not know how they act together to produce cellulose. The coordinated regulation of

cellulose synthases and other genes that provide the sugars, such as sucrose synthase, is also not well understood. How the glucan strands penetrate through the cell membranes and how crystallization happens are uncertain. Which factors control the movement of the rosette and consequently control the angle of microfibrils need to be investigated further. These uncertainties are mostly derived from the difficulty to isolate pure cellulose synthase proteins due to their membrane-embedded property. This is and will still be a big challenge in future cellulose biosynthesis research.

FIGURES

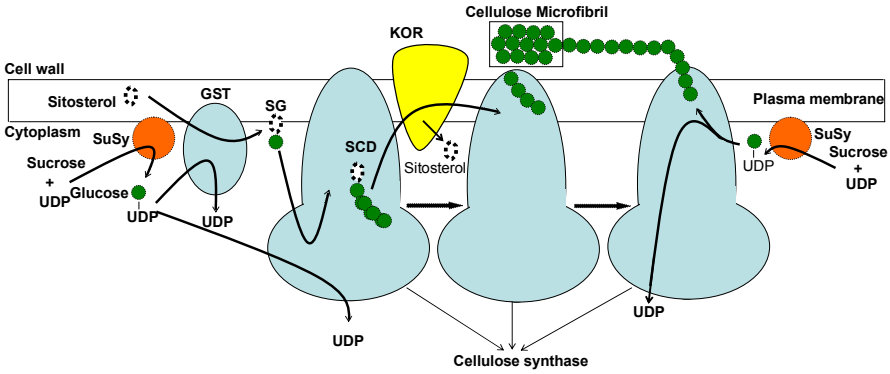


Figure 1.1 Proposed pathway of plant cellulose biosynthesis. GST: UDP-glucose sterol glucosyltransferase; SG: sitosterol-β-glucoside; SCD: sitosterol-cellodextrin; KOR: KORRIGAN; SuSy: sucrose synthase. SG is synthesized by GST using sitosterol and UDP-glucose as substrates. SCD is synthesized by cellulose synthase after the addition of glucose residuals to the end of SG. Sitosterol is then removed from SCD by KOR and cellulose synthase catalyzes the elongation of glucan chain by adding glucose residuals to the short glucan chain, which is released by KOR. The glucose residual, represented by the solid green dot, is derived from the UDP-glucose, which is synthesized by SuSy.

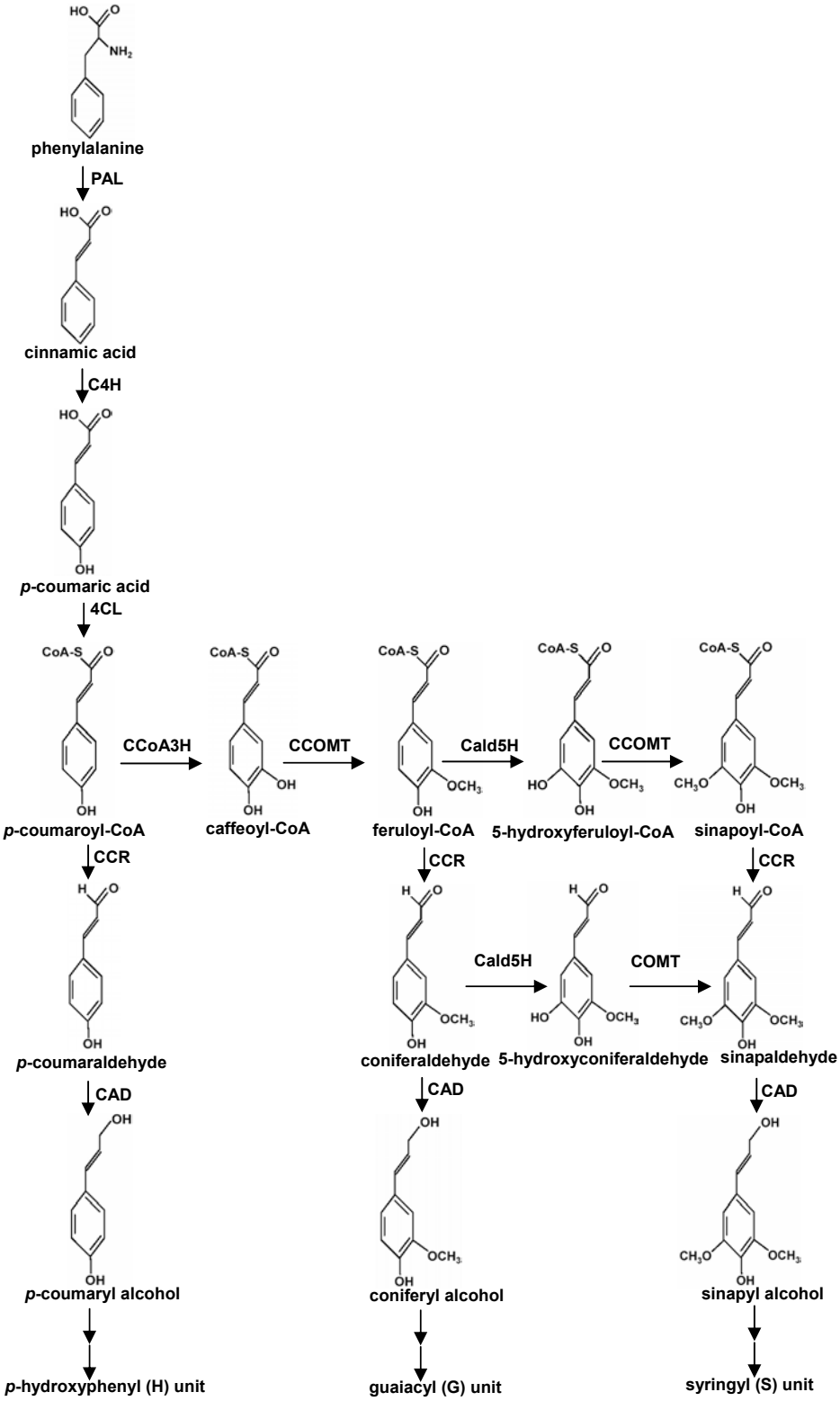


Figure 1.2 Monolignol biosynthesis pathways. PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate:CoA ligase; CCoA3H: coumaroyl-CoA 3-hydroxylase; CCOMT: caffeoyl CoA 3-O-methyltransferase; Cald5H: coniferyl aldehyde-5-hydroxylase; CCR: cinnamoyl-CoA reductase; COMT: caffeic acid O-methyltransferase; CAD: cinnamyl alcohol dehydrogenase.

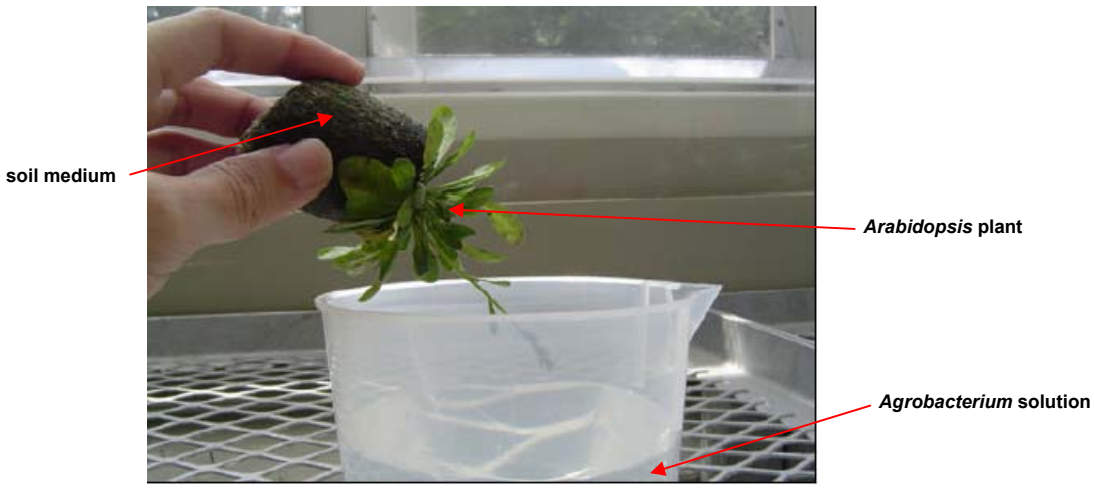


Figure 1.3 Transformation of *Arabidopsis* plants using the floral dip method (Clough and Bent, 1998). *Arabidopsis* plant grown on the soil medium was dipped into the *Agrobacterium* solution containing *Agrobacterium*, sucrose and Silwett L-77.

1.5 References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252, 1651-1656

Albrecht GO (2000) The increase in sucrose synthase activity correlates with a higher content of cellulose in wheat roots suffering from oxygen deficiency. Abstract 1020. In: Proceedings of Plant Biology 2000, 15-19 July, San Diego, CA. American Society of Plant Physiologists, Rochville. MD.

Aloni R (2001) Foliar and axial aspects of vascular differentiation-hypotheses and evidence. *J. Plant Growth Reg.* 20, 22-34

Angenendt P (2005) Progress in protein and antibody microarray technology. *Drug Discovery Today* 10, 503-511

Anterola AM, Lewis NG (2002) Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221-294

Basra A, Malik CP (1984) Development of the cotton fiber. *Int. Rev. Cytol.* 89, 65-113

Baucher M, Monties B, Van MM, Boerjan W (1998) Biosynthesis and genetic engineering of lignin. *Critical Reviews in Plant Sciences* 17, 125-197

Baud S, Vaultier MN, Rochat C (2004) Structure and expression profile of the sucrose synthase multigene family in *Arabidopsis*. *J. Ex. Bot.* 55, 397-409

Becker GJ (2005) Bioinformatics, computational biology, and the national centers for biomedical computing. *Journal of the American College of Radiology* 2, 398-400

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519-546

Borderies G, Jamet E, Lafitte C, Rossignol M, Januneau A, Boudart G, Monsarrat B, Esquerré-Tugayé MT, Boudet A, Pont-Lezica R (2003) Proteomics of loosely bound cell

wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis. *Electrophoresis* 24, 3421-3432

Boudet AM, Lapierre C, Grima-Pettenati J (1995) Biochemistry and molecular biology of lignification. *New Phytol.* 129, 203-236

Boudet AM (2000) Lignins and lignification: Selected issues. *Plant Physiol. Biochem.* 38, 81-96

Breyne P, Dreesen R, Cannoot B, Rombaut D, Vandepoele K, Rombauts S, Vanderhaeghen R, Inzé D, Zabeau M (2003) Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Mol. Gen. Genomics* 269, 173-179

Brooker MIH, Slee AV (1996) Eucalyptus. In Walsh, N.G. and Entwisle, T.J. (Eds.) Flora of Victoria, Volume 3, Dicotyledons: Winteraceae to Myrtaceae, Inkata Press, Melbourne

Brown TA (2002) Genomes (2nd ed) BIOS Scientific Publishers Ltd, Oxfrod.

Brunner AM, Busov VB, Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends in Plant Science* 9, 49-56

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

Busse JS, Evert RF (1999a) Pattern of differentiation of the first vascular elements in the embryo and seedling of *Arabidopsis thaliana*. *Int. J. Plant Sci.* 160, 1-13

Busse JS, Evert RF (1999b) Vascular differentiation and transition in the seedling of *Arabidopsis thaliana* (Brassicaceae). *Int. J. Plant Sci.* 160, 241-251

Campbell MM, Sederoff RR (1996) Variation in lignin content and composition: mechanisms of control and implications for the genetic improvement of plants. *Plant Physiol*. 110, 3-13

Cassab GI (1998) Plant cell wall proteins. *Annu. Rev. Plant Biol.* 49, 28-309

Chaffey NJ, Cholewa E, Regan S, Sundberg B (2002) Secondary xylem development in *Arabidopsis*: a model for wood formation. *Physiologia Plantarum* 114, 594-600

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735-743

Darley CP, Forrester AM, McQueen-Mason SJ (2001) The molecular basis of plant cell wall extension. *Plant Mol. Biol.* 47, 179-195

Delmer PD, Ravit E, Pat H, Monika D, Debbie JW, Liangcai P, Alison R, Doron H, (2001) Abs # 510: Expression analysis of the ten cellulose synthase (CesA) genes of *Arabidopsis thaliana*. Plant Biology 2001 July 21 - July 25, Providence, Rhode Island, USA

Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a ‘‘metabolic grid’’, or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 57, 1069–1084

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

Durrant WE, Rowland O, Piedras P, Hammond-Kosack KE, Jones JDG (2000) cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. *Plant Cell* 12, 963–977

Esau K (1965) Plant anatomy. New York: Wiley

Fenning TM, Gershenzon J (2002) Where will the wood come from? Plantation forests and the role of biotechnology. *Trends Biotechnol.* 20, 291–296

Fizames C, Munos S, Cazettes C, Nacry P, Boucherez J, Gaymard F, Piquemal D, Delorme V, Commes T, Doumas P, Cooke R, Marti J, Sentenac H, Gojon A (2004) The *Arabidopsis* root transcriptome by serial analysis of gene expression; gene identification using the genome sequence. *Plant Physiol.* 134, 67-80

Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL (1993) Multiplexed biochemical assays with biological chips. *Nature* 5, 555-556

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

Fukuda H (1996) Xylogenesis: Initiation, progression, and cell death. *Annu. Rev. Plant Biol.* 47, 299-325

Gibbings JG, Cook BP, Dufault MR, Madden SL, Khuri S, Turnbull CJ, Dunwell JM (2003) Global transcript analysis of rice leaf and seed using SAGE technology. *Plant Biotechnology Journal* 1, 271-285

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook, J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza* sativa L. ssp. *japonica*) *Science* 296, 92-100

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

Grattapaglia D (2004) Integrating genomics into *Eucalyptus* breeding. *Genet. Mol. Res.* 3, 369-379

Haigler CH, Ivanova-Datcheva M, Hogan PS, Salnikov VV, Hwang S, Martin K, Delmer DP (2001) Carbon portioning to cellulose synthesis. *Plant Mol. Biol.* 47, 29-51

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhle´n M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A transcriptional roadmap to wood formation. *PNAS* 98, 14732-14737

Israelsson M, Eriksson ME, Hertzberg M, Aspeborg H, Nilsson P, 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

Johnson AF, Baucom C, Brawner A, Golden K, Hubbard KM, Peter G, Retzel EF, Stambaugh A, Whetten RW, Sederoff RR (2002) The *Pinus taeda* L. (loblolly pine) EST sequencing project. Plant, animal and microbe genomes X Conference January 12-16, Town and Country Convention Center, San Diego, CA

Jones L, Ennos AR, Turner SR (2001) Cloning and characterization of *irregular xylem 4* (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*. *Plant J*. 26, 205-216

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

Kainuma K (2003) Completion of the decoding of rice genome sequence and prospects for genetically modified (GM) foods. *Foods Food Ingredients J. Jpn.* 208, 1

Kawaoka A, Ebinuma H (2001) Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry* 57, 1149-1157

Kleczkowski LA (1994) Glucose activation and metabolism through UDP-glucose pyrophosphorylase in plants. *Phytochemistry* 37, 1507-1515

Koch K (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Current Opinion in Plant Biology* 7, 235-246

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

Komatsu A, Moriguchi T, Koyama K, Omura M, Akihama T (2002) Analysis of sucrose synthase genes in citrus suggests different roles and phylogenetic relationships. *J. Ex. Bot.* 53, 61-71

Kornmann B, Preitner N, Rifat D, Fleury-Olela F, Schibler U (2001) Analysis of circadian liver gene expression by ADDER, a highly sensitive method for the display of differentially expressed mRNAs. *Nucleic Acids Res.* 29, E51-1

Lacombe E, Hawkins S, Doorsselaere JV, Piquemal J (1997) Cinnomyl-CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression, and phylogenetic relationships. *Plant J.* 11, 429-441

Larson PR (1994) The Vascular Cambium. Springer-Verlag, Berlin

Lev-Yadun S (1995) Short secondary vessel members in branching regions in roots of *Arabidopsis thaliana*. *Aust. J. Bot.* 43, 435-438

Lorenz WW, Dean JFD (2002) SAGE Profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). *Tree Physiology* 22, 301-310

Lu MZ, Xie HL, Du J, Cui KM (2003) Protein profiling on regenerating tissues from girdled trunk of *Populus tomentosa*. IUFRO Tree Biotechnology June 7-12 Umeå, Sweden

McCartney L, Steele-King CG, Jordan E, Knox JP (2003) Cell wall pectic (1-4)-β-D-galactan marks the acceleration of cell elongation in the *Arabidopsis* seedling root meristem. *Plant J.* 33, 447-454

Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W (2001) Unravelling cell wall formation in the woody dicot stem. *Plant Mol. Biol.* 47, 239-274

Moore KJ, Jung HG (2001) Lignin and fiber digestion. *J. Range Manage* 54, 420-430

Nicol F, His I, Jauneau A, Vernhettes S, Canut H, 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

Nieminen KM, Kauppinen L, Helariutta Y (2004) A weed for wood? *Arabidopsis* as a genetic model for xylem development. *Plant Physiol.* 135, 653-659

Park YW (2004) Enhancement of growth and cellulose accumulation by overexpression of xylogucanase in poplar. *FEBS Letters* 564, 183-187

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

Peng LC, Kawagoe Y, Hogan P, Delmer D (2002) Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. *Science* 295, 147-150

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

Poetz O, Schwenk JM, Kramer S, Stoll D, Templin MF, Joos TO (2005) Protein microarrays: catching the proteome. *Mechanisms of Ageing and Development* 126, 161-170

Qin L, Overmars H, Helder J, Popeijus H, Rouppe van der Voort J, Groenink W, van Koert P, Schots A, Bakker J, Smant G (2000) An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. *Mol. Plant-Microbe Interact* 13, 830-836

Qin Q, Bergmann CW, Rose JKC, Saladie M, Kolli VSK, Albersheim P, Darvill AG, York WS (2003) Characterization of a tomato protein that inhibits a xyloglucan-specific endoglucanase. *Plant J*. 34, 327-338

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

Reiter WD (2002) Biosynthesis and properties of the plant cell wall. *Current Opinion in Plant Biology* 5, 536–542

Richmond TA (2000) Higher plant cellulose synthases. *Genome Biology* 1, 1-6

Sanchez JR (2002). Panel de tecnologia, pasta y papel. In: Grupo de Investigacion AF-4 - Universidade de Vigo. Socioeconomia, Patologia, Tecnologia y Sostenibilidade del Eucalipto. Ed. Universidad de Vigo, Graficas Diumaró, Vigo, Spain, Vol. 1. pp. 187-192

Santos RL (1997) The *Eucalyptus* of California. Alley-Cass Publications, California

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

Scheible WR, Eshed R, Richmond T, Delmer D, Somerville C (2001) Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis ixr1* mutants. *PNAS* 98, 10079-10084

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470

Sibout R, Baucher M, Gatineau M, Doorsselaere JV, Mila I, Pollet B, Maba B, Pilate G, Lapierre C, Boerjan W, Jouanin L (2002) Expression of a poplar cDNA encoding a ferulate-5-hydroxylase/coniferaldehyde 5-hydroxylase increases S lignin deposition in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 40, 1087-1096

Smook GA (1992) Handbook for Pulp and Paper Technologists. Angus Wilde Publications, Vancouver

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 TT, Boerjan W, Gustafsson P, Uhlén M, Sundberg B, Lundeberg J (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. *Plant Biology* 95, 13330-13335

Sundberg B, Uggla C, Tuominen H (2000) Cambial growth and auxin gradients. In: Savidge RA, Barnett JR, Napier R, editors. Cell and molecular biology of wood formation. Oxford: BIOS Scientific Publishers. 169-188

Sutcliffe JG, Foye PE, Erlander MG, Hilbush BS, Bodzin LJ, Durham JT, Hasel KW (2000) TOGA: an automated parsing technology for analyzing expression of nearly all genes. *PNAS* 97, 1976-1981

Tamagnone L, Merida A, Parr A, MacKay S, Culianez-Macia FA, Roberts K, Martin C (1998) The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *The Plant Cell*

10, 135-154

Taylor G (2002) *Populus*: *Arabidopsis* for forestry. Do we need a model tree? *Ann. Bot.* 90, 681-689

Taylor NG, Howells RM, Huttly A, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. *PNAS* 100, 1450-1455

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

Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR (1999) The *irregular xylem 3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell* 11, 769-779

Theander O, Westerlund E (1993) Quantitative analysis of cell wall components. In: H.G. Jung. D.R. Hatfield and J.Ralph (Eds.) Forage Cell Wall Structure and Digestibility. American Society of Agronomy/SCCA/SSSA. Madison. WI. 83-104

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815

Tibbits WN, Boomsma DB, Jarvis S (1997) Distribution, biology, genetics and improvement programs for *Eucalyptus globulus* and *E. nitens* around the world. Proceedings of the 24th Southern Forest Tree Improvement Conference. 81-95

Timell TE (1986) Compression wood in gymnosperms. Springer-Verlag, Heidelberg

Van der Biezen EA, Juwana H, Parker JE, Jones JDG (2000) cDNA-AFLP display for the isolation of *Peronospora parasitica* genes expressed during infection in *Arabidopsis thaliana*. *Mol. Plant-Microbe Interact* 13, 895-898

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (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, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407-4414.

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

Williamson RE, Burn JE, Hocart CH (2001) Cellulose synthesis: mutational analysis and genomic perspectives using *Arabidopsis thaliana*. *Cell Mol. Life Sci.* 58, 1475-1490

Winter H, Huber JL, Huber SC (1997) Membrane association of sucrose synthase: changes during the graviresponse and possible control by protein phosphorylation. *FEBS Letters* 420, 151-155

Yang SH, van Zyl L, No EG, Loopstra CA (2004) Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). *Plant Science* 166, 1185-1195

Yates JR (2000) Mass spectrometry: from genomics to proteomics. *Trends Genet.* 16, 5-8

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza* sativa L. ssp. *indica*). *Science* 296, 79-92

Zhang XQ, Lund AA, Sarath G, Cerny RL, Roberts DM, Chollet R (1999) Soybean nodule sucrose synthase (nodulin-100): further analysis of its phosphorylation using recombinant and authentic root-nodule enzymes. *Archives of Biochemistry and*

Biophysics 371, 70-82

Chapter Two

Establishment of *Arabidopsis thaliana* as a model system for functional analysis of *Eucalyptus* wood formation genes

2.1 Introduction

Although *Eucalyptus* is an economically important fibre crop that is widely used in the pulp and papermaking industry, our understanding of the mechanism of wood formation in *Eucalyptus* trees is far from complete due to several obstacles. To fully understand this mechanism, the functions of wood formation genes will ultimately need to be investigated through transgenic analysis in *Eucalyptus* trees. However it is currently difficult to transform *Eucalyptus* trees. It takes at least 24 months to see a reliable wood phenotype of a wood formation gene, after transformation of the gene construct to *Eucalyptus* and successful regeneration of transgenic plants (Tzfira *et al.*, 1998). What is more, transformation efficiency is very low in *Eucalyptus* trees and requires a substantial investment in tissue culture facilities. A model system such as *Arabidopsis thaliana* provides an useful and robust tool for the functional analysis of wood formation genes in *Eucalyptus*.

Arabidopsis is a herbaceous species, but secondary xylem can be easily induced in hypocotyls and inflorescence stems. Under appropriate growth conditions that delay senescence and promote vegetative growth, such as repeated cutting of inflorescence stems and the use of short daylight cycle, *Arabidopsis* hypocotyls undergo secondary growth and the anatomy is comparable with that of an *Eucalyptus* seedling (Chaffey *et al.*, 2002). Ko *et al.* (2004) reported that *Arabidopsis* body weight is able to induce secondary xylem in inflorescence stems probably due to the induction of phytohormones such as auxin.

Large mutant populations have been constructed for *Arabidopsis* (http://www.arabidopsis.org) using non-transgenic tools, such as ethyl methane sulfonate (EMS) mutagenesis, which usually induces G to A transitions, and fast

neutron bombardment, that causes small deletions (Weigel and Glazebrook, 2002). T-DNA insertion mutagenesis represents the major transgenic technology for functional genomics of *Arabidopsis* (http://www.arabidopsis.org). Introduction of heterologous transposons into the *Arabidopsis* genome is also frequently used to identify gene functions (Weigel and Glazebrook, 2002).

Numerous mutants involved in vascular development have been identified in *Arabidopsis* (http://www.arabidopsis.org). For instance, *irregular xylem 4* (*irx4*) plants contain a mutant allele of the *Cinnamoyl-CoA reductase1* (*AtCCR1*) gene in *Arabidopsis* (Jones *et al.*, 2001). This mutant has collapsed cell walls and decreased lignin content, because *CCR* is an important lignin biosynthesis gene. Similarly, the secondary cell wall related cellulose synthase mutant of *Arabidopsis*, *irregular xylem1* (*irx1*, Taylor *et al.*, 2000) is caused by mutations in *Arabidopsis cellulose synthase 8* (*AtCesA8*). This mutant has remarkably decreased cellulose content in secondary cell walls. Vascular development mutants such as these can be used to confirm the function of homologous genes in other species by complementation of the *Arabidopsis* mutation.

In this study, we evaluated the possibility of using *Arabidopsis thaliana* as a model system for functional analysis of *Eucalyptus* wood formation genes. In particular, we wanted to determine the possibility of changing the cell wall chemistry of *Arabidopsis* plants in a predictable fashion using genes isolated from *Eucalyptus* trees. To achieve this, the *Eucalyptus CCR* gene (*EUCCR*) (Lacombe *et al.*, 1997), which is the ortholog of *AtCCR1*, was transformed into the *irx4* mutant and wild-type Col-0 *Arabidopsis*. The *Eucalyptus cellulose synthase 1* (*EgCesA1*) gene (Ranik and Myburg, unpublished), the ortholog of *AtCesA8*, was transformed into the *irx1* mutant. The traditional (Clough and Bent, 1998) and improved (Briones *et al.*, 2004) floral dip transformation methods were

used. The resulting transgenic plants were subjected to microscopic and chemical analysis. This chapter describes the establishment of the *Arabidopsis* model system in our laboratory and serves partly as complementary materials and methods for Chapter three, which is prepared as a short research paper.

2.2 Materials and Methods

2.2.1 *Arabidopsis* growth conditions

The same methods and conditions were used to grow *Arabidopsis thaliana* L. Heynh. ecotype Columbia (Col-0), Landsberg erecta (Ler-0), *irx1* and *irx4* plants. The seeds for these plants were ordered from Nottingham *Arabidopsis* Stock Centre (http://arabidopsis.info). Seeds were surface sterilized with 1.5% (w/v) sodium hypochlorite and inoculated on Petri plates containing 1% (w/v) Murashige and Skoog medium (Murashige and Skoog, 1962) in 0.8% (w/v) gelled agar supplemented with 2% (w/v) sucrose. The young seedlings were transferred to soil medium (Jiffy Products International AS) and grown in growth chamber at 22°C and 8 h photoperiod (100 µmol/m^2/s).

2.2.2 Vector construction

The construction of plant transformation vectors required several intermediate steps. The pEUCCR plasmid, containing *EUCCR* cDNA, was kindly provided by Jacqueline Grima-Pettenati (UMR UPS/CNRS 5546, France). pEgCesA1 contained a full-length copy of *EgCesA1* cDNA, which was isolated by Ranik and Myburg (unpublished) in this lab. Gene-specific primers containing suitable restriction sites for cloning into

intermediate vector pRTL2 were designed for the two genes (Table 2.1). Two recombinant expression cassettes containing *EUCCR* and *EgCesA1* cDNAs, respectively, were under the control of a double Cauliflower Mosaic Virus (CaMV) 35S promoter and CaMV 35S terminator. The recombinant cassettes were subsequently cloned into the *Agrobacterium*-based plant transformation vector, pCAMBIA 3301. The detailed information about vector construction is provided below.

2.2.2.1 Construction of the expression vectors

Primer design and PCR amplification of the EUCCR and EgCesA1

EUCCR-specific primers, CCRcDNA5UTR-F (5'UTR-specific and containing an *Eco*RI site) and CCRcDNA3UTR-R (3'UTR-specific and containing a *Bam*HI site) (Table 2.1) were designed to amplify *EUCCR* cDNA from pEUCCR. pEUCCR consists of a plasmid backbone, the pBluescript sk^+ plasmid, and *EUCCR* cDNA. The PCR reaction mixture included 0.16 units of Expand Taq polymerase (Roche Diagnostics GmbH), 1× reaction buffer, 15 ng of pEUCCR plasmid DNA as template, 0.25 mM dNTPs and 0.4 µM of each primer in a 20 µl reaction volume. The thermal cycler (BIO-RAD) was programmed to include one denaturing cycle at 95°C for 1 min, followed by 25 cycles of denaturation at 94°C for 30 sec, annealing for 30 sec at 59°C and then primer extension at 72°C for 1 min. A final step at 72°C for 20 min was included for the elongation.

The primers, 5+7*Eco*RI (5'UTR-specific and containing an *Eco*RI site) and 3'UTR *Sma*I (3'UTR-specific and containing *Xma*I site) (Table 2.1) were used to amplify *EgCesA1* from pEgCesA1, which consists of the TOPO plasmid (TOPO TA Cloning Kit, Invitrogen) and *EgCesA1* cDNA. The annealing was performed at 55°C and primer

extension was performed for 3 min per cycle.

Colony PCR and sequence confirmation

The PCR products were cloned into the pTZ57R vector (InsT/Aclone PCR Cloning Kit, Fermentas). Colony PCR was then performed to confirm the insertion of the PCR product into the pTZ57R vector. Colony PCR after the white/blue screening was conducted using the universal vector primers, M13-wide-F and M13-40-R (Table 2.1). Thirty PCR cycles were carried out with denaturation at 94°C for 30 sec, annealing at 52°C for 30 sec, and extension at 72°C for 1 min. The reaction had an initial denaturation step of 95°C for 1 min and a final elongation step of 72°C for 7 min. Positive clones were picked to perform sequencing reactions to identify colonies without PCR mutations.

Sequencing was performed using the BigDye cycle sequencing kit (V3.0, Applied Biosystems, Foster City, CA) and the samples were subjected to automated analysis with an ABI Prism model 3100 sequencer (Applied Biosystems). Sequence analysis was done using the ABI Prism SeqScape Software Version 2.0 (Applied Biosystems). The final sequence comparison was performed using ClustalW (http://www.ebi.ac.uk/clustalw).

Restriction digestion and ligation reaction

The intermediate vector, pRTL2, and the pTZ57R vector with the *EUCCR* fragment were digested with *Eco*RI and *Bam*HI (Roche) in restriction buffer B (Roche) at 37°C. Ligation of the excised *EUCCR* fragment into the pRTL2 vector was performed with a 3:1 insert-to-vector molar ratio at 22°C overnight. Similarly, pRTL2 and the pTZ57R vector with the *EgCesA1* fragment were digested with *Eco*RI (Roche) and *Xma*I (NEB)

at 37°C. Ligation of the digested *EgCesA* fragment into the pRTL2 vector was performed with a 3:1 insert-to-vector molar ratio at 22°C overnight in the water bath. Both constructs were transformed into competent *E. coli* DH5α cells. Colony screening was performed on LB plates containing 100 µg/ml ampicillin. Colony PCR was then used to amplify the fragment with predicted size. The expression vector was subsequently re-digested with the corresponding enzymes described above to confirm the size of the insert.

2.2.2.2 Construction of the binary vectors

Restriction digestion and ligation reaction

The binary vector, pCAMBIA 3301 (www.cambia.org), and the expression vector containing *EUCCR* cDNA were digested with *Pst*I (Fermentas) in buffer O (Fermentas) at 37°C. The *EUCCR* expression cassette was ligated into pCAMBIA 3301 with a 3:1 insert-to-vector molar ratio at 22°C. The expression vector which contained *EgCesA1* was digested with *Hind*III (Roche) in buffer B (Roche) at 37°C. The *EgCesA1* expression cassette was ligated into pCAMBIA 3301 as described above. Both constructs were transformed into competent *E. coli* DH5α cells. Colony screening was performed on LB plates containing 50 µg/ml kanamycin.

Colony PCR and nucleotide confirmation

Colony PCR was used to confirm the insertion of the expression cassettes into pCAMBIA 3301. The binary vector was re-digested with the corresponding enzymes described above to confirm the insert size. The expression cassettes, including the junction points in the binary vector, were sequenced to confirm the integrity of the inserts.

2.2.3 Production of transconjugant *Agrobacterium tumefaciens*

The *Agrobacterium tumefaciens* strains LBA 4404 and C58C1 were used in this study. The transformation of *Agrobacterium* requires the production of competent *Agrobacterium* cells and 'freeze-thaw' transformation (An *et al.*, 1988) with binary vectors. The LBA 4404 strain was transformed with the binary vector containing *EUCCR* and with the empty pCAMBIA 3301, while C58C1 was transformed with the binary vector containing *EgCesA1* and with the empty pCAMBIA 3301. The empty pCAMBIA 3301 vector containing a β-glucuronidase (GUS) gene (http://www.cambia.org) was used as a transformation control.

2.2.3.1 Preparation of competent *Agrobacterium* cells

Small-volume cultures were prepared by inoculating 5 ml YEP medium containing 50 µg/ml rifampicin. To prepare C58C1 small-volume cultures, an additional antibiotic, gentamycin, was added at 100 µg/ml. The culture was grown at 28°C for 2 days until the mid-log phase. Two milliliters of the small-volume culture were used to inoculate a larger volume of 50 ml YEP medium containing the same antibiotics at the same concentration as described above. The culture was incubated at 28°C with shaking at 200 rpm for two to three days. The sample was then centrifuged at 3000 ×*g* for 15 min and the resulting bacterial pellets resuspended in 1 ml ice-cold 10 mM $CaCl_2$. The resuspension solution was aliquoted into 100 µl volumes and frozen in liquid nitrogen.

2.2.3.2 Transformation of competent *Agrobacterium* cells

One microgram of each binary vector was pipetted onto 100 µl frozen competent *Agrobacterium* cells. The DNA-bacteria mixture was incubated in a 37°C water bath for 5 min, followed by the addition of 1 ml YEP medium containing 50 µg/ml rifampicin

and 20 µg/ml kanamycin. To transform C58C1 cells, an additional antibiotic, gentamycin, was included at 100 µg/ml to the YEP medium. The culture was incubated for 4 hours at 28°C with shaking at 200 rpm. The culture was then centrifuged for 2 min at 12,000 ×*g* and the resulting pellets resuspended in 100 µl YEP medium. YEP agar plates containing appropriate antibiotics were subsequently incubated at 28°C for two to three days.

2.2.4 Transformation of *Arabidopsis thaliana* plants

2.2.4.1 Preparation of transformation culture

Single *Agrobacterium* colonies containing the binary vectors confirmed by PCR were picked. Colonies were inoculated into 15 ml of YEP medium containing the appropriate antibiotics. The resulting cultures were incubated at 28°C overnight with shaking at 200 rpm. These precultures were kept at 4°C for future use. The transformation culture was grown with a 100:1 YEP medium-to-preculture volume ratio at 28°C for 24 hours with shaking at 250 rpm. Appropriate antibiotics (see above) were added into the final medium. The bacterial cells were used to transform *Arabidopsis* plants as described below.

2.2.4.2 Floral dip transformation

Primary inflorescence stems of *Arabidopsis* plants were cut to induce more inflorescence stems. The plants were ready for transformation when most secondary stems were about 5 cm long, the plants had many immature flower clusters and not many fertilized siliques were visible. Jiffy peat pellets in which plants were growing were saturated with the tap water one day before transformation. YEP medium containing *Agrobacterium* cells was then centrifuged at 5000 ×*g* for 20 min. The cells

were resuspended to OD_{600}=0.90 (for the traditional floral dip method) and OD_{600}=2.0 (for the improved floral dip method) in 5% (w/v) sucrose (Sigma). Before dipping, the surfactant, Silwett L-77 (LETHEL, Round Rock, U.S.A), was added to a concentration of 0.05% (v/v) and mixed well.

Wild-type *Arabidopsis* (Col-0) was transformed with the binary vectors containing *EUCCR*, or the empty pCAMBIA 3301, which were contained in *Agrobacterium* strain LBA 4404, using the traditional floral dip method. Above-ground parts of plants were dipped in *Agrobacterium* solution for about five seconds, with gentle agitation. The rosette was also dipped to expose the shorter axillary inflorescences to the *Agrobacterium* solution.

Irx4, *irx1* and wild-type *Arabidopsis* Ler-0 were transformed with the binary vector containing *EgCesA1*, or with empty pCAMBIA 3301, which were contained in *Agrobacterium* C58C1, using the improved floral dip method. A drop of 10-20 µl of the concentrated *Agrobacterium* solution was pipetted onto each flower bud.

The dipped plants were then laid on their side and left in the greenhouse at 24°C for about 16 hours. The next day, plants were stood upright in the greenhouse without watering. Plants were watered sparingly for the first two days. After one week, the plants were dipped for a second time using the method as described above. The plants were then grown for another two to three weeks until T1 seeds could be harvested.

2.2.5 Selective screening of transgenic *Arabidopsis*

T1 transgenic seedlings were grown and selected on seedling soil by spraying the plants with 230 µM Basta (glufosinate ammonium). The surviving plants were allowed to self-pollinate and T2 transgenic plants containing single T-DNA insertions were

identified using a segregation analysis. T2 lines revealing 3 (resistant to Basta):1 (sensitive to Basta) segregation ratios were selected for progression to T3 lines. T3 homozygous lines were selected by spraying the T3 plants grown on seedling soil with the same concentration of Basta as described above.

2.2.6 Leaf PCR analysis

Rosette leaves from 6-week-old plants were cut to perform leaf PCR. Leaf discs were collected using the narrow end of a 1000-µl pipette tip and the small leaf discs were then put into PCR-tubes with reaction mixture. The following PCR mixture was used to confirm the presence of the β-glucuronidase (GUS) gene in T3 control lines containing empty pCAMBIA 3301: primer GUS-SPC (F) (1 µM), primer GUS-SPC (R) (1 µM) (Table 2.1), 1× PCR buffer, 0.20 mM dNTPs, 1.5 mM $MgCl_2$, and, 0.2 U of ExSel polymerase in 20 µl. PCR was performed for thirty-five cycles with denaturation at 94°C for 30 sec, annealing at 51°C for 30 sec, and extension at 72°C for 30 sec. The reaction had an initial denaturation step of 94°C for 3 min and a final elongation step of 72°C for 10 min.

The following mixture was used to confirm the presence of *EUCCR* in T3 overexpression lines: primer 1428 (1 µM), primer EgunCCR1-116F (1 µM) (Table 2.1), 1× PCR buffer, 0.2 mM dNTPs, and 0.2U of Supertherm polymerase (Hoffman-La-Roche) in 20 µl. PCR was performed for thirty-five cycles with denaturation at 94°C for 30 sec, annealing at 51°C for 30 sec, and extension at 72°C for 1.5 min. The reaction had an initial denaturation step of 94°C for 3 min and a final elongation step of 72°C for 10 min.

2.2.7 Histochemical GUS assay

Rosette leaves from 6-week-old plants were cut for *E. coli* β-glucuronidase (GUS) assays (Jefferson, 1989). Leaf tissues were immersed in the GUS staining solution containing 1.92 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc, Sigma) in 200 mM sodium phosphate buffer (pH 7.0), containing 20% (v/v) methanol, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 10 mM Na_2EDTA and 0.1% (v/v) Triton X-100. Samples were vacuum-infiltrated briefly and incubated for 1 hour to overnight at 37°C. The resulting plant tissues were subsequently bleached several times by washing with 70% (v/v) ethanol in H_2O.

2.2.8 RT-PCR analysis

Fresh mature primary inflorescence stems of *Arabidopsis thaliana* were collected and frozen in liquid N_2 and stored at −80°C until used. Total RNA was extracted using the Concert Plant RNA Reagent (Invitrogen) according to the manufacturer's instructions. Genomic DNA in RNA samples was degraded using 1U DNaseI (Roche) per microgram DNA with 1 µl of RNasin (Promega) and 5× reverse transcription buffer (Roche) in a total volume of 50 µl for 30 min at 37°C. The resulting RNA samples were purified using the RNeasy kit (Qiagen). cDNA was then synthesized using the ImProm-II reverse transcription system (Promega). Primer anchor-P, anchoring on poly(A) tails, was used to synthesize first-strand cDNA. Primers ACT-F and ACT-R, specific to the *Arabidopsis Actin2* gene, were used to amplify the *Actin2* control gene during RT-PCR analysis (Fiers *et al.*, 2004). Primers EUCCR-A and EUCCR-B were used for analysis of the expression pattern of *EUCCR*. Primers AtCCR1-A and AtCCR1-B were used to amplify a fragment of the endogenous *Arabidopsis CCR1* gene

(*AtCCR1*). The resulting RT-PCR products were sequenced as described above.

The PCR reaction mixture included 0.16 units of Expand Taq polymerase, 1× reaction buffer, 1.0 µl of diluted (1/10) cDNA as template, 0.25 mM dNTPs and 0.4 µM of each primer in a 20 µl reaction volume. The thermal cycler was programmed to include one denaturing cycle at 95°C for 1 min. For EUCCR-A/EUCCR-B and AtCCR1-A/ AtCCR1-B primer pairs, it was followed by 20 cycles of denaturation at 94°C for 30 sec and annealing for 30 sec at 55°C and and 62°C, respectively. For ACT-F/ACT-R primer pair, it was followed by 25 cycles of denaturation at 94°C for 30 sec and annealing for 30 sec at 56°C. All the primers were extended at 72°C for 30 sec. A final step at 72°C for 7 min was included for the elongation. The PCR products were sequenced.

The same *Actin2*-specific primer pair (which spans an intron) was also used to detect genomic DNA contamination in total RNA samples. Leaf PCR was performed to test if this primer pair is able to amplify the fragment of the *Actin2* genomic copy. Leaf disc samples were prepared using the same method as described as above. The following PCR reaction mixture was used for leaf PCR: primer ACT-F (1 µM), primer ACT-R (1 µM), 1× PCR buffer, 0.20 mM dNTPs, and 0.2U of ExSel polymerase in 20 µl. PCR was performed for thirty-five cycles with denaturation at 94°C for 30 sec, annealing at 51°C for 30 sec, and extension at 72°C for 1 min. The reaction had an initial denaturation step of 94°C for 5 min and a final elongation step of 72°C for 10 min.

2.2.9 Microscopic analysis of cell wall phenotypes

Hypocotyls of sixteen-week-old *Arabidopsis* and the basal parts of mature inflorescence stems were used for microscopic analysis. To induce secondary xylem in hypocotyls, *Arabidopsis* plants were grown in a growth chamber at 22°C and 8 hours photoperiod

(100 µmol/m^2/s) and inflorescence stems were cut repeatedly. The tissues were sectioned to 16 µM thickness using a Leica cryotome (model CM1100) followed by phloroglucinol-HCl staining (Speer, 1987). Each sectioned sample was immersed in a 1% (w/v) solution of phloroglucinol (Sigma) in 92% (v/v) ethanol and then in 25% (v/v) hydrochloric acid. The samples were subsequently mounted in a drop of mounting solution after reddening of the section occurred. A cover slide was placed over the sample and the cell wall phenotype was detected using a light microscope (Axon) at a magnification of 20× magnification.

Microscopic analysis was also performed at four developmental stages, namely 3, 8, 12 and 16 weeks of Col-0 hypocotyls to observe the induction of secondary xylem in *Arabidopsis*. The same methods as described above were used to grow plants, induce secondary xylem and stain the cell walls.

2.3 Results

2.3.1 Vector construction

2.3.1.1 PCR amplification and cloning of *EUCCR* and *EgCesA1*

The *EUCCR* full-length cDNA of 1.1 kb (Figure 2.1) and the full-length *EgCesA1* cDNA of 2.97 kb (Figure 2.2) were amplified. Ten to twelve white colonies were screened by colony PCR analysis to confirm the insertion of the *EUCCR* cDNA fragment (Figure 2.3 A) and the *EgCesA1* cDNA fragment (Figure 2.3 B) in pTZ57R vectors using universal primers, M13-wide-F and M13-40-R primer (Table 2.1). The fragments of approximately 1.25 kb and 3.1 kb were amplified from positive clones containing *EUCCR* (Figure 2.3 A) and *EgCesA1* (Figure 2.3 B), respectively. Five

positive clones were selected to sequence the *EUCCR* cDNA and four positive clones were selected to sequence the *EgCesA1* cDNA. For each gene, one full-length clone without PCR induced errors was selected for restriction digestion. The two resulting vectors were designated pTZCCR and pTZCesA1.

2.3.1.2 Construction of the expression cassettes

The 1.1 kb *EUCCR* fragment (Figure 2.4) digested from pTZCCR with *Eco*RI and *Bam*HI was then ligated into the 3.6 kb digested pRTL2 vector (Figure 2.4). As shown in Figure 2.5 A, colony PCR confirmed the success of the ligation reaction by the amplification of a 1.45 kb fragment using two pRTL2-specific primers, pEuRTL-F and pEuRTL-R (Table 2.1). This clone was designated pRTLCCR. Figure 2.6 showed the re-digestion result of pRTLCCR with *Eco*RI and *Bam*HI and bands of predicted size. No mutations were detected according to sequencing results of the expression cassette

The 2.97 kb *EgCesA1* fragment digested from pTZCesA1 with *Eco*RI and *Xma*I was then ligated into the 3.6 kb digested pRTL2 vector. As shown in Figure 2.5 B, colony PCR confirmed the success of the ligation reaction by the amplification of the predicted 4.2 kb fragment using M13-40-R and M13-wide-F primers (Table 2.1). One positive clone was designated pRTLCesA1 and no mutations were detected in the expression cassette.

2.3.1.3 Construction of the binary vectors

After the digestion of pRTLCCR and pCAMBIA 3301 with *Pst*I, a 2.1 kb fragment (Figure 2.1; Figure 2.7 A), which contains the expression cassette, was then ligated into the 11.3 kb linearized pCAMBIA 3301 vector (Figure 2.7 B). As shown in Figure 2.8 A, colony PCR with vector primers, M13-40-R and M13-wide-F confirmed the success of

the ligation reaction. One positive clone designated pCAMCCR was used for plant transformation. The re-digestion of pCAMCCR with *Pst*I confirmed the insert size (Figure 2.9).

No mutations were detected in the pCAMCCR. The final plant transformation vector pCAMCCR, which contains *EUCCR* cDNA, was 13.4 kb (Figure 2.1). Sequencing results also confirmed the direction of the expression cassette (Figure 2.1).

A 4 kb fragment (Figure 2.2), which contained the *EgCesA1* expression cassette, was digested from pRTLCesA1 with *Hind*III and ligated into the 11.3 kb linearized pCAMBIA 3301. As shown in Figure 2.8 B, colony PCR with gene-specific primers, EgCesA-F and EgCesA-R confirmed the success of the ligation reaction. One positive clone was designated pCAMCesA1 and was used for plant transformation.

No mutations were detected in pCAMCesA1. The final plant transformation vector pCAMCesA1, which contained the full-length *EgCesA1* cDNA, was 15.3 kb in size (Figure 2.2). Sequencing results also confirmed the direction of the expression cassette (Figure 2.2).

2.3.2 Transformation of *Arabidopsis* plants

In this study, only the transformation of Col-0 plants with pCAMCCR and pCAMBIA 3301 (empty vector), and the transformation of *irx1* with pCAMCesA1 succeeded. As shown in Table 2.2, using traditional floral dip method the transformation efficiency for Col-0 was about 0.5%, but for *irx1* transformed with pCAMCesA1 it was less than 0.01% and only three transgenic lines were obtained using improved floral dip method. Here the transformation efficiency means the number of Basta-resistant plants (T1)/ the number of T1 seeds. The study of the determination of cell wall phenotypes of *irx1*

plants transformed with pCAMCesA1 is still in progress. Therefore, only results of the analysis of transgenic plants which contained overexpressed *EUCCR* will be discussed below. These results include the selection of transgenic plants, leaf PCR analysis, histochemical GUS assay, RT-PCR analysis, and microscopic analysis of cell wall phenotype. The rest of the results are provide and discussed in Chapter three.

2.3.3 Selection of transgenic plants

The chi-square test is a statistical method which is widely used to compare an observed frequency distribution or table with the expected frequency distribution or table, assuming the null hypothesis is true. We used this test to analyze the segregation of transgenic T2 lines which contained single T-DNA insertions. About four and twenty T1 lines were advanced to T2 lines which contained empty vector and *EUCCR*, respectively. After spraying with the herbicide, Basta, four T2 lines, 3301A2 (containing empty vector), CCR-A2, CCR-B2 and CCR-C2 were selected based on the segregation ratio (Table 2.3). These T2 lines segregated with a 3 (resistant):1 (sensitive) ratio, which was supported by the statistical test (Table 2.3). Line 3301A2 was a T2 line that contained empty pCAMBIA 3301 in the Col-0 background, while CCR-A2, CCR-B2 and CCR-C2 were T2 lines containing *EUCCR* cDNA in the Col-0 background.

Four T3 homozygous lines, 3301A3, CCR-A3, CCR-B3 and CCR-C3 were subsequently derived from 3301A2, CCR-A2, CCR-B2 and CCR-C2, respectively. The surviving lines that did not segregate after spraying with Basta, were considered as T3 homozygous lines.

2.3.4 Leaf PCR analysis

To confirm the insertion of T-DNA, leaf PCR was performed on the four T3 homozygous lines, 3301A3, CCR-A3, CCR-B3 and CCR-C3 which were described above. As shown in Figure 2.10, using GUS-specific primers, a 670 bp expected PCR product was amplified from the 3301A3 line which contained the empty vector and PCR products of 530 bp expected size were successfully amplified from CCR-A3, CCR-B3 and CCR-C3 lines using *EUCCR*-specific primers.

2.3.5 Histochemical GUS assay

GUS assays were performed to confirm the expression of the transgenes in T3 homozygous lines, 3301A3, CCR-A3, CCR-B3 and CCR-C3. As illustrated in Figure 2.11, GUS was successfully expressed in all of the transgenic plants, but no GUS activity was detected in wild-type plants. These results confirmed that T-DNA regions had been inserted into the *Arabidopsis* genome and that the transgenes had not been silenced.

2.3.6 RT-PCR analysis

Semi-quantitative RT-PCR was performed to confirm the transcription of *EUCCR* in transgenic *Arabidopsis* T2 and T3 lines. The results, which will be discussed in next chapter, revealed that the *EUCCR* gene was successfully expressed ectopically in wild-type Col-0. In this experiment, the *Arabidopsis Actin2* gene was used as internal control by amplifying a 240 bp fragment from stem cDNA. The same *Actin2*-specific primer pair was used to detect genomic DNA contamination by amplifying a 680 bp intron-spanning fragment from the genomic DNA of Col-0. As shown in Figure 2.12 A,

a genomic DNA fragment of *Actin2* was amplified using leaf PCR. In Figure 2.12 B and C, which were the results of RT-PCR for T2 and T3 plants, respectively, no obvious contaminating bands were revealed after the amplification from the cDNA samples. This meant that the *Actin2*-specific primer pair could be used to detect genomic DNA contamination in cDNA samples and that the cDNA samples used for RT-PCR, were DNA free.

2.3.7 Microscopic analysis of cell wall phenotype

Phloroglucinol-HCl was used to analyze the secondary xylem development of hypocotyls at four development stages, namely 3, 8, 12 and 16 weeks. The results (Figure 2.13) revealed that secondary xylem of hypocotyls of Col-0 increases along with the age of the plants. The older the plant is, the more the secondary xylem. Plants at 16 weeks contained large amounts of secondary xylem in the hypocotyls (Figure 2.13).

Lignin deposition was visualized using phloroglucinol-HCl to evaluate the increase of lignin content in the xylem cell walls of three T3 homozygous lines, CCR-A3, CCR-B3 and CCR-C3 which were transformed with the *EUCCR* gene. Staining in the mature ten-week-old stems and sixteen-week-old hypocotyls of these lines containing secondary xylem did not reveal a remarkable change, compared with the control lines (Figure 2.14).

2.4 Discussion

This project was set up to establish a functional genetic model system for wood formation genes that were isolated in our gene discovery work in *Eucalyptus*. In order to

establish and validate this model, we looked for *Arabidopsis thaliana* mutants of vascular development, for which we could obtain the corresponding *Eucalyptus* genes. We also required mutant plants that have phenotypes that are easy to evaluate under the microscope, so that complementation could be easily confirmed. The *Arabidopsis irregular xylem* (*irx*) mutants fulfilled these requirements: (1) we had full-length homologous *Eucalyptus* genes of these *irx* mutants; (2) an easy distinct phenotypic assay to evaluate and (3) as an added advantage, the mutant phenotype and complementation phenotype could be confirmed at the biochemical level. The *irx1*/*irx4* mutants furthermore allowed us to study lignin and cellulose biosynthesis, particularly carbon flux reallocation where any of these pathways were disturbed. Carbon flux reallocation is further discussed in Chapter Three.

In this study, pCAMBIA 3301 was used to construct binary vector containing *Eucalyptus* wood formation genes using traditional subcloning methods. However, sometimes it was difficult to get the ligation products due to this big vector although all the ligation reaction worked finally. It is especially difficult if blunt ends were used in the ligation reaction. The subcloning methods with much higher efficiency, such as gateway system (Karimi *et al.*, 2002; Curtis and Grossniklaus 2003), are now being used in this lab.

The genetic background of *irx1* and *irx4* is Ler-0. This might be the reason why the transformation efficiency is so low for *irx1* with pCAMCesA1, and why the rest of the transformations failed even after using improved floral dip method. The efficiency to transform Ler-0 was previously found to be extremely low using the floral dip method (Clough and Bent, 1998). Therefore, the efficiency to transform *Arabidopsis* Ler-0 and mutants within the Ler-0 background needs to be increased by testing other motheds,

such as the bombardment of plant material with DNA-coated microprojectiles (Klein *et al.*, 1987).

The *Eucalyptus CCR* gene (*EUCCR*) was successfully transformed into wild-type *Arabidopsis thaliana* using the traditional floral dip method and secondary xylem was induced in the hypocotyls of these transgenic plants. Combined with the results described in Chapter Three, *EUCCR* altered the cell wall chemistry in inflorescence stems of transgenic *Arabidopsis* in a predicted fashion. These results revealed that this widely used model plant can also be used to model the function of wood formation genes isolated in *Eucalyptus* trees. In the future, the *irx1* plants transformed with *EgCesA1* need to be analyzed using chemical and microscope methods.

No apparent differences were found between the control and overexpression plants using microscopic analysis (Figure 2.14). This might be due to the limited increase of lignin content (12%) as determined using TGA methods, which will be discussed in Chapter three. The threshold where the differences turn to be apparent could be found using relatively simple methods, for instance, by comparing the plants with different percentages of lignins using phloroglucinol-HCl staining.

More work needs to be done to validate the use of this model system for wood formation in *Eucalyptus* tree. We are going to backcross *EUCCR* from overexpressed T3 lines containing *EUCCR* into the *irx4* mutant genetic background. Basta selection, and analysis of the complementation such as microscopic and chemical analysis, will be performed. The *irx1* plants transformed with *EgCesA1* will be advanced to T3 lines and microscopic and chemical analysis will be performed to analyze the complementation phenotypes.

FIGURES

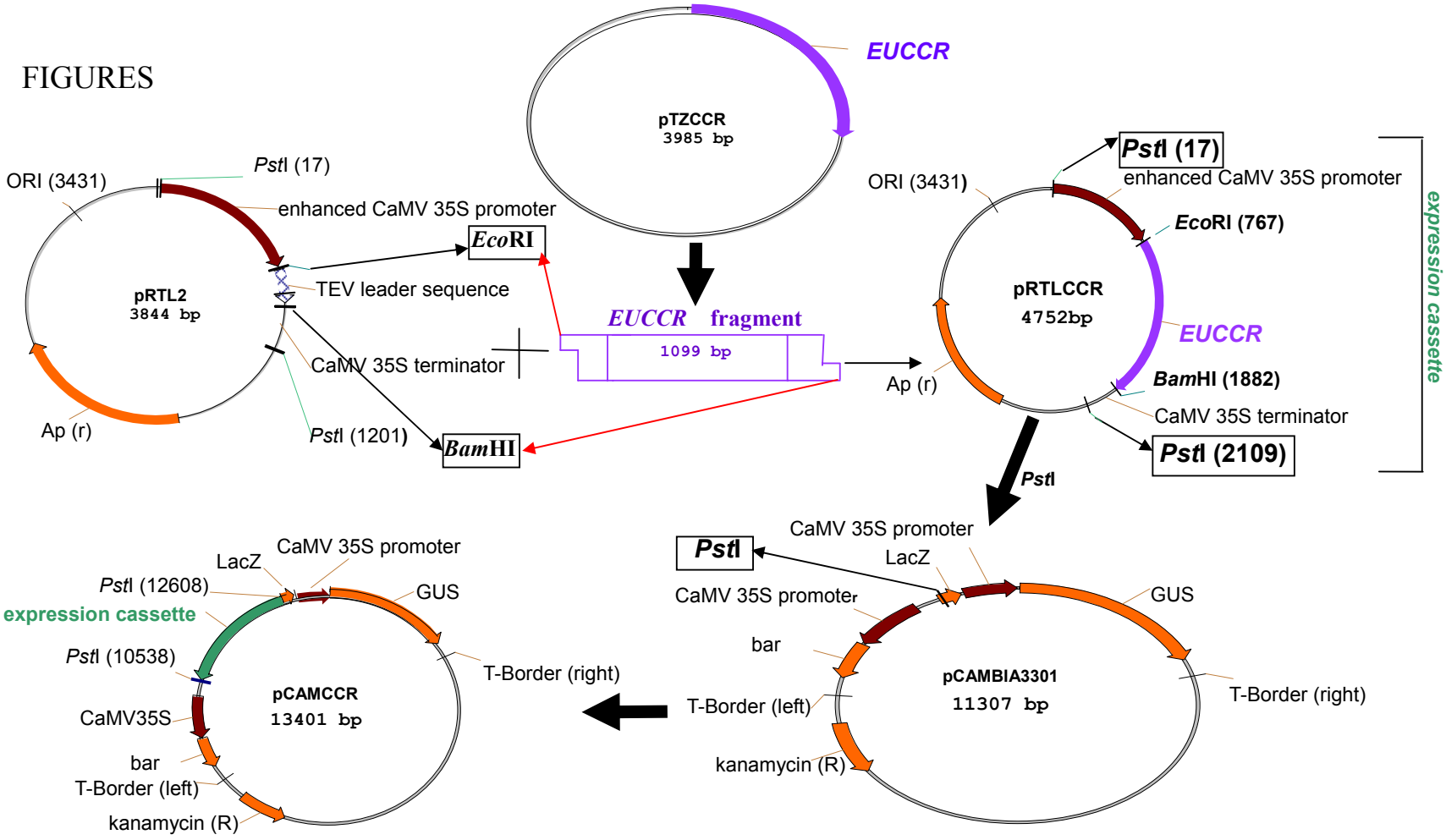


Figure 2.1 Schematic representation of the three-step construction of the pCAMCCR binary vector: 1: *EUCCR* fragment amplification; 2; expression vector construction and 3: binary vector construction.

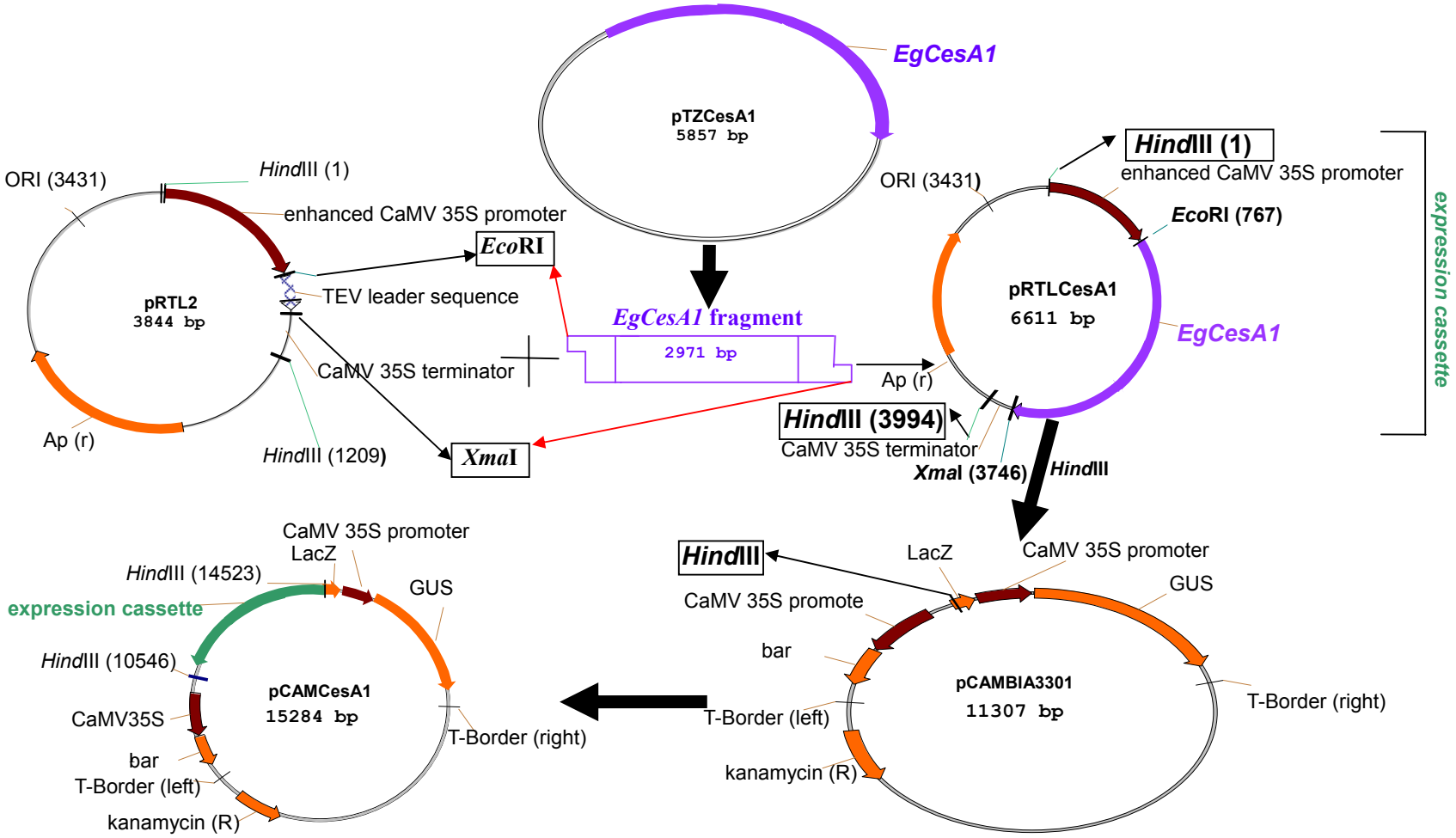


Figure 2.2 Schematic representation of the three-step construction of the pCAMCesA1 binary vector: 1: ***EgCesA1*** fragment amplification; 2; expression vector construction and 3: binary vector construction.

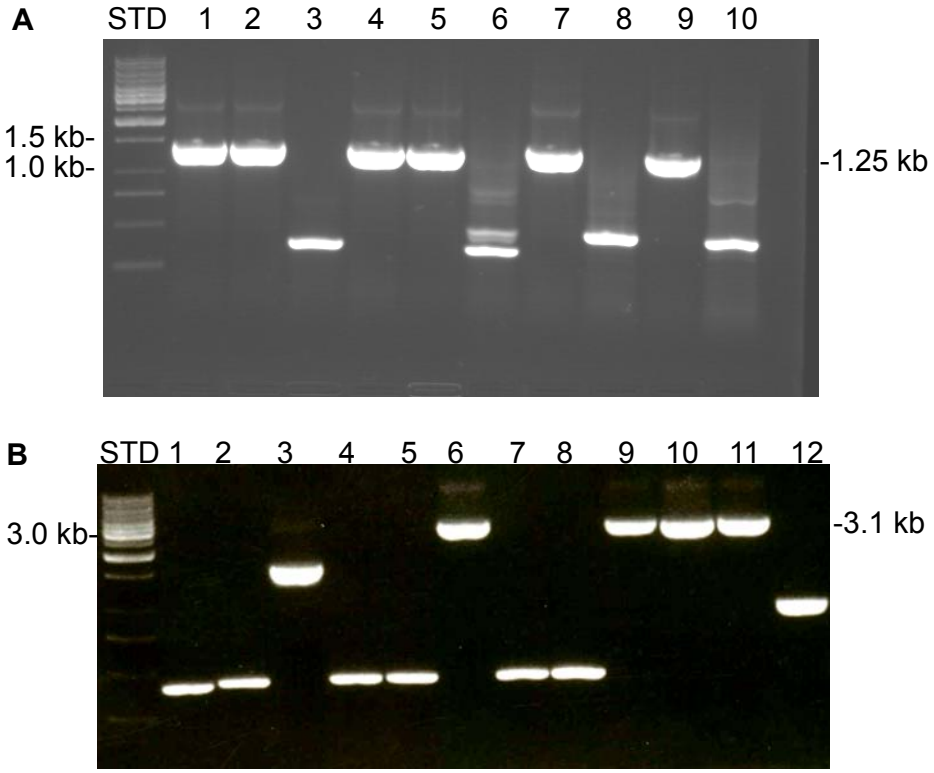


Figure 2.3 Colony PCR confirmation of A) *EUCCR* cDNA insertion and B) *EgCesA1* insertion into the pTZ57R vector. STD: 1 kb DNA size standard. The expected sizes of full-length positive clones are indicated on the right of each gel image. Numbered lanes contain colony PCR products of different clones.

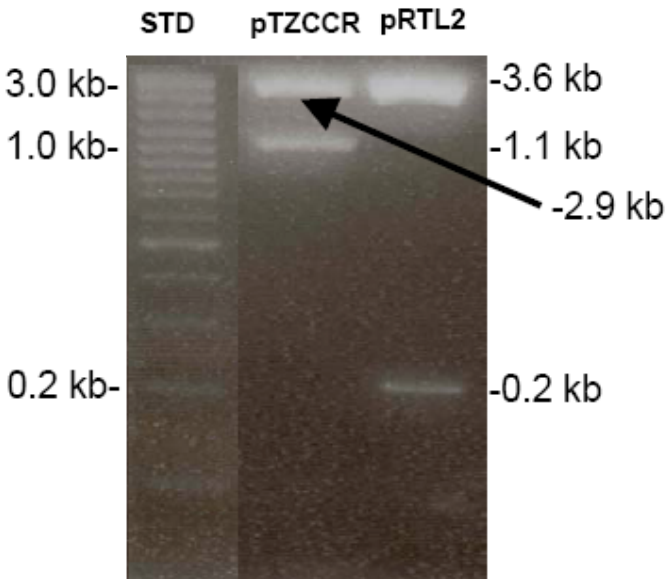


Figure 2.4 Restriction digestions of the pTZCCR vector and pRTL2 vector with *Eco*RI and *Bam*HI. STD: 100 bp size standard; pTZCCR: digested pTZCCR plasmid; pRTL2 digested pRTL2 plasmid. The expected sizes of restriction products are shown on the right of the gel image.

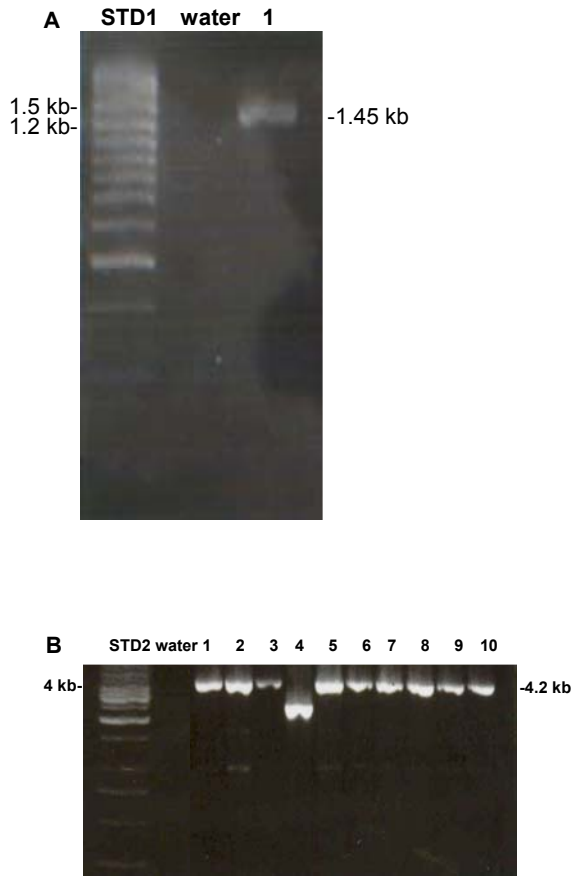


Figure 2.5 Colony PCR confirmation of the insertion of A) *EUCCR* cDNA and B) *EgCesA1* into the pRTL2. STD1: 100 bp size standard; STD2: 1 kp size standard; water: water as control. Numbered lanes contain colony PCR products derived from positive clones. Expected full-length sizes are indicated on the right of each gel image.

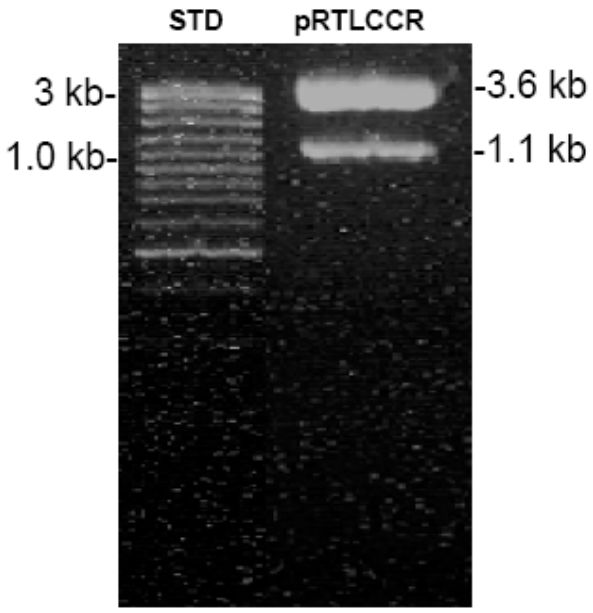


Figure 2.6 Restriction digestion analysis of pRTLCCR. STD: 100 bp size standard; pRTLCCR: pRTLCCR plasmid digested with *Eco*RI and *Bam*HI. Expected restriction fragment sizes are indicated on the right of the gel image.

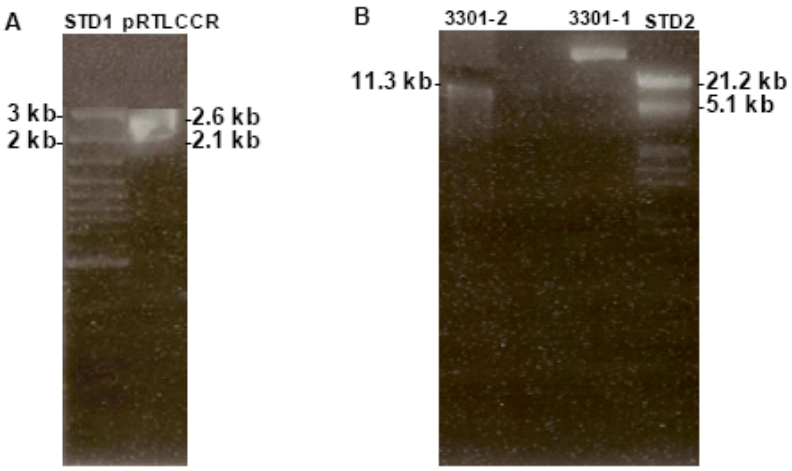


Figure 2.7 Restriction digestions of A) pRTLCCR and B) pCAMBIA 3301 vectors with *Pst*I. STD1: 100 bp size standard; STD2: lambda DNA /*Eco*RI+*Hind*III marker; pRTLCCR: digested pRTLCCR, the fragment at the size of 2.1 kb was excised and used for ligation; 3301-1: pCAMBIA 3301 plasmid DNA; 3301-2: digested pCAMBIA 3301.

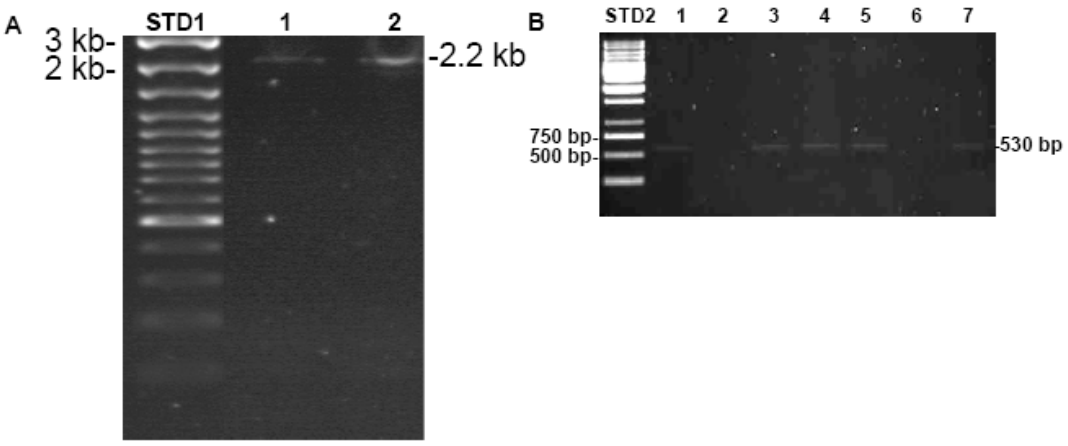


Figure 2.8 Colony PCR confirmation of the insertion of expression cassettes containing A) *Eucalyptus* CCR cDNA and B) *EgCesA1* cDNA in pCAMBIA 3301. Primer pairs, M13-40-R/M13-wide-F and EgCesA-F/EgCesA-R, were used to peform PCR for *Eucalyptus* CCR cDNA and *EgCesA1* cDNA, respectively. STD1: 100 bp size standard; STD2: 1 kb size standard. Numbered lanes contain colony PCR products derived from positive clones. Expected full-length sizes are indicated on the right of each gel image.

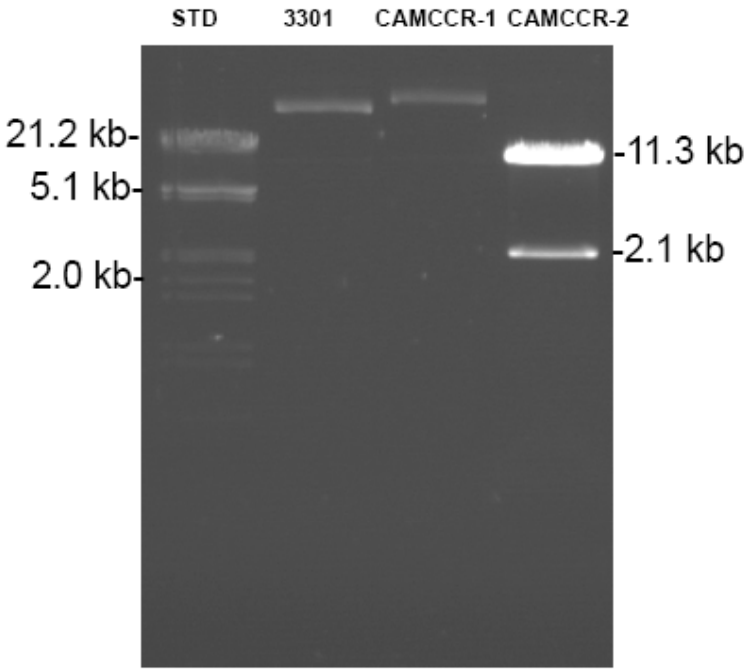


Figure 2.9 Re-digestion analysis of pCAMCCR. STD: lambda DNA/*Eco*RI+*Hind*III size standard; 3301: pCAMBIA 3301 plasmid; CAMCCR-1: pCAMCCR plasmid DNA; CAMCCR-2: pTRLCCR plasmid DNA digested with *Pst*I.

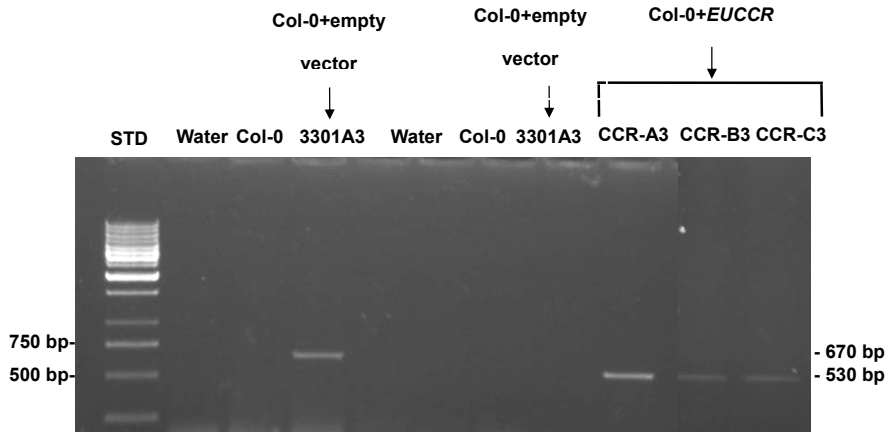


Figure 2.10 Leaf PCR analysis of transgenic plants. Leaf PCR for T3 homozygous control line (3301A3) containing empty pCAMBIA 3301 vector was performed using GUS-specific primers. *EUCCR*-specific primers were employed to amplify the fragments in three T3 overexpressing lines (CCR-A3, CCR-B3, and CCR-C3) containing *EUCCR*. *EUCCR*-specific primers were also used perform PCR for control line (3301A3) containing empty pCAMBIA 3301 vector. STD: 1 kb size standard; Water: water as negative control.

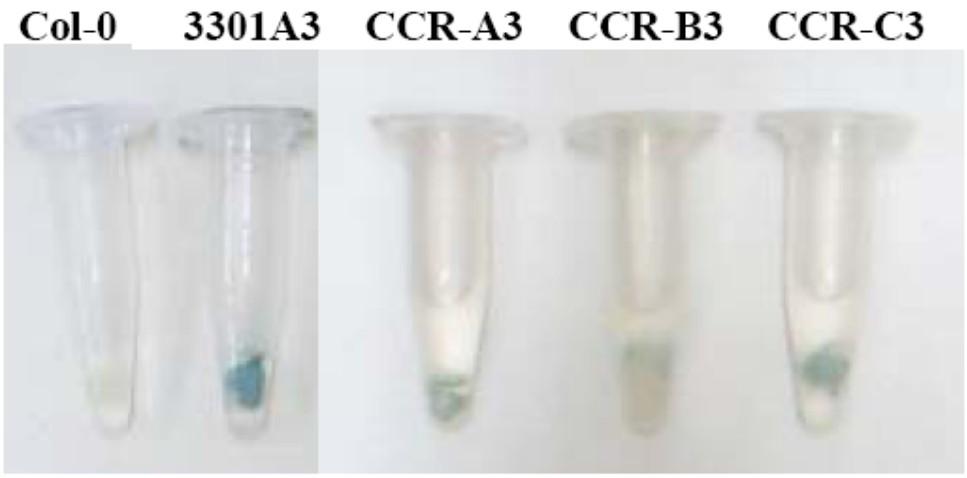


Figure 2.11 GUS assay for the transgenic plants. Col-0: wild-type *Arabidopsis*; 3301A3: Col-0 transformed with empty pCAMBIA 3301; CCR-A3, CCR-B3, CCR-C3: three T3 lines containing pCAMCCR.

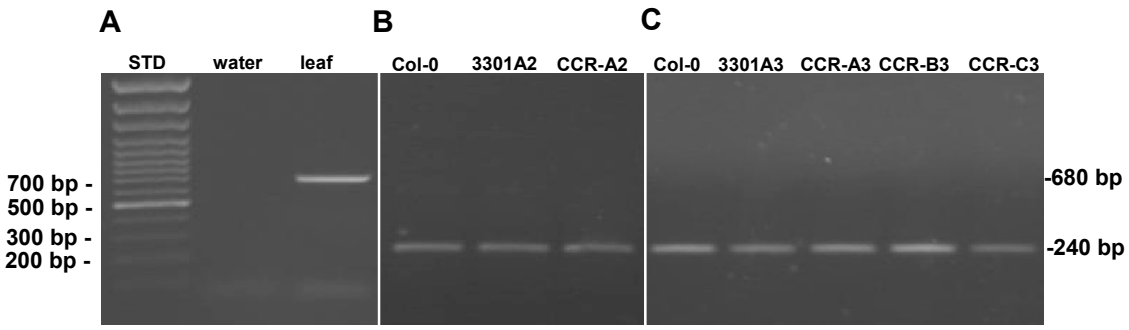


Figure 2.12 Genomic DNA contamination analysis. A) leaf PCR; B) RT-PCR for T2 lines and C) RT-PCR for T3 lines using *Arabidopsis Actin2*-specific primers. STD: 100 bp size standard; Col-0: wild-type *Arabidopsis*; 3301A2: Col-0 transformed with empty pCAMBIA 3301 (T2 line); CCR-A2: T2 line containing pCAMCCR; 3301A3: Col-0 transformed with empty pCAMBIA 3301 (T3 line); CCR-A3, CCR-B3, CCR-C3: three T3 lines containing pCAMCCR.

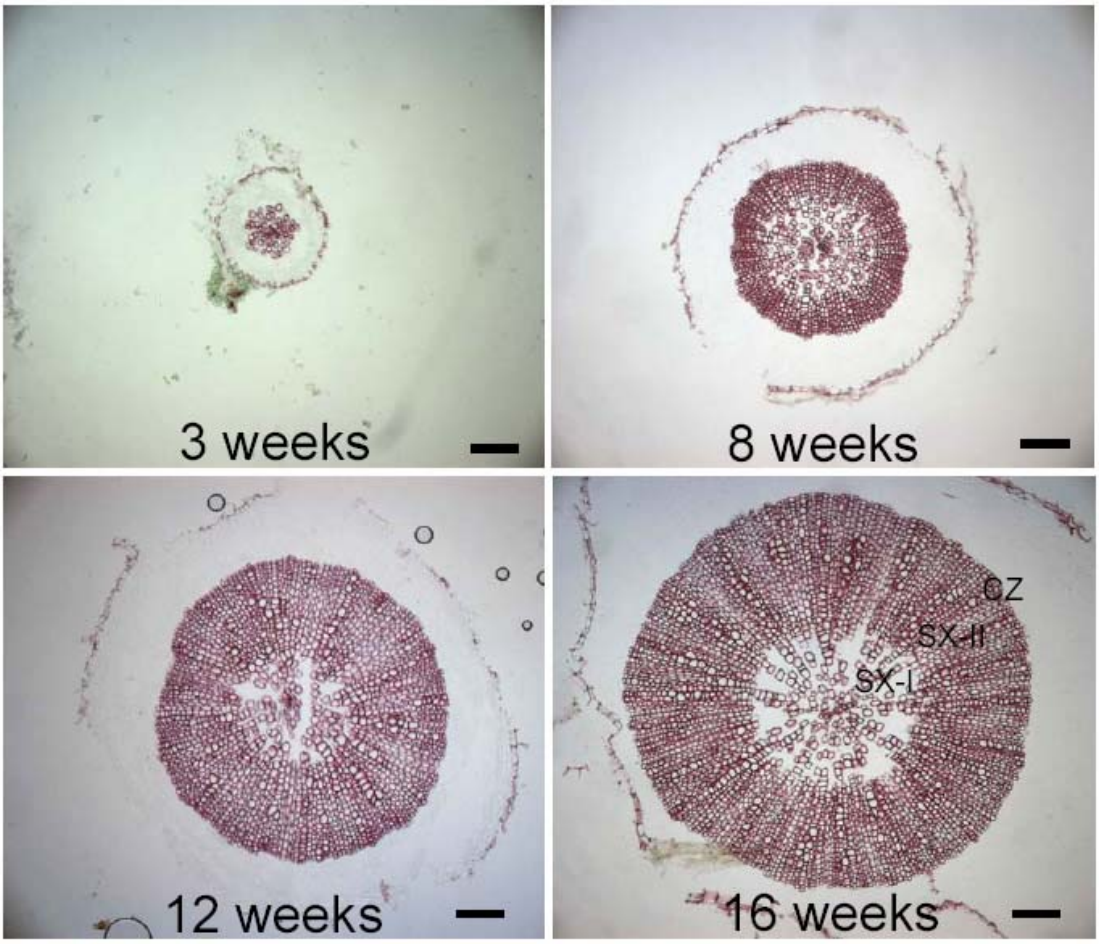


Figure 2.13 Phloroglucinol-HCl staining of lignified cell walls in hypocotyls of wild type *Arabidopsis,* Col-0, at four different developmental stages to reveal the developmental changes of induced secondary xylem. SX-I: secondary xylem at stage I; SX-II: secondary xylem at stage II; CZ: cambial zone. Scale bars equal 200 µm.

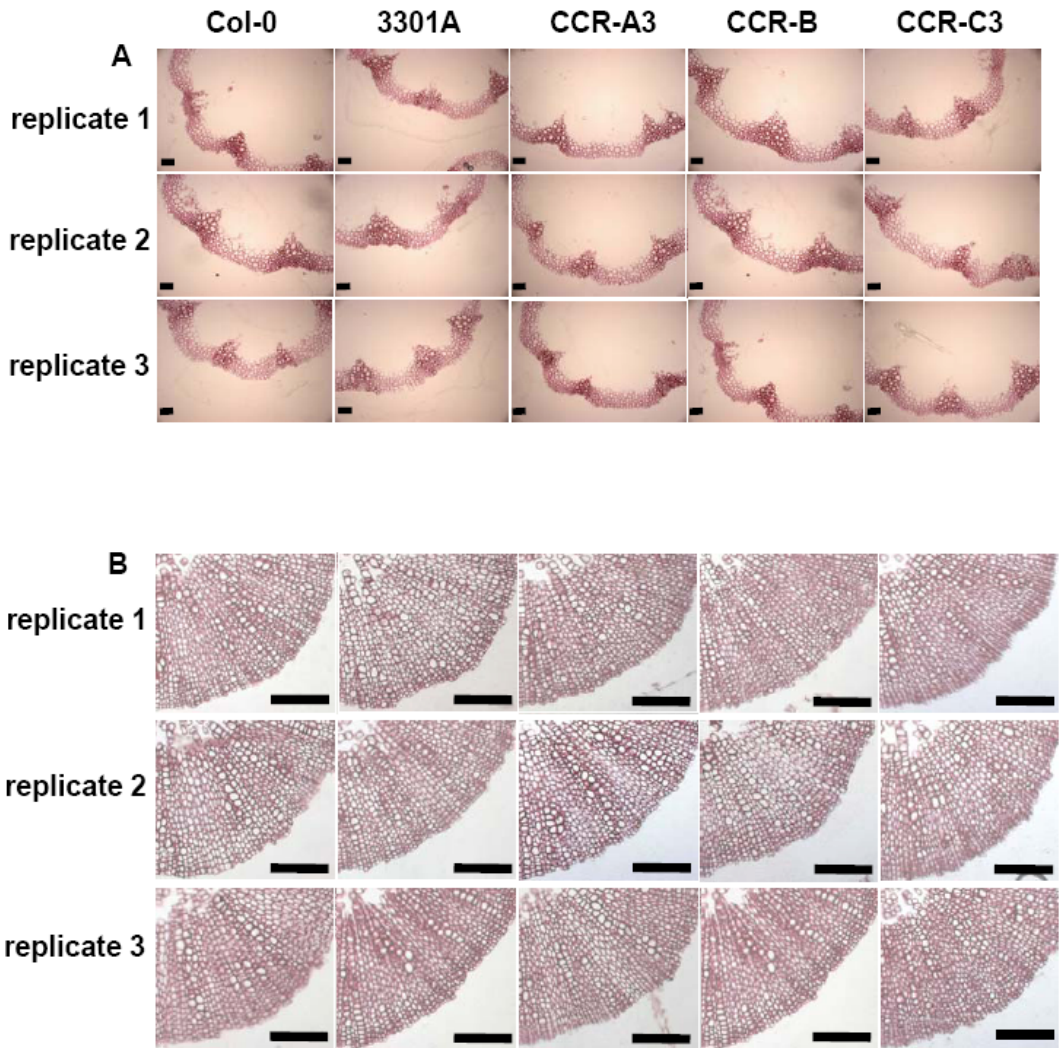


Figure 2.14 Lignin staining in control and transgenic plants. Cross-sections of ten-week-old inflorescence stems (A) and sixteen-week-old hypocotyls (B) of three mature replicate plants of each line were stained with phloroglucinol-HCl to reveal lignified tissues. Col-0: wild type *Arabidopsis*; 3301A3: Col-0 transformed with empty pCAMBIA 3301; CCR-A3, CCR-B3, CCR-C3: three lines containing pCAMCCR. Bars indicate a scale of 200 µm.

TABLES

Table 2.1 Oligonucleotide primers used in this chapter (restriction sites were underlined).

Primer name	Primer sequence (5'➔3')
CCRcDNA5'UTR-F	GCGAATTCAAGCGCAACCTCCA
CCRcDNA3'UTR-R	GCGGATCCTGCGCACCGTGATGGATCTA
5+7 *Eco*RI	CCGGAATTCGGCCAACATGATGGAATCCG
3' UTR *Sma*I	TCCCCCGGGCAACACTGGCGCAATTCAAC
GUS-SPC (F)	CATGTCGCGCAAGACTGTAA
GUS-SPC (R)	TCCGGTTCGTTGGCAATACT
1428	CATTGATCGTCGACTGAAGG
EgunCCR1-116F	CCTGGATTGTCAAGCTTCTC
anchor-P	GACCACGCGTATCGATGGCTCA*T_{17}**V
ACT-F	TCTCGTTGTCCTCCTCACTT
ACT-R	AATCCAGCCTTCACCATACC
EUCCR-A	CCTGGATTGTCAAGCTTCTC
EUCCR-B	CATTGATCGTCGACTGAAGG
AtCCR1-A	AAGCGCGTGGTCATCACCTC
AtCCR1-B	GGCGTTGATCGTCGGCTGTA
M13-40-R	GTTTTCCCAGTCACGAC
M13-wide-F	GGATAACAATTTCACACAGG
EgCesA-F	TACCTGGCTAACCGTAGCAC
EgCesA-F	TGCGCTATCCTTAGCATGAG
pEuRTL-F	CGTTCCAACCACGTCTTCAA
pEuRTL-R	CTGCAGGTCACTGGATT
EgunCCR-220F	AGCGAGAGGCTGACGCTGTA
EgunCCR-648F	TGCGAGCATCATCCACATCC
3301-spec-F	CAAGCTGCTCTAGCCAATAC
3301-spec-R	CTCAAGCTGCTCTAGCATTC

* T_{17}: TTTTTTTTTTTTTTTTT; **V: A, G or C.

Table 2.2 Transformation efficiency of the floral dip method using different transgene constructs.

Constructs	Lines	Transformation methods	Efficiency (%)
pCAMBIA 3301	Col-0	traditional	0.5
	irx1	improved	0
	irx4	improved	0
pCAMCCR	Col-0	traditional	0.5
	irx4	improved	0
pCAMCesA1	*irx1*	improved	0.007

Transformation efficiency: the number of Basta-resistant plants (T1)/ the number of T1 seeds;

Traditional: traditional floral dip method (Clough and Bent, 1998);

Improved: improved floral dip method (Briones *et al.*, 2004).

Table 2.3 Chi-square test for selected T2 lines containing single T-DNA insertions (3:1 segregation ratio).

Lines	3301A2		CCR-A2		CCR-B2		CCR-C2	
Sensitivity	R	S	R	S	R	S	R	S
Observed counts (plant numbers)	106	26	131	40	116	42	121	38
Expected counts (plant numbers)	99	33	128.25	42.75	118.5	39.5	119.25	39.75
P-value	0.159		0.627		0.646		0.749	

R: Basta-resistant; **S**: Basta-sensitive

References

An G, Ebert PR, Mitra A, Ha SB (1992) Binary vector. In: S.B. Gelvin, R. A. Schilperoort (Eds.). *Plant Molecular Biology Manual*. A3, 1-19, Kluwer Academic Publishers, Dordrecht

Briones V, Cabrera-Ponce JL, Herrera-Estrella L (2004) Improving transformation efficiency of *Arabidopsis thaliana* by modifying the floral dip method. *Plant Mol. Biol. Rep*. 22, 63-70

Chaffey NJ, Cholewa E, Regan S, Sundberg B (2002) Secondary xylem development in *Arabidopsis*: a model for wood formation. *Physiologia Plantarum* 114, 594-600

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J*. 16, 735-743

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol*. 133, 462–469

Fiers M, Hause G, Boutilier K, Casamitjana-Martinez E, Weijers D, Offringa R, Geest L, Campagne M, Liu CM (2004) Mis-expression of the *CLV3*/*ESR*-like gene *CLE19* in *Arabidopsis* leads to a consumption of root meristem. *Gene* 327, 37-49

Jefferson RA (1989) The GUS Reporter Gene System. *Nature* 342, 837-838

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

Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* 7, 193-195

Klein TM, Wolf ED, Wu R, Sanford JC (1987) High velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327, 70-73

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

Plant Physiol. 135, 1069-1083

Lacombe E, Hawkins S, Doorsselaere JV, Piquemal J (1997) Cinnamoyl coA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *Plant J*. 11, 429-441

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant* 15,473-496

Speer EO (1987) A method of retaining phloroglucinol, proof of lignin. *Stain Technology* 62, 279-280

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

Tzfira T, Zuker A, Altman A (1998) Forest-tree biotechnology: genetic transformation and its application to future forests. *TIBTECH* 16, 439-446

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

Chapter Three

Overexpression of the *Eucalyptus* cinnamoyl-CoA reductase (CCR) gene in *Arabidopsis* alters lignin content and composition of *Arabidopsis* stems

Honghai Zhou, David K. Berger, Shawn D. Mansfield, Alexander A. Myburg

Prepared as a manuscript for the journal *Journal of Plant Physiology*

3.1 Abstract

Cinnamoyl-CoA reductase (CCR; EC 1.2.1.44) catalyzes the first step in the phenylpropanoid pathway specifically dedicated to the monolignol biosynthetic branch. A *Eucalyptus* CCR gene, *EUCCR*, was previously isolated from *Eucalyptus gunnii*. We evaluated the effect of overexpression of *EUCCR* on the cell wall chemistry of transgenic *Arabidopsis* plants. The *Eucalyptus* CCR gene was placed under the control of the CaMV 35S promoter and transformed into Col-0 wild-type *Arabidopsis* plants. In T2 overexpressing plants with a single T-DNA insertion, total lignin was increased approximately 5% in inflorescence stems compared with those of control plants. HPLC analysis revealed that total carbohydrate content was decreased approximately 8% in transgenic stems, suggesting the reallocation of carbon from cellulose to lignin. Thioacidolysis analysis indicated that G-lignin subunit content was increased (16%), while S-unit content was decreased (21%) resulting in an overall reduction of the S/G ratio. These results are consistent with CCR down-regulation phenotypes in tobacco and support the use of *Arabidopsis* to model the role of *Eucalyptus* genes in xylem cell wall formation.

Keywords: *Arabidopsis***,** *EUCCR*, lignin, functional genetics, G monomer, S monomer

3.2 Introduction

Eucalyptus is a very important plantation hardwood genus and eucalypt wood fibre is widely used in the pulp and papermaking industry (Tumbull, 1991). However, our understanding of wood formation processes, such as lignin biosynthesis in *Eucalyptus* and other tree species, is far from complete. Lignin of angiosperms such as *Eucalyptus* mainly consist of a mixture of monolignols, monomethoxylated guaiacyl (G), bimethoxylated syringyl (S) and traces of nonmethoxyleated *p*-hydroxyphenyl (H), derived from coniferyl, sinapyl and *p*-coumaryl alcohol, respectively (Anterola and Lewis, 2002). Lignin has to be removed from cellulose pulp, the crude material. Chiang and Funaoka (1990) reported that high S lignin levels facilitate industrial kraft delignification during the pulping process. Lignin content and methoxyl content are also negatively associated with forage digestibility for ruminant animals (Sewalt *et al.*, 1996). Lignin is therefore a target for genetic engineering in trees (Hu *et al.*, 1999; Li *et al.*, 2003).

Monolignol biosynthesis is a multi-step process involving several enzymes and metabolites (Figure 3.1). This process starts with the common phenylpropanoid pathway. Monolignols are then produced via two reductive steps catalyzed by Cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD). CCR catalyzes the conversion of cinnamoyl CoA esters to their corresponding cinnamaldehydes at the first specific step dedicated to the monolignol biosynthetic branch. At the biochemical and genetic levels, the entire pathway of lignin biosynthesis has not been characterized in a single tissue or a single species (Boerjan *et al.*, 2003; Raes *et al.*, 2003). The currently accepted model of monolignol biosynthesis envisages a metabolic grid leading to G and S units. Dixon *et al.* (2001) suggested an alternative model in which metabolic

channeling allows for independent pathways to G and S lignin, as indicated by the results of genetic manipulation experiments and enzyme specificity studies, which are inconsistent with the classical model of monolignol biosynthesis. It is clear that, despite extensive research on aspects of lignin biosynthesis in various species, our knowledge of the regulation and nature of this process is still lacking.

The difficulty of studying the function of wood formation genes in *Eucalyptus* is mainly derived from the fact that it is not easy to obtain transgenic *Eucalyptus* trees. Transformation efficiency is very low in *Eucalyptus* trees and requires a substantial investment in tissue culture facilities. Even if successful transformation and regeneration is achieved, it can take another 24 months to obtain a reliable transgene phenotype in *Eucalyptus* trees (Tzfira *et al.*, 1998). Model systems, such as *Arabidopsis*, are useful for functional analysis of wood formation genes identified in *Eucalyptus*.

Besides the general advantages of *Arabidopsis* for model studies (Meyerowitz, 2001; Chen *et al.*, 2004), this herbaceous species has advantages that allow it to be used for functional genetic analysis of wood formation genes. Secondary xylem can be induced and studied in hypocotyls (Chaffey *et al.*, 2002) and inflorescence stems (Ko *et al.*, 2004) of *Arabidopsis* plants. Chaffey *et al.* (2002) reported that the anatomy of the induced secondary xylem cells in hypocotyls is comparable to that of young woody stems of poplar or *Eucalyptus* seedlings. Recently, a study by Ko *et al.* (2004) revealed that increased *Arabidopsis* body weight is able to induce secondary xylem in inflorescence stems. The ability to harvest several inflorescence stems from each *Arabidopsis* plant means that enough experimental material can be obtained for gene expression studies and chemical analysis. Numerous vascular developmental mutants have been identified in *Arabidopsis* (www.arabidopsis.org). This facilitates

complementation studies using homologous genes from woody plants such as *Eucalyptus* or poplar (Sibout *et al.*, 2002). Finally, the knock-out phenotypes of unknown genes discovered in woody systems can be evaluated in induced *Arabidopsis* plants if homologs of these genes are present in *Arabidopsis*.

In this study, we investigated the use of *Arabidopsis thaliana* as a model system to analyze the function of wood formation genes in *Eucalyptus*. To achieve this, we specifically overexpressed the *Eucalyptus CCR* (*EUCCR*) gene in wild-type *Arabidopsis thaliana* (Col-0) plants to evaluate the effects of ectopic expression of this gene on monolignol biosynthesis and carbon flux between different cell wall components in transgenic plants. Chemical analysis performed on the transgenic plants confirmed alteration of the cell wall chemistry and usefulness of *Arabidopsis* to model *Eucalyptus* wood formation genes.

3.3 Materials and Methods

3.3.1 Plant materials

Arabidopsis thaliana L. Heynh. ecotype Columbia (Col-0, Nottingham *Arabidopsis* Stock Centre) seeds were surface sterilized with 1.5% (w/v) sodium hypochlorite for 30 min and sown on Petri dishes containing 1% (w/v) MS medium (Murashige and Skoog, 1962) in 0.8% (w/v) gelled agar supplemented with 2% (w/v) sucrose. The seeds were then kept at 4°C for 3 days to break the dormancy, followed by incubation at 25°C under a 16/8 hours day/night cycle (150 µmol/m^2/s) for two weeks. The young seedlings were transferred to soil medium (Jiffy Products International AS) and grown in a growth

chamber at 22°C and 8 h photoperiod (100 µmol/m^2/s) to maximize vegetative growth.

3.3.2 Vector construction

Cloning was performed according to standard molecular cloning protocols (Sambrook *et al.*, 1989). PCR primers 5'GCGAATTCAAGCGCAACCTCCA3' and 5'GCGGATCCTGCGCACCGTGATGGATCTA3' matching the 5'UTR and 3'UTR of the *EUCCR* cDNA (Genbank accession number: X79566) and introducing *EcoR*I and *BamH*I restriction sites (underlined), respectively, were designed to amplify the 1.1 kb full-length *EUCCR* cDNA (Lacombe *et al.*, 1997). The PCR product was cloned into the pTZ57R vector (InsT/Aclone PCR Cloning Kit, Fermentas). Five positive clones were sequenced and BLAST (www.ncbi.nlm.nih.gov/BLAST) analysis was performed to confirm their integrity. A clone without PCR mutations was selected for restriction digestion with *Eco*RI/*Bam*HI. The restriction fragment was then ligated into pRTL2 (Restrepo *et al.*, 1990) containing a dual CaMV 35S promoter and CaMV 35S terminator. The resulting expression cassette was cloned as a *Pst*I fragment in the pCAMBIA 3301 binary vector (www.cambia.org). The expression cassette including junction points in the final binary vector, pCAMCCR, was sequenced to confirm the direction of cloning and integrity of the clone. pCAMCCR was transformed into *Agrobacterium* strain LBA 4404 (Larebeke *et al.*, 1974) using the chemically based direct transformation method (An *et al.*, 1992).

3.3.3 Transformation and selection of transgenic lines

Agrobacterium strains containing the binary vectors pCAMBIA 3301 or pCAMCCR were used to transform wild-type Col-0 plants grown on soil medium using the floral dip method (Clough and Bent, 1998). T1, T2 and T3 transgenic lines were screened for

herbicide resistance by spraying the young seedlings grown on seedling soil with 230 µM Basta (glufosinate ammonium).

3.3.4 RT-PCR analysis of *EUCCR* in transgenic *Arabidopsis thaliana* plants

Fresh mature primary inflorescence stems of control and transgenic plants were collected into liquid N_2 and stored at −80°C until used. Total RNA was extracted using the Concert Plant RNA Reagent (Invitrogen) according to manufacturer's instructions. Genomic DNA in RNA samples was degraded using 1U of DNaseI (RNase free, Roche) per microgram of DNA with 1 µl of RNasin (Promega) and 5x reverse transcription buffer (Roche) in a total volume of 50 µl for 30 min at 37°C. Total RNA was purified using the RNeasy kit (Qiagen). cDNA was then synthesized from total RNA using the ImProm-II reverse transcription system (Promega). An anchored poly(A) tail primer, 5'GACCACGCGTATCGATGGCTCATTTTTTTTTTTTTTTV3' (V=A, C or G), was used to synthesize the first-strand cDNA. Primers 5'TCTCGTTGTCCTCCTCACTT3' and 5'AATCCAGCCTTCACCATACC3', specific to the *Arabidopsis Actin2* gene, were used to amplify this control gene for relative quantification (Fiers *et al.*, 2004). Primers 5'CCTGGATTGTCAAGCTTCTC3' and 5'CATTGATCGTCGACTGAAGG3' were used to amplify *EUCCR* from inflorescence cDNA. A fragment of the endogenous *Arabidopsis CCR1* gene (*AtCCR1*) was amplified using primers 5'AAGCGCGTGGTCATCACCTC3' and 5' GGCGTTGATCGTCGGCTGTA3'.

3.3.5 Chemical analysis of *Arabidopsis thaliana* stems

A modified micro-Klason method (Huntley *et al.*, 2003), HPLC analysis and thioacidolysis were used for T2 lines and a thioglycolic acid assay (Campbell and Ellis,

1992) for T3 lines to analyze the changes in cell wall chemistry due to the ectopic expression of *EUCCR* in *Arabidopsis*. To compare the differences in chemical components between control lines (containing empty vectors) and transgenic lines (containing the *EUCCR* construct), a two-tailed *t* test was performed.

Freeze-dried *Arabidopsis* stems (200 mg) were ground to pass a 40-mesh screen using a Wiley mill and extracted in a Soxhlet apparatus with acetone for 6 hours. Lignin content of mature *Arabidopsis* inflorescence stems was determined according to a modified micro-Klason method (Huntley *et al.*, 2003). Briefly, 100 mg of acetone-extracted ground stem sample was digested with 72% (v/v) H_2SO_4 for 2 hours, and then hydrolysed in 4% (v/v) H_2SO for 1 hour at 121°C. The filtrate was measured for acid-soluble lignin by absorbance at 205 nm according to TAPPI Useful Method UM250, while the total weight of extracted, non-hydrolyzed components was determined gravimetrically for acid-insoluble lignin. Two independent 100 mg samples were analyzed.

For each sample, carbohydrate components of the filtrates, expressed as microgram carbohydrate per milligram of the dry weight of the sample, was analyzed by high performance anion-exchange chromatography on a CarboPac PA-1 column using a Dionex HPLC system (Dionex) equipped with a pulsed amperometric detector.

Thioacidolysis was employed on the extracted, ground *Arabidopsis* stems to determine the compositional ratio of monolignols, and was performed according to Lapierre *et al.* (1999), with the volumes scaled to accommodate 20 mg of starting material.

Thioglycolic acid (TGA) assay (Campbell and Ellis, 1992) was carried out to qualify the total lignin content in T3 homozygous lines. In brief, crude cell-wall fractions were

obtained by centrifugation after extracting soluble material from air-dried mature inflorescence stems in 70% (v/v) ethanol at 70°C for 1 hour (Reiter *et al.*, 1993). The dry weight of the cell-wall preparation was measured. The assay was conducted on four replicates. The cell-wall material was then saponified in 1 M NaOH for 3 hours followed by the extraction of base-insoluble material after the centrifugation. Base-insoluble material was incubated for 3 hours at 80°C in the solution containing distilled water, containing concentrated HCl and thioglycolic acid. The insoluble material was then incubated in 1 M NaOH for 12 hours on a rotating shaker at room temperature. Insoluble material was collected by centrifugation. The supernatant was transferred to a fresh 1.5 ml tube. Concentrated HCl (200 µl) was added to the transferred supernatant prior to the incubation for 4 hours at 4°C. The precipitate was resuspended in 1 M NaOH followed by measurement of total lignin content at 280 nm.

3.4 Results

3.4.1 Construction of binary vector pCAMCCR

A schematic outline of the T-DNA region of the final binary vector, pCAMCCR, is provided in Figure 3.2. The T-DNA region contained the expression cassette with the *Eucalyptus CCR* gene under the control of the CaMV 35S promoter. The CaMV 35S terminator was used to terminate the transcription of the *Eucalyptus CCR* cDNA. As shown in Figure 3.2, a *35S::GUS* gene construct was also included. The *Bar* gene, which confers resistance to the herbicide Basta was included in the T-DNA region.

3.4.2 Screening and segregation analysis of transgenic plants

After Basta selection, surviving T1 lines were allowed to self-pollinate. Four T2 lines, 3301A2, CCR-A2, CCR-B2 and CCR-C2 containing single T-DNA insertions were identified using a segregation analysis. T2 line 3301A2 contains empty pCAMBIA 3301 in the Col-0 background; whereas CCR-A2, CCR-B2 and CCR-C2 are three independent T2 lines containing *EUCCR* in the Col-0 background. The T2 lines were advanced to T3 homozygous lines, 3301A3, CCR-A3, CCR-B3 and CCR-C3, respectively.

3.4.3 Transcript levels of *EUCCR* in T2 and T3 plants

We confirmed the expression of *EUCCR* in inflorescence stems of transgenic *Arabidopsis* by semi-quantitative RT-PCR. A total of 25 PCR cycles were used to amplify *Actin2* and *AtCCR1*, while 15 cycles were sufficient to amplify *EUCCR* after optimization. The results revealed high *EUCCR* transcription levels in overexpressing T2 (Figure 3.3A) and T3 (Figure 3.4A) plants compared with those of control plants, indicating the constitutive expression of *EUCCR* in transgenic *Arabidopsis*. No obvious changes were observed in transcription levels of the endogenous *AtCCR1* in T2 (Figure 3.3C) and T3 plants (Figure 3.4C). The differences between the transcription levels of *EUCCR* in CCR-A3, CCR-B3 and CCR-C3, were not significant (Figure 3.4A).

3.4.4 Lignin determination

In order to test the effect of the overexpression of *EUCCR* on cell wall chemistry of *Arabidopsis* plants, we quantified the lignin content in inflorescence stems of control and transgenic plants. Total micro-Klason lignin content in primary stems of the T2 overexpressing line was found to be 5% higher than in the T2 control line (Figure 3.5).

Acid insoluble lignin levels were decreased, while acid soluble lignin levels were significantly (67%, at $\alpha = 0.05$) increased in the T2 line in comparison to 3301A2. Total lignin content determined using the TGA assay in the three T3 homozygous overexpressing lines, CCR-A3, CCR-B3 and CCR-C3, was significantly (at $\alpha = 0.01$) increased by 12.4%, 11.9% and 12.7 % , respectively compared with 3301A2 (Figure 3.6).

3.4.5 Analysis of lignin-derived monomers

Lignin composition was investigated by thioacidolysis, expressed as percentage of dry weight of each monomer. The key reaction of thioacidolysis is the generation of G and S monomers from G and S units, respectively, by the cleavage of labile β-O-4 ether bonds, which are the target of the industrial delignification processes due to its frequency and labile properties (Higuchi, 1990). The results showed an increase of 16% in G lignin units and a decrease of 21% in S lignin units in the *CCR* overexpressing plant (Figure 3.7A). The S/G ratio was therefore significantly altered (deceased by 32%) relative to Col-0 (Figure 3.7B).

3.4.6 Carbohydrate analysis

As the *CCR* gene has an effect on lignin content in T2 plants, which may impact on carbon flux towards carbohydrate biosynthesis, we also measured cellulose content (expressed as glucose) and non-cellulose fractions of cell walls of T2 plants. Figure 3.8A shows that the glucose content of cell wall material in overexpressing plants was significantly decreased (by 10%, at $\alpha = 0.05$) and the total sugar level in stems was lower (8%) in T2 plants compared to 3301A2. Although we could not directly distinguish between glucose derived from cellulose and hemi-cellulose, we assumed

that most of the glucose was derived from cellulose. No significant difference was revealed in the total non-cellulosic components of stem cell walls between 3301A2 and T2 plants, although significant (at α = 0.05 level) differences were observed in the levels of xylose with lower contents in the stem material of T2 plants compared to 3301A2 (Figure 3.8B).

3.5 Discussion

In this study we investigated the use of *Arabidopsis thaliana* as a model system to analyze the functions of genes involved in wood development in *Eucalyptus*. In particular, the effect of overexpressing the *Eucalyptus CCR* gene on cell wall chemistry of inflorescence stems of *Arabidopsis* was analyzed. Using chemical analysis methods, such as TGA, thioacidolysis and HPLC, it was found that the components of the cell walls in T2 and T3 transgenic overexpressing *Arabidopsis* were indeed altered in a predicted fashion.

The overexpression of the *Eucalyptus CCR* gene in *Arabidopsis* leads to an increase in total lignin content of *Arabidopsis* stems. This result is consistent with phylogenetic and functional analysis of *CCR* genes from other plants species, such as tobacco (Piquemal *et al*., 1998) and *Arabidopsis thaliana* (Jones *et al.*, 2001). As shown by Jones *et al.* (2001), *EUCCR*, *Arabidopsis CCR* (*AtCCR1*) and tobacco *CCR* (*TobCCR*) genes group closely together in a phylogenetic tree indicating that these genes are most probably orthologs. AtCCR1 deficient plants (*irx4* mutants), have 50% less lignin than wild type *Arabidopsis* (Jones *et al.*, 2001). Down-regulation of *TobCCR*, using antisense constructs induced a remarkable reduction (47%) in lignin content (Piquemal *et al*.,

1998).

The S/G ratio was dramatically decreased due to the increased amounts of G monomer and decreased amounts of S monomer in overexpressing T2 plants (Figure 3.7). This result and results from other studies (Piquemal *et al*., 1998) suggest that EUCCR and other homologs might be G-lignin specific proteins, although CCR is traditionally considered as a common protein for S and G lignin biosynthesis (Boudet, 2000; Boudet *et al.*, 2003). Piquemal *et al*. (1998) reported an increased S/G ratio due to the decreased amounts of G units (but constant amounts of S units) in tobacco plants with down-regulated CCR activity. Bioinformatic analysis of the promoters of lignin genes has also suggested a possible link between *AtCCR1* homologs and G-lignin synthesis. The presence of an AC element in the promoter regions of some lignin biosynthesis genes was found to be related to G lignin biosynthesis (Raes *et al.*, 2003). The promoter regions of both *AtCCR1* (Raes *et al.*, 2003) and *EUCCR* (Lacombe *et al.*, 2000) contain conserved AC elements. There are three possible explanations for these findings: (1) distinct CCR isoforms exist that specifically utilize different CoA esters; (2) some kind of metabolic channelling places a form of CCR in a complex with other enzymes committed to G lignin biosynthesis; (3) the pathway to S lignin somehow bypasses CCR (Dixon *et al.*, 2001).

Carbon flux between lignin and carbohydrate biosynthesis was altered by overexpressing *EUCCR* in *Arabidopsis*. Changes in carbon flux associated with lignin and cellulose mutants are well-known phenomena. Cellulose and lignin are the major sinks for carbon on earth (Delmer and Haigler, 2002). It has been reported that in *Populus tremuloides* the down-regulation of another lignin biosynthesis gene, 4-coumarate:coenzyme A ligase (4CL), can result in a remarkable decrease in lignin

content with a corresponding increase in cellulose content (Hu *et al.*, 1999). Here, the cellulose level, expressed as glucose, was decreased with a corresponding increase in lignin content (Figure 3.5; Figure 3.8). The decreased S monomer with increased G monomer (Figure 3.7A) could also be due to the carbon flux from S lignin to G lignin, due to increased G lignin-specific CCR activity. The possible carbon flux from G lignin to S lignin was found in the other study (Li *et al.*, 2003).

This study supports the metabolic channel model (Dixon *et al.*, 2001), where CCR is specific for G lignin and excluded from the S lignin synthesis pathway, because in this study G monomer was increased and S monomer decreased in transgenic plants overexpressing *EUCCR* (Figure 3.7A). We, however, cannot exclude the possibility that distinct CCR isoforms are used for S lignin biosynthesis. This implies that it could be possible to overexpress these isoforms to increase S/G ratio for industry benefit if they can be isolated in the future. The results of the present study are also beneficial for agriculture, because low S/G ratio facilitates forage digestibility for ruminant animals (Sewalt *et al.*, 1996). The carbon flux that exists between cellulose and lignin implies the possibility that lignin content might be simultaneously decreased if cellulose content can be increased by genetic engineering.

The results obtained from this study support the use of *Arabidopsis* to model the function of *Eucalyptus* wood formation genes in this herbaceous species. Obviously, this will significantly enhance the studies on wood formation process in *Eucalyptus* and other woody species. However, some genes, such as transcription factors and cell signalling genes in *Eucalyptus* may not function properly in *Arabidopsis*. Some woody plants-specific processes such as dormancy and ray cells formation (Taylor, 2002), can not be analyzed in *Arabidopsis*. Therefore, other model systems are needed as to model

these processes.

3.6 Acknowledgements

The authors would like to thank J. Grima-Pettenati for kindly providing the *EUCCR* cDNA clone, as well as E. Cowley and M. Mphahlele for excellent technical support. 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

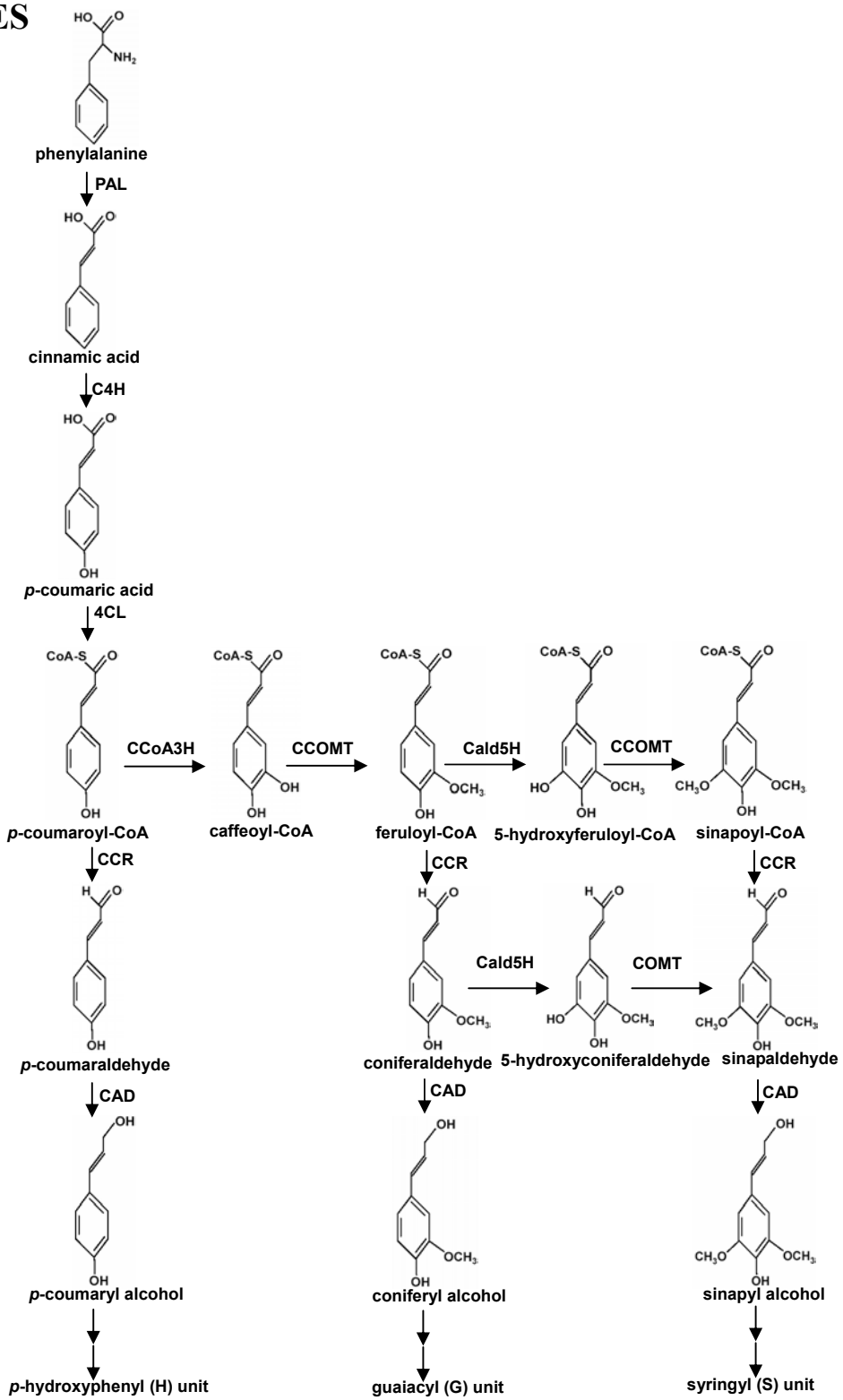


Figure 3.1. Pathways of monolignol biosynthesis. PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate:CoA ligase; CCoA3H: coumaroyl-CoA 3-hydroxylase; CCOMT: caffeoyl CoA 3-O-methyltransferase; Cald5H: coniferyl aldehyde-5-hydroxylase; CCR: cinnamoyl-CoA reductase; COMT: caffeic acid O-methyltransferase; CAD: cinnamyl alcohol dehydrogenase.

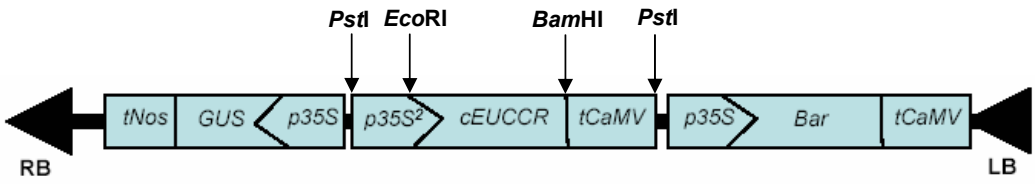


Figure 3.2. Schematic representation of the pCAMCCR construct. Only the T-DNA region is shown. LB and RB: *Agrobacterium* T-DNA left and right borders, respectively; tNos: terminator of the nopaline synthase gene; *GUS*: β-Glucoronidase gene; p35S: 35S CaMV promoter; $p35S^2$: double enhanced 35S CaMV promoter; cEUCCR: coding region of the *Eucalyptus CCR* gene; tCaMV: terminator of the 35S CaMV; Bar: coding sequence of *Bar* gene conferring resistance to herbicide, Basta.

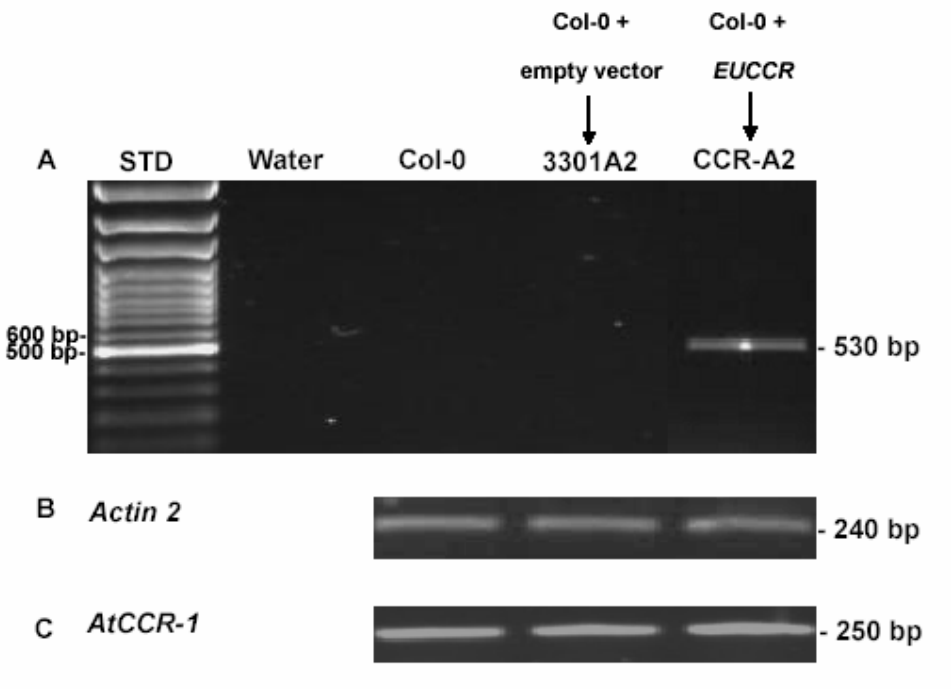


Figure 3.3. RT-PCR analysis of *EUCCR* in T2 *Arabidopsis* plants. Total RNA was isolated from inflorescence stems of wild-type Col-0 (Col-0), T2 control plants with empty pCAMBIA 3301 vector (3301A2) and a T2 overexpressing line (CCR-A2) containing *EUCCR*. RT-PCR was performed using A) *Eucalyptus CCR*-specific primers; B) *Actin 2*-specific primers; C) *AtCCR1*-specific primers. STD: 100 bp size standard; Water: water as negative control.

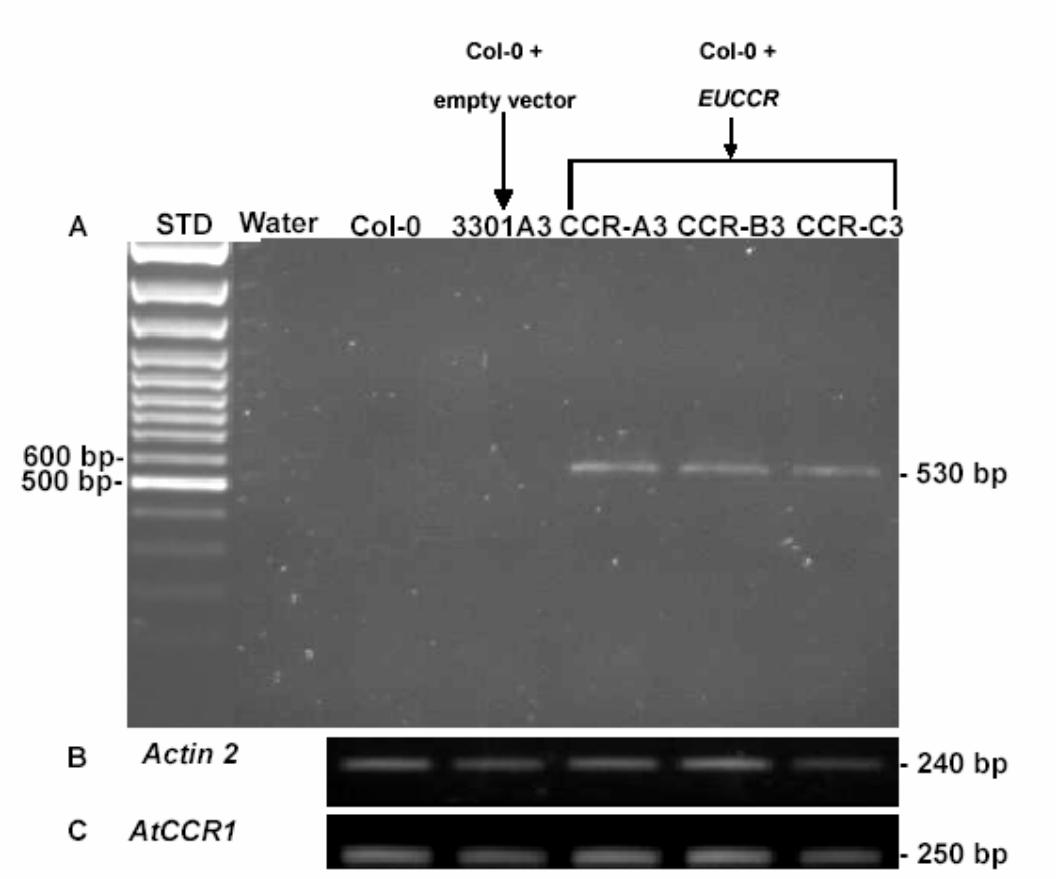


Figure 3.4. RT-PCR analysis of *EUCCR* in T3 homozygous *Arabidopsis* plants. Total RNA was isolated from inflorescence stems of the wild-type Col-0 (Col-0), T3 homozygous plants containing empty pCAMBIA 3301 vector (3301A3) and three T3 overexpressing lines (CCR-A3, CCR-B3, and CCR-C3) containing *EUCCR*. RT-PCR was performed using A) *Eucalyptus CCR*-specific primers; B) *Actin 2*-specific primers; C) *AtCCR1*-specific primers. STD: 100 bp size standard; Water: water as negative control.

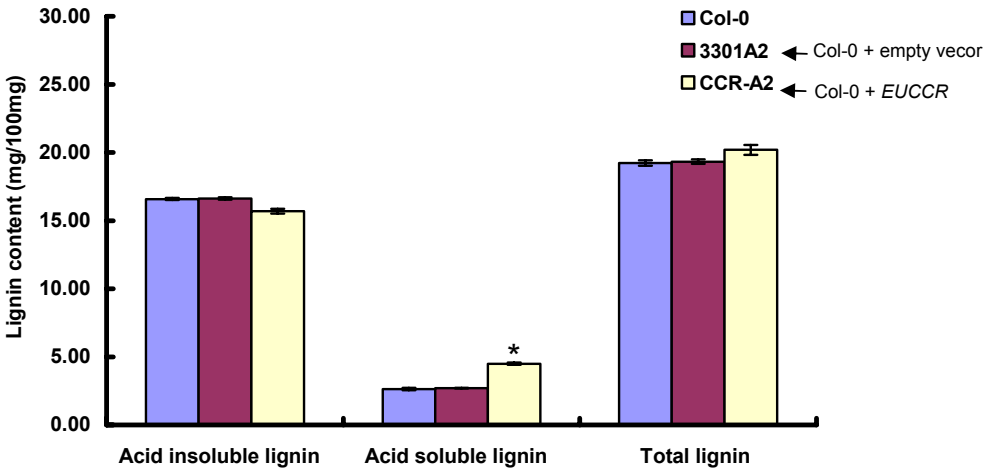


Figure 3.5. Micro-Klason lignin content determination. Acid insoluble, acid soluble and total lignin content were measured in the wild-type Col-0 (Col-0), T2 control plants transformed with empty pCAMBIA 3301 vector (3301A2) and T2 overexpressing plants (CCR-A2). Error bars denote the standard deviations of two replicates. *: significant difference between transgenic and control (3301A2) lines at 0.05 level of significance.

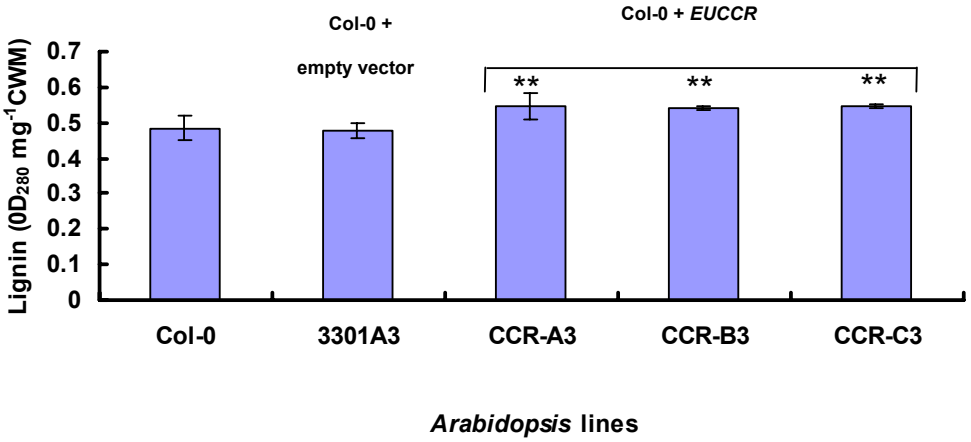


Figure 3.6. Total lignin content determined using the thioglycolic acid assay. Relative total lignin content was measured at OD_{280} per mg of cell wall material (CWM) in the wild-type Col-0 (Col-0) line, T3 homozygous positive control plants transformed with empty pCAMBIA3301 vector (3301A3) and three T3 overexpressing lines (CCR-A3, CCR-A3, and CCR-C3) containing *EUCCR*. Error bars denote the standard deviations of four replicates. **: significant difference between transgenic and control (3301A3) lines at 0.01 level.

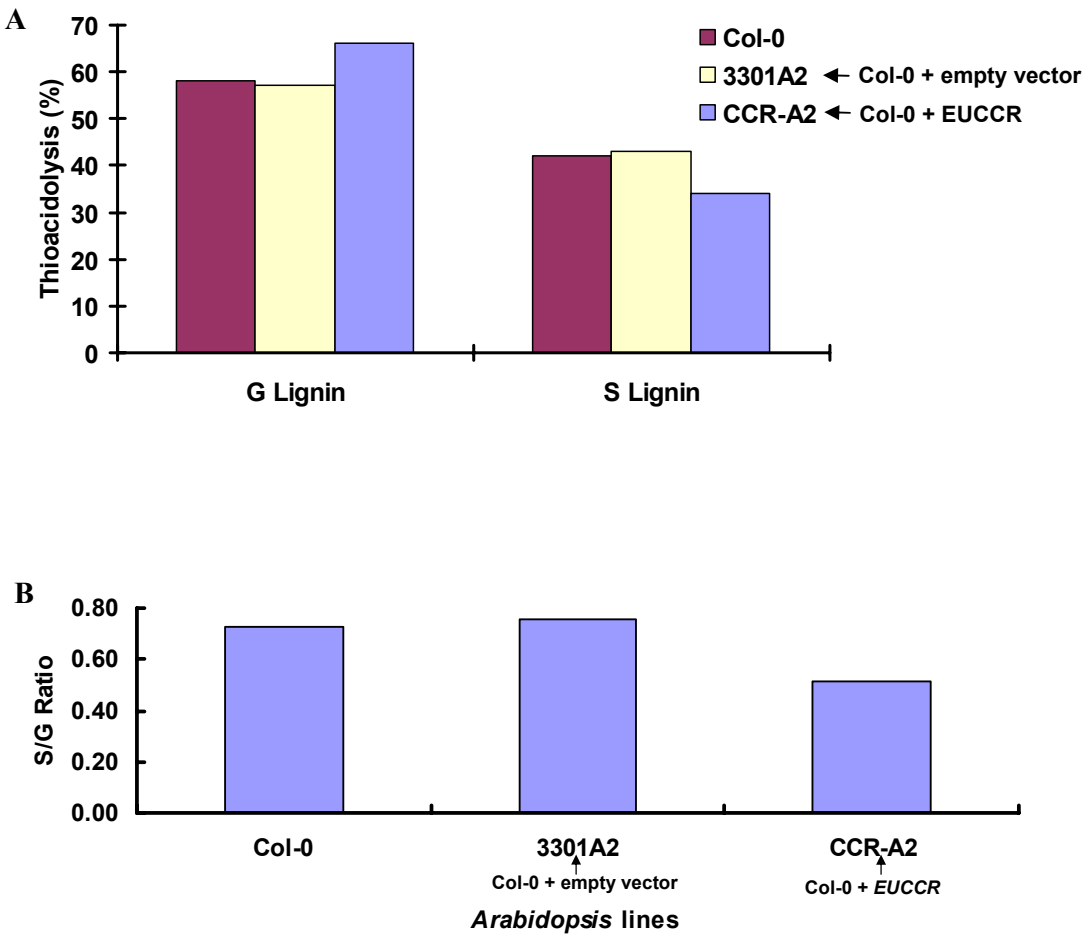


Figure 3.7. Thioacidolysis analysis of lignin monomer composition. (**A**) The weight percentages of G and S units and (**B**) S/G ratio in the wild-type Col-0 (Col-0), T2 positive control plants transformed with empty pCAMBIA3301 vector (3301A2) and T2 overexpressing plants (CCR-A2).

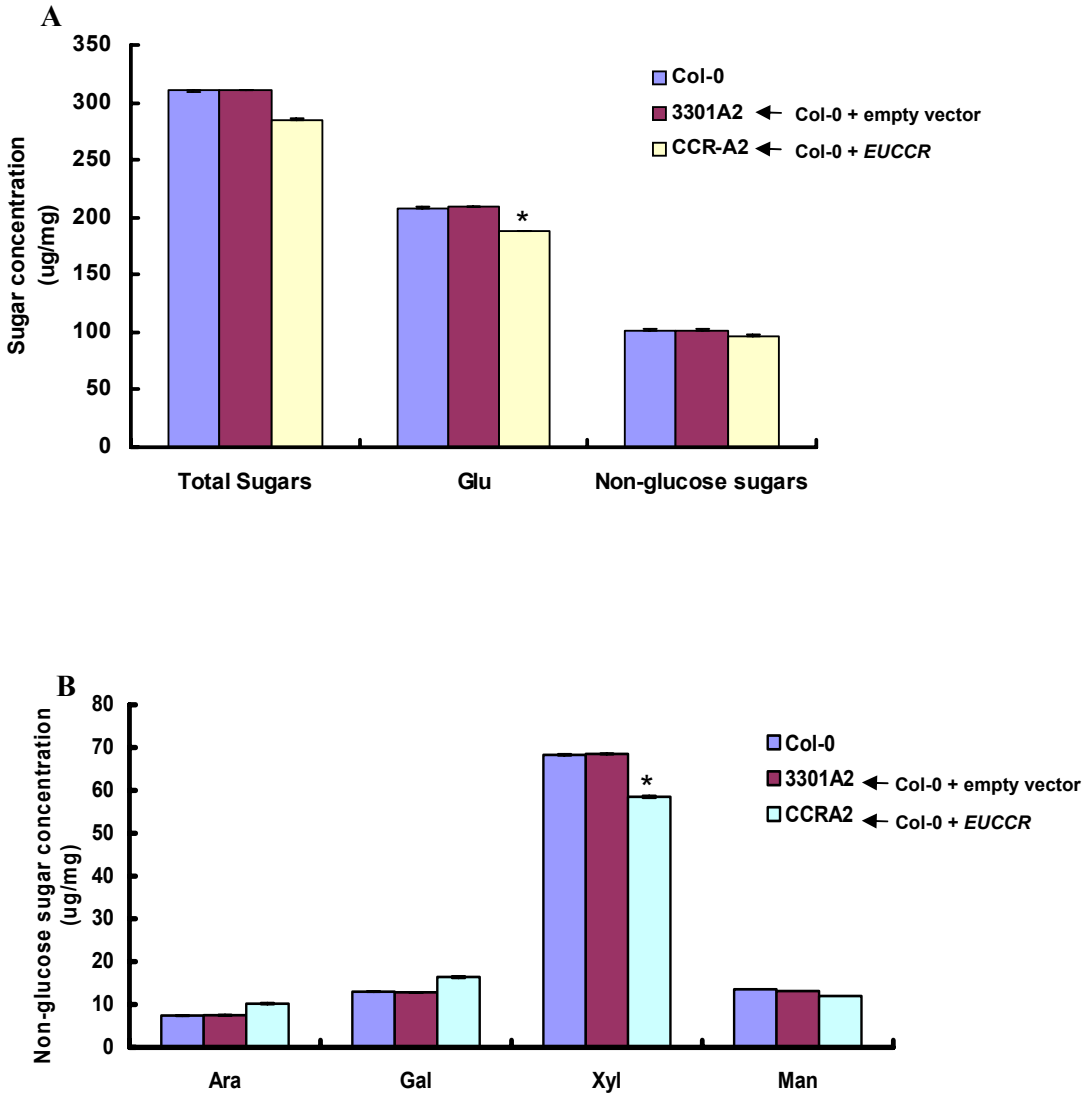


Figure 3.8. Carbohydrate concentration evaluated with HPLC. A) Total sugars, Glucose (Glu) and total non-glucose sugars, and B) four individual non-glucose sugars, Arabinose (Ara), Galactose (Gal), Xylose (Xyl) and Mannose (Man), were measured in the wild-type Col-0 (Col-0), T2 positive control plants transformed with empty pCAMBIA3301 vector (3301A2) and T2 overexpressing plants (CCR-A2). Error bars denote the standard deviations of two replicates. *: significant difference between transgenic and control (3301A2) lines at 0.05 level.

3.7 References

An G, Ebert PR, Mitra A, Ha SB (1992) Binary vector. In: S.B. Gelvin, R. A. Schilperoort (Eds.). *Plant Molecular Biology Manual*. A3, 1-19, Kluwer Academic Publishers, Dordrecht

Anterola AM, Lewis NG (2002) Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221-294

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annu. Rev. Plant. Biol.* 54, 519-546

Boudet AM (2000) Lignins and lignification: Selected issues. *Plant Physiol. Biochem.* 38, 81-96

Boudet AM, Kajita S, Grima-Pettenati J, Goffner D (2003) Lignins and lignocellulosics: a better control of synthesis for new and improved uses. *Trends in Plant Science* 8, 576-581

Campbell MM, Ellis BE (1992) Fungal elicitor-mediated responses in pine cell cultures: cell wall-bound phenolics. *Phytochemistry* 31, 737-742

Chaffey NJ, Cholewa E, Regan S, Sundberg B (2002) Secondary xylem development in *Arabidopsis*: a model for wood formation. *Physiologia Plantarum* 114, 594-600

Chen ZJ, Wang J, Tian L, Lee HS, Wang JJ, Chen M, Lee JJ, Josefsson C, Madlung A, Watson B, Pires JC, Lippman Z, Vaughn M, Colot V, Doerge RW, Martienssen RA, Comai L, Osborn TC (2004) The development of an *Arabidopsis* model system for genome-wide analysis of polyploidy effects. *Biological Journal of the Linnean Society* 82, 689-700

Chiang VL, Funaoka M (1990) The difference between guaiacyl and guaiacyl-syringyl lignins in their responses to kraft delignification. *Holzforschung* 44, 173-176

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735-743

Delmer DP, Haigler CH (2002) The regulation of metabolic flux to cellulose, a major sink

for carbon in plants. *Metabolic Engineering* 4, 22-28

Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a ‘‘metabolic grid’’, or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 57, 1069-1084

Fiers M, Hause G, Boutilier K, Casamitjana-Martinez E, Weijers D, Offringa R, Geest L, Campagne M, Liu CM (2004) Mis-expression of the *CLV3*/*ESR*-like gene *CLE19* in *Arabidopsis* leads to a consumption of root meristem. *Gene* 327, 37-49

Higuchi T (1990) Lignin biochemistry: biosynthesis and biodegradation. *Wood Sci. Technol.* 24, 23-63

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth of transgenic trees. *Nature Biotechnology* 17, 808-812

Huntley SK, Ellis D, Gilbert M, Chapple C, Mansfield SD (2003) Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. *J. Agric. Food Chem.* 51, 6178-6183

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

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

Lacombe E, Hawkins S, Doorsselaere JV, Piquemal J (1997) Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *Plant J.* 11, 429-441

Lacombe E, Doorsselaere JV, Boerjan W, Boudet AM, Grima-Pettenati J (2000) Characterization of *cis*-elements required for vascular expression of the *Cinnamoyl CoA Reductase* gene and for protein-DNA complex formation. *Plant J.* 23, 663-676

Lapierre C, Pollet B, Petit-Conil M, Toval G, Romero J, Pilate G, Leple JC, Boerjan W,

Ferret VV, De Nadai V, Jouanin L (1999) Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiol.* 119, 153-164

Larebeke NV, Engler G, Holsters M, Esacker SV, Schilperoort RA, Schell J (1974) Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability. *Nature* 252, 169-170

Li L, Zhou Y, Cheng X, Sun J, Marita JM, Ralph J, Chiang VL (2003) Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. *PNAS* 100, 4939-4944

Meyerowitz EM (2001) Prehistory and history of *Arabidopsis* research. *Plant Physiol.* 125, 15-19

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant* 15,473-496

Sambrook J., Tritsch EF, Maniatis T (1989) 'Molecular cloning: a laboratory manual (2nd ed)'. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sibout R, Baucher M, Gatineau M, Doorsselaere JV, Mila I, Pollet B, Maba B, Pilate G, Lapierre C, Boerjan W, Jouanin L (2002) Expression of a poplar cDNA encoding a ferulate-5-hydroxylase/coniferaldehyde 5-hydroxylase increases S lignin deposition in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 40, 1087-1096

Piquemal J, Lapierre C, Myton K, O'Connell A, Schuch W, Grima-Pettenati J, Boudet AM (1998) Down-regulation of Cinnamoyl-CoA Reductase induces significant changes of lignin profiles in transgenic tobacco plants. *Plant J.* 13, 71-83

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

Reiter WD, Chapple C, Somerville CR (1993) Altered growth and development in a fucose-deficient cell wall mutant of *Arabidopsis*. *Science* 261, 1032-1035

Restrepo MA, Freed DD, Carrington JC (1990) Nuclear transport of plant potyviral proteins. *The Plant Cell* 2, 987-998

Sewalt VJH, Glasser WG, Fontenot JP, Allen VG (1996) Lignin impact on fibre degradation: 1-Quinone methide intermediates formed from lignin during *in vitro* fermentation of corn stover. *Journal of the Science of Food and Agriculture* 71, 195-203

Taylor G (2002) *Populus*: *Arabidopsis* for forestry. Do we need a model tree? *Ann. Bot.* 90, 681-689

Tumbull JW (1991) Future use of eucalypts: opportunities and problems. In: Ed. A. P. G. Schonau. IUFRO Symp Intensive Forestry: The Role of Eucalypts. Southern Africa Institute of Forestry, Pretoria 2-27

Tzfira T, Zuker A, Altman A (1998) Forest-tree biotechnology: genetic transformation and its application to future forests. *TIBTECH* 16, 439-446

Chapter Four

Cloning, characterization and expression profiling of two novel *Eucalyptus* sucrose synthase genes isolated from xylem

Honghai Zhou, Frank Maleka, Alexander A. Myburg

Prepared as a manuscript for the journal *Functional Plant Biology*

4.1 Abstract

One of the functions of sucrose synthase (SuSy) is to supply the immediate substrate, UDP-glucose, for cellulose biosynthesis in higher plants. To clone candidate sucrose synthase genes with this function in wood-forming tissues of *Eucalyptus* trees, we applied degenerate PCR, rapid amplification of cDNA 3'end (3'RACE), 5'RACE and genome walking techniques. Two full-length cDNAs encoding SuSy genes, *EgSuSy1* and *EgSuSy3*, were isolated from immature xylem tissue of *Eucalyptus grandis*. A putative pseudogene (designated *EgSuSy2*) was detected during genomic analysis. The *EgSuSy1* cDNA was 2498 bp in length with an open reading frame of 2418 bp encoding 805 amino acids with a predicted molecular mass of 92.3 kDa. The 2528 bp full-length *EgSuSy3* cDNA contained the same length of open reading frame as *EgSuSy1,* but encoded a polypeptide with a predicted molecular mass of 92.8 kDa. Gene structure analysis, quantitative real-time reverse transcriptase-polymerase chain reaction (qRT-PCR) and phylogenetic analysis revealed that *EgSuSy1* and *EgSuSy3* may have functions related to cellulose biosynthesis during secondary cell wall formation in *Eucalyptus*.

Keywords: Sucrose synthase; UDP-Glucose; *Eucalyptus*; cellulose; wood

4.2 Introduction

Sucrose synthase (SuSy; UDP-glucose: D-fructose 2-glucosyltransferase; EC 2.4.1.13) is a homotetrameric enzyme that plays a crucial role in energy metabolism by channeling sucrose into various pathways (Kleczkowski, 1994). Sucrose synthase catalyzes the reversible conversion of sucrose and UDP to UDP-glucose (UDP-Glc) and fructose, but it functions to degrade sucrose under most conditions, particularly in nonphotosynthetic tissues (Kleczkowski, 1994). UDP-Glc can be channeled into cell wall components such as cellulose and callose (Amor *et al.*, 1995), as well as glycolysis, or starch synthesis (Geigenberger and Stitt, 1993).

Cellulose biosynthesis is a complex biological process by which cellulose synthase obtains UDP-Glc from sucrose that was hydrolyzed by SuSy to construct primary and secondary cell walls (Delmer & Haigler, 2002). Primary cell walls of woody species contain about 20% to 30% of cellulose, whereas cellulose in secondary cell walls contributes to about 40% to 50% of the dry weight of wood (Mellerowicz *et al.*, 2001). Cellulose is therefore a main component of wood and a raw material for the pulp and papermaking industry.

Some studies have focused on the role of SuSy in cellulose biosynthesis. Albrecht and Mustroph (2003) reported the increase (150% to 200%) of cellulose in wheat roots, including in endodermal cells with secondary cell walls, under hypoxia. Hypoxia has been demonstrated to lead to the increase of sucrose synthase activity (Albrecht *et al.*, 2003). *Sh1* is one of the sucrose synthase genes of maize and the cells of the *sh1* mutant have particularly thin, fragile, primary walls due to wall degeneration (Chourey *et al.*, 1998). The role of SuSy in cellulose biosynthesis is further supported by studies in transgenic carrot plants with suppressed SuSy activity where it was found that these

plants contained only about 63% of the normal crystalline cellulose (Tang and Sturm, 1999).

It has been proposed that there are two different forms of SuSy in higher plants (Amor *et al.*, 1995). One is the membrane-associated (P-SuSy), the other is soluble and abundant in the cytoplasm (S-SuSy) (Salnikov *et al.*, 2001). P-SuSy has a special role in contributing UDP-Glc to secondary wall cellulose biosynthesis and to callose synthesis, whereas S-SuSy may function to produce products used in general metabolism and for synthesis of storage polymers (Haigler *et al.*, 2001). Koch (2004) postulated that P-SuSy is directly associated with the plasma membrane-embedded cellulose synthase complex. The data at present supports the idea that the same gene encodes different P-and S-SuSy forms which are probably regulated by the phosphorylation status of the sucrose synthase enzyme (Haigler *et al.*, 2001). However, the precise mechanism that controls the different forms is still unknown.

Sucrose synthase genes have been isolated from a number of plants besides those described above. Two *Arabidopsis thaliana* sucrose synthase genes, *AtSUS1* (Chopra *et al.*, 1992) and *AtSUS2 (*Martin *et al.*, 1993), have been isolated. Three pea sucrose synthase genes, *Sus1*, *Sus2* and *Sus3* were isolated from a pea leaf cDNA library (Barratt *et al.*, 2001). To study the relationship between the sucrose synthase gene and sugar production, Lingle and Dyer (2001) isolated one *SuSy* from sugarcane. Kalluri (2003) isolated *SuSy* cDNA from poplar to analyze the relationship between this gene and cellulose biosynthesis in poplar.

In most of studies it was found that there are at least two non-allelic isoforms of SuSy that perform different physiological functions. In this study, we isolated two novel sucrose synthesis genes from *Eucalyptus* immature xylem to study the relationship

between *SuSy* and cellulose biosynthesis in *Eucalyptus*. Exon/intron organization, transcription patterns and phylogenetic relationships of the two *Eucalyptus* sucrose synthase genes were investigated.

4.3 Materials and Methods

4.3.1 Plant Materials

After removal of the bark to expose the xylem, the following tissues were collected from a 4-year-old *Eucalyptus grandis* tree into liquid nitrogen and stored at -80ºC until use: *immature xylem*, the outer layers of exposed xylem composed of vascular cambium cells and developing xylem tissue, *xylem*, deeper layers of xylem cells with advanced maturity after complete removal of immature xylem; *phloem*, the inner surface of the bark with vascular cambium and developing phloem cells, *cork*, bark material after the removal of phloem comprising cork, cork cambium and mature phloem, *young leaves*, unfolding young leaves; and *mature flowers*.

4.3.2 cDNA synthesis

Total RNA was extracted using a modified CTAB-based method for xylem tissues (Chang *et al.*, 1993). Genomic DNA in RNA samples was degraded using 1U of DNaseI (Roche Diagnostics GmbH) per microgram of DNA with 1 µl of RNasin (Promega, Madison, WI) and 1× reverse transcription buffer (Roche) in a total volume of 50 µl for 30 min at 37ºC. The resulting RNA samples were purified using the RNeasy kit (Qiagen, Valencia, CA). cDNA was then synthesized using the ImProm-II reverse transcription system (Promega). A polyT anchor primer (Table 4.1) was used to

synthesize the first-strand cDNA.

4.3.3 Degenerate PCR

Consensus-degenerate hybrid oligonucleotide primers (CODEHOP), SuSydeg-F and SuSydeg-R (Table 4.1), were designed according to the method of Rose *et al.* (1998). The primers were used to amplify *SuSy* gene fragments from both cDNA and genomic DNA. Genomic DNA was purified from young leaves of *Eucalyptus grandis* using the DNeasy Plant kit (Qiagen). Degenerate PCR was performed in a total reaction volume of 20 µl containing 1× ExSel buffer, 2.5 mM dNTPs (each), 0.4 µM each of SuSydeg-F and SuSydeg-R primers, 0.2U of ExSel polymerase, and 10 ng of template DNA (genomic DNA or cDNA). Touch-down PCR was performed with denaturation at 95°C for 30 sec, annealing at 62°C for 30 sec (with a decrease of 1°C after every cycle), and extension at 72°C for 1 min for 7 cycles. The next 5 cycles were performed with denaturation at 95°C for 30 sec, annealing at 48°C for 30 sec (with a decrease of 1°C after every cycle), and extension at 72°C for 1 min. The same denaturation and extension step was used for next 30 cycles, with an annealing temperature of 56°C. The reaction had an initial denaturation step of 95°C for 1 min and a final elongation step of 72°C for 20 min. PCR products were cloned into the pTZ57R vector of the InsT/Aclone PCR Cloning Kit (MBI Fermentas, Hanover, MD) and sequenced.

4.3.4 Rapid amplification of cDNA ends (RACE) and genome walking

To get the 3'ends of *Eucalyptus* sucrose synthase genes, rapid amplification of cDNA 3'ends (3'RACE) was performed according to Frohman *et al.* (1988). Gene-specific 3'RACE primers were used together with the generic 3' Anchor primer to amplify the 3'ends of the respective cDNAs. Nested 3'RACE primers (Table 4.1) were used in a

secondary PCR with the 3' Anchor primer to obtain a specific PCR product, if necessary.

To obtain the 5'ends of the *SuSy* cDNAs, 5'RACE was performed using the First Choice RLM-RACE kit according to the manufacturer's instructions (Ambion, Austin, TX). For 5'RACE, mRNA was first purified from total RNA of immature xylem of *Eucalyptus grandis* using the Oligotex mRNA kit (Qiagen) and used for cDNA synthesis. Nested, gene-specific 5' RACE primers (Table 4.1) were used to perform primary and secondary 5'RACE PCR.

Alternatively, the 5' ends of the genes were isolated using the Universal Genome walker kit (Clontech, Palo Alto, CA) for genome walking (Triglia *et al.*, 1988) according to the manufacturer's instructions. Nested gene-specific primers (Table 4.1) were used for genome walking. The final PCR products of 3'RACE, 5'RACE and genome walking were cloned into the pTZ57R vector with the InsT/Aclone PCR Cloning Kit (Fermentas) and sequenced.

4.3.5 Amplification of the full-length cDNAs and genomic sequence

UTR (untranslated region)-specific primers (Table 4.1) downstream of the predicted non-coding exons, were used to amplify the full-length cDNAs from immature xylem cDNA of *Eucalyptus grandis*. PCR was performed for thirty cycles with denaturation at 94°C for 30 sec, annealing at 58°C and 56°C for 3 min for the amplification of *EgSuSy1* and *EgSuSy3*, respectively, and extension at 72°C for 1 min. The reaction had an initial denaturation step of 95°C for 1 min and a final elongation step of 72°C for 10 min. The same primer pairs were used to amplify the full-length genomic copies of the genes and the primer pairs were annealed for 4 min. The genomic DNA used as template was

purified from young leaves of *Eucalyptus grandis*. At least three independent positive clones for each full-length cDNA were sequenced. One clone was sequenced for each full-length genomic copy.

4.3.6 Phylogenetic analysis

Phylogenetic analysis was performed using *MEGA* version 3.1 (Kumar *et al.*, 2004). The CLUSTAL function of this package was used to align the translated *Eucalyptus* cDNA sequences to homologous sucrose synthase proteins characterized in other plants. A neighbour-joining analysis was then conducted to create a bootstrap consensus tree. The statistic reliability of the tree topology was based on 1000 bootstrap replicates. Nodes with bootstrap scores of less than 50% were rejected.

4.3.7 Quantitative real-time RT-PCR analysis

Quantitative real-time RT-PCR (qRT-PCR) was performed to accurately quantify the transcript levels of the *Eucalyptus SuSy* genes. Gene-specific primers (Table 4.1) were used to amplify short amplicons for qRT-PCR analysis. *Eucalyptus* ADP ribosylation factor (GenBank accession no. **AY770746**) was used as internal control (Ranik *et al.*, unpublished). The reactions were performed on a LightCycler Instrument (Version 1.2, Roche Diagnostics GmbH) with the LightCycler-FastStar DNA Master SYBR green I kit (Roche) according to the manufacturer's instruction. A standard curve was created using a serial dilution of *Eucalyptus* ADP ribosylation factor cDNA. Here we assumed that PCR efficiency of the two *Eucalyptus SuSy* genes were the same as the internal control. qRT-PCR was performed with 10 ng of cDNA of each tissue as the template in a total volume of 20 µl. The reactions were incubated at 95ºC for 10 min to activate the hot-start DNA polymerase, followed by 35 cycles of 1 s at 95^{o}C, 10 s at 60^{o}C, and 20 s

at 72°C. The specificity of the RT-PCR amplification was checked using a heat dissociation protocol (from 65°C to 95°C) following the final cycle of the PCR. PCR products were purified and sequenced. The results obtained from different tissues were standardized to the expression level of the constitutive *Eucalyptus* ADP ribosylation factor gene. The second derivative maximum method was used to perform relative quantification by the use of the LightCycler software (version 3.5.3, Roche).

4.3.8 Nucleotide sequencing and analysis

Sequencing reactions were performed using the BigDye cycle sequencing kit (Applied Biosystems, Foster City, CA) and the samples were subjected to automated sequencing with an ABI Prism model 3100 sequencer (Applied Biosystems). Sequence analysis was done using the ABI Prism SeqScape Software Version 2.0 computer software program (Applied Biosystems). Sequence data were compared with the databases at the National Center for Biotechnology Information (NCBI) using the BLASTX program (www.ncbi.nlm.nih.gov/blast).

4.4 Results

4.4.1 Isolation and sequence analysis of two full-length sucrose synthase cDNAs

All of the primers mentioned in this section are listed in Table 4.1. Of the degenerate PCR products derived from *Eucalyptus* immature xylem cDNA, 21 positive clones were sequenced and they were found to represent two different sucrose synthase family

members (*EgSuSy1* and *EgSuSy3*). A PCR fragment obtained by degenerate PCR from *E. grandis* genomic DNA was found to represent another family member (designated *EgSuSy2*). We attempted to detect and clone the 3' end of the cDNA of *EgSuSy2* using 3'RACE. EgSuSy2-3'RACE and 3' Anchor primers were used for the primary PCR, while the EgSuSy2-3'RACE-nest and 3' Anchor primers were used for the secondary PCR. However, we failed to obtain the predicted PCR product from 3'RACE (data not shown). In addition, when we performed genomic walking towards the 5' end using gene-specific primers 5'EgSuSy1-GSP1 and 5'EgSuSy1-GSP2 (Table 4.1), we found that the putative start codon (ATG) was missing for *EgSuSy2* (data not shown). Therefore, we propose that *EgSuSy2* is a pseudogene.

The 3' ends of *EgSuSy1* and *EgSuSy3* were obtained by 3'RACE using EgSuSy1-3'RACE and EgSuSy3-3'RACE with the 3' Anchor primer (Table 4.1). For *EgSuSy3* it was necessary to do a nested PCR with EgSuSy3-3'RACE-nest together with the 3' Anchor primer (Table 4.1). We were able to amplify the 5' end of *EgSuSy3* using nested 5'RACE primers (5'EgSuSy3-GSP-1 and 2, Table 4.1). However, we had to use genome walking (with 5'EgSuSy1-GSP1 and 2) to clone the 5' end of *EgSuSy1*. Finally, UTR-specific primers (Table 4.1) were used to amplify the full-length cDNAs of *EgSuSy1* and *EgSuSy3*.

The full-length *EgSuSy1* cDNA was 2498 bp in length and contained 41 bp of 5' untranslated region (UTR), 39 bp of 3'UTR and an open reading frame of 2418 bp. The cDNA encoded 805 amino acids with a predicted molecular mass of 92.3 kDa. The 2528 bp full-length *EgSuSy3* cDNA contained 70 bp of 5'UTR, 40 bp of 3'UTR and the same length of open reading frame as *EuSUSY1* encoding a polypeptide with a predicted molecular mass of 92.8 kDa. The predicted molecular mass of the two *Eucalyptus*

sucrose synthases were approximately the same as that of sucrose synthases characterized in other plant species.

4.4.2 Exon/intron organization of *EgSuSy1* and *EgSuSy3*

The full-length genomic copies of *EgSuSy1* and *EgSuSy3* were sequenced to analyze the structures of the two genes. The total lengths of *EgSuSy1* and *EgSuSy3*, from putative start codons to stop codons, were 3832 bp and 3933 bp, respectively. The exon/intron boundaries were determined according to consensus splice sites using EMBOSS (http://bio.dfci.harvard.edu/Tools/EMBOSS/). The presence of a non-coding exon in the 5'UTRs of the two genes was not confirmed in this study. As revealed in Figure 4.1, the two genes contained 13 coding exons and both had the same exon/intron organization, although the length of each corresponding intron was different. The gene structures of *EgSuSy1* and *EgSuSy3* were similar to the conserved structure of sucrose synthase genes in the dicot SUS1 group (Komatsu *et al.*, 2002). Interestingly, the structures, including the length of each exon and the total numbers of coding exons, of the two genes were exactly the same as those of *CitSUS1* (*Citrus SUS1*, Komatsu *et al.*, 2002).

4.4.3 Phylogenetic analysis

Comparison of the amino acid sequences of the two *Eucalyptus SuSy* genes to that of homologous genes, *AtSUS1* (Baud *et al.*, 2004), *AtSUS4* (Baud *et al.*, 2004), *CitSUS1* (Komatsu *et al.*, 2002), *Gossypium hirsutum SS3* (Ruan *et al.*, 1997; Ruan *et al.*, 2003) and *PtrSuSy1* (Joshi *et al.*, 2004) revealed that *EgSuSy1* and *EgSuSy3* shared high amino acid identity (79%-85%) with other plant sucrose synthases (Figure 4.2; Table 4.2). *EgSuSy1* and *EgSuSy3* shared 86% identity (Table 4.2) and 91% similarity. As shown in Figure 4.2, comparisons of their amino acid sequences also revealed that the

N-termini of all of these seven proteins contain the highly conserved phosphorylation site, Ser, in the putative phosphorylation domain (LTRVH**S**LRER), which might be involved in the shift between P-SuSy and S-SuSy (Amor *et al.*, 1995; Huber *et al.*, 1996; Zhang *et al.*, 1999; Komatsu *et al.*, 2002).

A phylogenetic dendrogram of amino acid sequences of fifty sucrose synthases previously analyzed (Komatsu *et al.*, 2002) was generated using the *MEGA* software program to investigate the phylogenetic relationships between these sucrose synthases. Plant SuSy proteins have been divided into four groups (A new group, a monocot group, a dicot SUS1 group, and a dicot SUSA group, Komatsu *et al.*, 2002; Baud *et al.*, 2004) as shown in Figure 4.3. The deduced amino acid sequences of *EgSuSy1* and *EgSuSy3* showed a high degree of similarity to members of the large dicot SUS1 group (Figure 4.3). The two *Eucalyptus* genes were grouped into one clade with *AtSUS1*, *AtSUS4*, *CitSUS1*, *Gossypium hirsutum SS3* and *PtrSuSy1* (Joshi *et al.*, 2004).

4.4.4 Expression patterns of *EgSuSy1* and *EgSuSy3*

qRT-PCR analysis of *EgSuSy1* using the EgSuSy1-3'RACE and EgSuSy1-3'UTR-R primers (Table 4.1) produced a single gene fragment of 490 bp. Similarly, for *EgSuSy3,* qRT-PCR analysis with EgSuSy3-2086-F primer and EgSuSy3-3'UTR-R (Table 4.1) yielded a single gene fragment of 370 bp. Control qRT-PCR reactions based on the *Eucalyptus* ADP ribosylation factor (with primers Egr-ADP-RF-F and Egr-ADP-RF-R, Table 4.1) produced a fragment of 220 bp.

qRT-PCR analysis revealed that *EgSuSy1* was highly expressed in xylem, immature xylem, flower and young leaf tissues, but expressed at low levels in phloem and cork tissues (Figure 4.4). The expression pattern of *EgSuSy3* was similar to that of *EgSuSy1*,

although *EgSuSy1* was expressed at two to seven fold higher levels than *EgSuSy3* in all of the tissues investigated. Melting curve analysis revealed that only one RT-PCR product had been amplified for each gene (data not shown). This was corroborated by sequencing results which proved that the RT-PCR products were specific for *EgSuSy1* and *EgSuSy3* (data not shown).

4.5 Discussion

To isolate sucrose synthase genes possibly involved in cellulose biosynthesis in secondary xylem of *Eucalyptus* trees, degenerate PCR was performed on xylem cDNA and genomic DNA isolated from an *E. grandis* tree. Two *Eucalyptus* sucrose synthase genes, *EgSuSy1* and *EgSuSy3* were cloned from *Eucalyptus* xylem tissue. A fragment of a putative *Eucalyptus* sucrose synthase pseudogene, *EgSuSy2*, was also isolated from *Eucalyptus* genomic DNA. Quantitative RT-PCR analysis, phylogenetic analysis and exon/intron organization revealed that the two genes might both be involved in cellulose biosynthesis in primary and secondary cell walls of *Eucalyptus* trees.

The two *Eucalyptus SuSy* genes exhibited similar expression profiles across primary and secondary tissues of an *E. grandis* tree. The expression level of *EgSuSy3* was much lower than that of *EgSuSy1* in all of the tissues investigated (Figure 4.4). Both genes were expressed in young leaf, immature xylem and flower tissues enriched in primary cell wall biosynthesis, and in xylem tissues where massive amounts of secondary cell wall material is deposited. Similar patterns were found in *Arabidopsis* for *AtSUS1* and *AtSUS4* (Baud *et al*., 2004) and cotton for *SS3* (Ruan *et al.*, 1997). Kalluri (2003) using *in situ* hybridization of *PtrSuSy1* in poplar found that this homolog was expressed in

mature xylem tissues and other cells with developing primary cell walls. Although *EgSuSy1*, *EgSuSy3* and other homologs described above were transcribed in tissues enriched for both primary and secondary cell wall biosynthesis, the more accurate functions of these genes need to be studied at the protein level.

Phylogenetic analysis revealed that *EgSuSy1* and *EgSuSy3* were most closely related to the members of the Dicot SUS1 group, *AtSUS1*, *AtSUS4*, *CitSUS1*, *SS3* and *PtrSuSy1* (Figure 4.3). *CitSUS1* has been shown to supply UDP-Glc for cellulose synthesis during secondary cell wall formation of citrus (Komatsu *et al.*, 2002). The cotton sucrose synthase, *SS3*, might be involved in the secondary cell wall formation in the cotton fibres (Amor *et al.*, 1995; Ruan *et al.*, 1997), which are composed of almost pure cellulose in secondary cell walls (Haigler *et al.*, 2001). Suppression of *SS3* in cotton resulted in a fibreless phenotype (Ruan *et al.*, 2003). *EgSuSy1* and *EgSuSy3* were grouped together (Figure 4.3). This result, combined with expression levels of these two genes, high identity (86%, Table 4.2) and similarity (91%) between them and the same gene structures of the two genes, suggest that the two genes are recently diverged paralogs in the *Eucalyptus grandis* genome and that they may perform the same, redundant functions.

Exon/intron organization, transcriptional pattern and phylogenetic analysis suggest that *EgSuSy1* and *EgSuSy3* might be structurally and functionally conserved with *AtSUS1*, *AtSUS4*, *CitSUS1*, cotton *SS3* and *PtrSuSy1*. However the gene functions of these sucrose synthases is still not fully understood. For instance, the fibreless phenotype of cotton observed after suppression of *SS3* may be also due to reduced hexose levels (Ruan *et al.*, 2003). No mutations have been reported in *AtSUS1* or *AtSUS4* suggesting that they might be performing redundant functions and this redundant protein activity is

enough to channel UDP-Glc to cellulose to keep the plant growing normally. More detailed functions of *EgSuSy1*, *EgSuSy3*, *AtSUS1*, *AtSUS4*, *CitSUS1*, *SS3* and *PtrSuSy1* could be obtained from the double mutant of *AtSUS1* and *AtSUS4* using methods such as RNA interference (RNAi) technology.

The results from this study indicated that *EgSuSy1* and *EgSuSy3* might be the candidate genes that primarily supply UDP-Glc to cellulose in plant cell walls. Their high expression levels in secondary tissues may merely reflect the high rate of cellulose production in these tissues. This would also explain their high expression levels in rapidly developing young leaves. Cellulose biosynthesis was increased in bacteria transformed with modified mung bean *SuSy,* because of the enhanced recycling of UDP (Nakai *et al.*, 1999), which inhibits the reactions catalyzed by cellulose synthase (Ross *et al.*, 1991). It may be possible to produce transgenic *Eucalyptus* trees with enhanced cellulose content by the overexpression of *EgSuSy1*/*EgSuSy3* or by changing the phosphorylation status of *EgSuSy1*/*EgSuSy3* and, if necessary, by the simultaneous up-regulation of cellulose synthase genes and/or gene(s) that supply sucrose to SuSy.

4.6 Acknowledgements

The authors would like to thank N. Creux for assistance with genome walking and M. Ranik for help with transcription profiling. 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). The

nucleotide sequences of *EgSuSy1* and *EgSuSy3* reported here will be released in Genbank under the accession numbers DQ227993 and DQ227994, respectively.

FIGURES

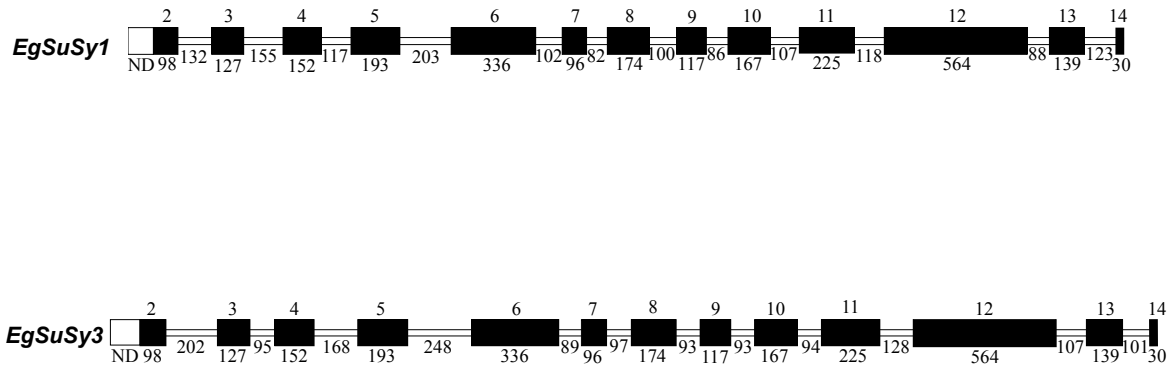
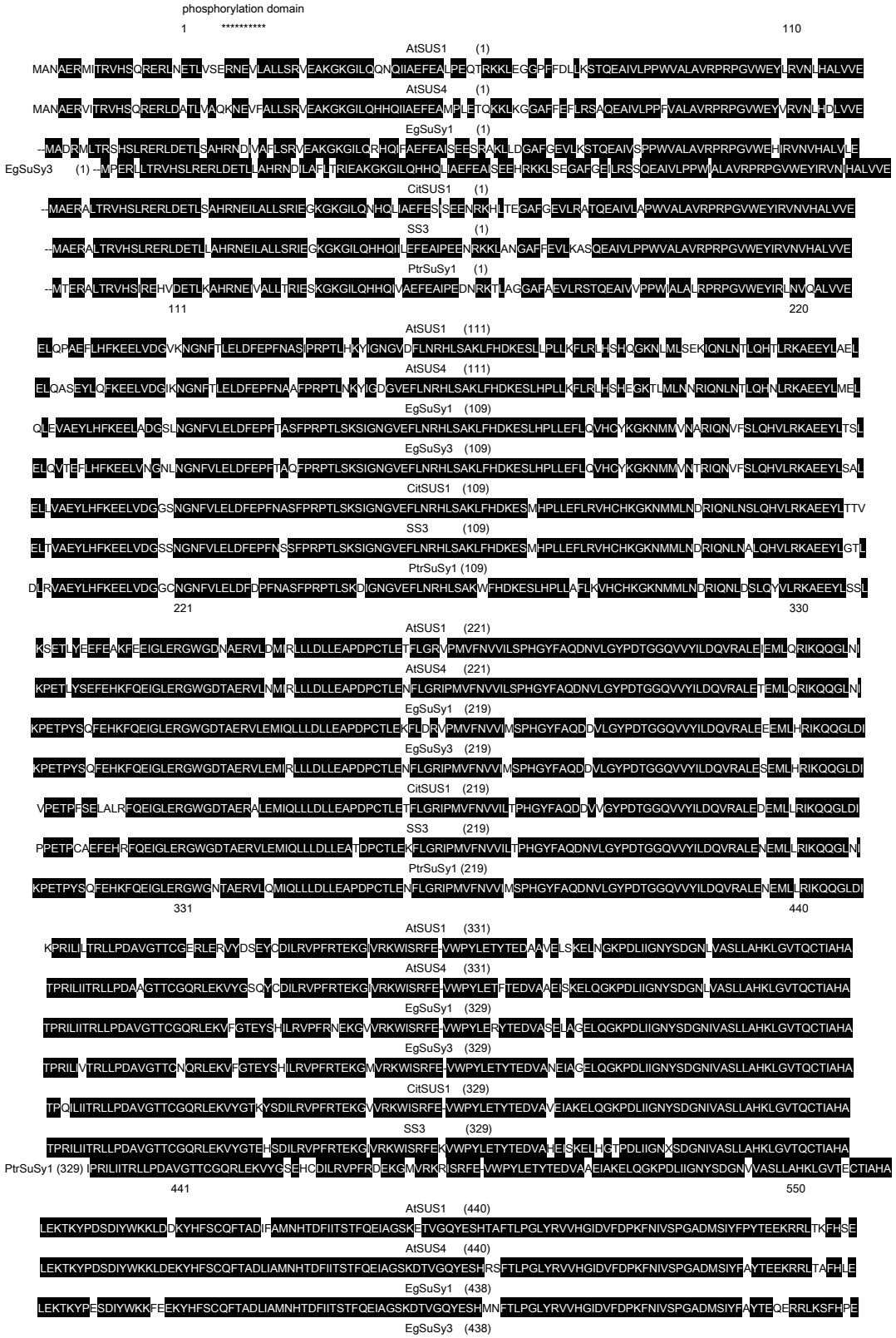


Figure 4.1. Exon/intron organization of *EgSuSy1* and *EgSuSy3*. The black boxes represent the exons and double lines connecting them are the introns. The numbers of each exon are indicated above it. The length, in base pairs, of each exon and intron is indicated below it. The possible presence of a non-coding exon 1 in 5'UTR was not considered. ND: length not determined.


phosphorylation domain
1 ********** 110
AtSUS1 (1)
AtSUS4 (1)
EgSuSy1 (1)
EgSuSy3 (1)
CitSUS1 (1)
SS3 (1)
PtrSuSy1 (1)
111 220
AtSUS1 (111)
AtSUS4 (111)
EgSuSy1 (109)
EgSuSy3 (109)
CitSUS1 (109)
SS3 (109)
PtrSuSy1 (109)
221 330
AtSUS1 (221)
AtSUS4 (221)
EgSuSy1 (219)
EgSuSy3 (219)
CitSUS1 (219)
SS3 (219)
PtrSuSy1 (219)
331 440
AtSUS1 (331)
AtSUS4 (331)
EgSuSy1 (329)
EgSuSy3 (329)
CitSUS1 (329)
SS3 (329)
PtrSuSy1 (329)
441 550
AtSUS1 (440)
AtSUS4 (440)
EgSuSy1 (438)
EgSuSy3 (438)

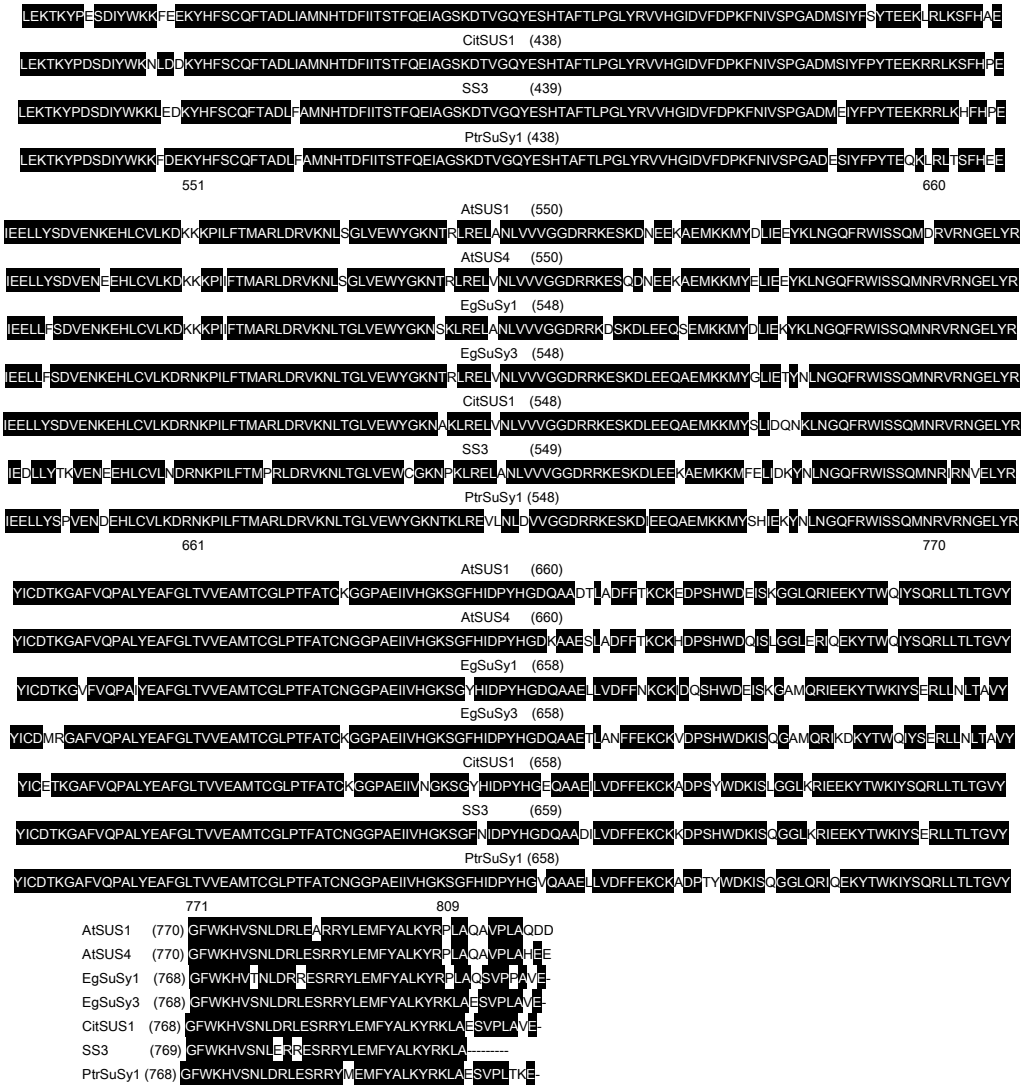


Figure 4.2. Multiple alignments of representative plant sucrose synthases. Species symbols at the beginning of gene names: At = *Arabidopsis thaliana*, Eg = *Eucalyptus grandis*, Cit = *Citrus unshiu* and Ptr = *Populus tremuloides.* Identical amino acids are indicated with black shading. The phosphorylation domain (LTRVH**S**LRER) is indicated above the sequences with asterisks.

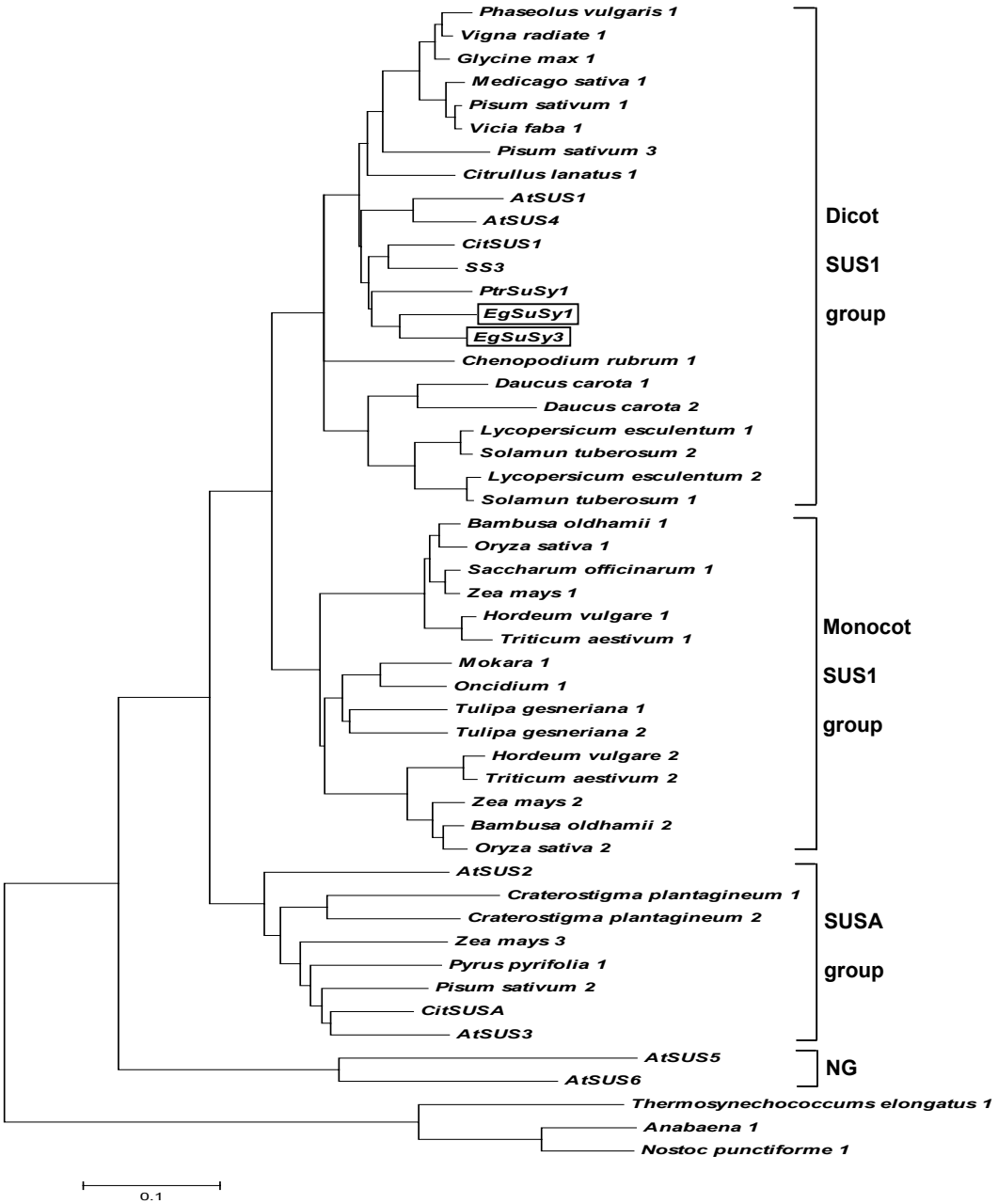


Figure 4.3. Phylogenetic dendrogram of amino acid sequence of 50 sucrose synthases. The statistic reliability of the tree topology was based on 1000 replicates for bootstrap analysis. Nodes with bootstrap scores of less than 50% were rejected. EMBL/Genbank database accession numbers of the displayed sucrose synthase genes are as follows: *Anabaena* 1, **AJ010639**; AtSUS1, **AK222090**; AtSUS2, **NM_124296**; AtSUS3, **NM_116461**; AtSUS4, **NM_114187**; AtSUS5, **NM_123077**; AtSUS6, **NM_105997**; *Bambusa oldhamii* 1, **AAL50571**; *Bambusa oldhamii* 2, **AAL50570**; *Chenopodium rubrum* 1, **X82504**; *Citrullus lanatus* 1, **BAA89232**; CitSUS1, **AB022092**; CitSUSA, **AB022091**; *Craterostigma plantagineum* 1, **AJ131999**; *Craterostigma plantagineum* 2, **AJ132000**; *Daucus carota* 1, **X75332**; *Daucus carota* 2, **Y16091**; EgSuSy1, **DQ227993**; EgSuSy3, **DQ227994**; *Glycine max* 1, **AF030231**; SS3, **U73588**; *Hordeum vulgare* 1, **X65871**; *Hordeum vulgare* 2, **X69931**; *Lycopersicum esculentum* 1, **L19762**; *Lycopersicum esculentum* 2, **AJ011319**; *Medicago sativa* 1, **AF049487**; *Mokara* 1, **AF530568**; *Nostoc punctiforme* 1, **AJ316589**; *Oncidium* 1, **AAM95943**; *Oryza sativa* 1, **X64770**; *Oryza sativa* 2, **X59046**; *Phaseolus vulgaris* 1, **AF315375**; *Pisum sativum* 1, **AJ012080**; *Pisum sativum* 2, **AJ001071**; *Pisum sativum* 3, **AJ311496**; PtrSuSy1, **AY341026;** *Pyrus pyrifolia* 1, **BAB20799**; *Saccharum officinarum* 1, **AF263384**; *Solamun tuberosum* 1, **U24088**; *Solamun tuberosum* 2, **U24087**; *Thermosynechococcums elongates* 1, **NP_681838**; *Triticum aestivum* 1, **AJ001117**; *Triticum aestivum* 2, **AJ000153**; *Tulipa gesneriana* 1, **X96938**; *Tulipa gesneriana* 2, **X96939**; *Vicia faba* 1, **X69773**; *Vigna radiate* 1, **D10266**; *Zea mays* 1, **X02400**; *Zea mays* 2, **L22296**; *Zea mays* 3, **AY059416**. NG: new group.

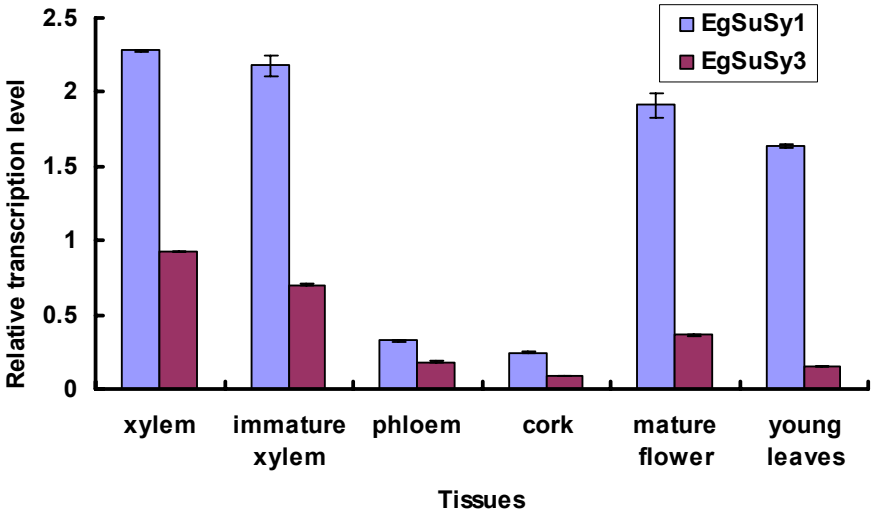


Figure 4.4. Relative expression levels of *EgSuSy1* and *EgSuSy3* in different *Eucalyptus* tissues. Relative expression levels were measured by quantitative real-time RT-PCR and standardized to the expression level of the constitutive ADP ribosylation factor. Error bars denote the standard deviations of three independent replicates.

TABLES

Table 4.1. Oligonucleotide primers used in this study.

Primer name	Primer sequence (5'⟶3')
EgSuSy161-3'RACE	TCTACCGCTACATCTGTGA
polyT anchor	GACCACGCGTATCGATGGCTCATTTTTTTTTTTTTTTV
3' Anchor	GACCACGCGTATCGATGGCTCA
SuSydeg-F	TTTGTGCTTGAATTGGAYTTYGARCCNTTY
SuSydeg-R	TCGGAATCAGGGTACTTNGTYTTYTC
EgSuSy1-3'RACE	TCTACCGCTACATCTGTGAC
EgSuSy2-3'RACE -nest	GTATCAGGATGCAGCAGGTGA A
EgSuSy2-3'RACE	GCCTTAGAGTTGCAAGGCTATCCA
EgSuSy3-3'RACE-nest	CGCCAATGAAATTGCAGGAGAAC
EgSuSy3-3'RACE	GAAGTGTGGCCTTACTTGGA
5'EgSuSy3-GSP-1	CCTTTCGGCAAAGCTCTTCC
5'EgSuSy3-GSP-2	AGTGGAGTTCCTTAACCGCC
EgSuSy1-5'UTR-F	GGTGGTTGTTGCAGTTCT
EgSuSy1-3'UTR-R	TACCAGAAGACGGAAGCATTGG
EgSuSy3-5'UTR-F	TCTGTCGGGTTCCGTTTTGC
EgSuSy3-3'UTR-R	CAGATCTCAAGAGGCAGACA
EgSuSy3-2086-F	GGTCCTGCTGAGATCATTGT
Egr-ADP-RF-F	TTCTGGTGCCATGCTGAGAA
Egr-ADP-RF-R	GATGCTGTGTTGCTCGTCTT
5'EgSuSy1-GSP1	CCAACATGTCCTGAGGAAGGCG
5'EgSuSy1-GSP2	CTCTCCGCTAAGCTCTTCCATGAC
5'EgSuSy2-GSP1	GACCTTTATGTTTGTGCGTAC
5'EgSuSy2-GSP2	GCAATGGCTCCAAGCAATCT

V: A, G, or C; Y: T or C; R: A or G; N: A, T, G or C.

Table 4.2. Comparison of amino acid sequences of seven similar sucrose synthases. The numbers indicate percentage identity between the sucrose synthases.

	AtSUS1	AtSUS4	CitSUS1	EgSuSy1	EgSuSy3	SS3	PtrSuSy1
EgSuSy1	79	80	83	100	86	82	82
EgSuSy3	80	82	85	86	100	83	84

The following are the EMBL/Genbank accession numbers of the corresponding sucrose synthases: AtSUS1-**AK222090**, AtSUS4-**NM_114187**, CitSUS1-**AB022092**, EgSuSy1-**DQ227993**, EgSuSy3-**DQ227994**, *Gossypium hirsutum* SS3-**U73588** and PtrSuSy1-**AY341026**.

4.7 References

Albrecht G, 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

Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *PNAS* 92, 9353-9357

Barratt DHP, Barber L, Kruger NJ, Smith AM, Wang TL, Martin C (2001) Multiple, distinct isoforms of sucrose synthase in pea. *Plant Physiol.* 127, 655-664

Baud S, Vaultier MN, Rochat C (2004) Structure and expression profile of the sucrose synthase multigene family in *Arabidopsis*. *J. Ex. Bot.* 55, 397-409

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

Chopra S, Del-favero J, Dolferus R, Jacobs M (1992) Sucrose synthase of *Arabidopsis*: genomic cloning and sequence characterization. *Plant Mol. Biol.* 18, 131-134

Chourey PS, Taliercio EW, Carlson SJ, Ruan YL (1998) Genetic evidence that the two isozymes of sucrose synthase present in developing maize endosperm are critical, one for cell wall integrity and the other for starch biosynthesis. *Mol. Gen. Genet*. 259, 88-96

Delmer DP, Haigler CH (2002) The regulation of metabolic flux to cellulose, a major sink for carbon in plants. *Metabolic Engineering* 4, 22-28

Frohman MA, Dush MK and Martin GR (1988) Rapid production of full-length cDNA from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *PNAS* 85, 8998-9002

Geigenberger P, Stitt M (1993) Sucrose synthase catalyses a readily reversible reaction in developing potato tubers and other plant tissues. *Planta* 189, 329-339

Haigler CH, Ivanova-Datcheva M, Hogan PS, Salnikov VV, Hwang S, Martin K,

Delmer DP (2001) Carbon portioning to cellulose synthesis. *Plant Mol. Biol.* 47, 29-51

Huber SC, Huber JL, Liao PC, Gage DA, McMichael Jr RW, Chourey PS, Hannah LC, Koch K (1996) Phosphorylation of serine-15 of maize leaf sucrose synthase (occurrence *in vivo* and possible regulatory significance). *Plant Physiol.* 112: 793-802

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

Kalluri UC (2003) Comparative expression analysis of cellulose biosynthesis related genes from aspen trees. PhD thesis. Houghton, MI, USA: Michigan Technological University

Kleczkowski LA (1994) Glucose activation and metabolism through UDP-glucose pyrophosphorylase in plants. *Phytochemistry* 37, 1507-1515

Koch K (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Current Opinion in Plant Biology* 7, 235-246

Komatsu A, Moriguchi T, Koyama K, Omura M, Akihama T (2002) Analysis of sucrose synthase genes in citrus suggests different roles and phylogenetic relationships. *J. Ex. Bot.* 53, 61-71

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5, 150-163

Lingle SE, Dyer JM (2001) Cloning and expression of sucrose synthase-1 cDNA from sugarcane. *J. Plant Physiol.* 158, 129-131

Martin T, Frommer WB, Salanoubat M, Willmitzer L (1993) Expression of an *Arabidopsis* sucrose synthase gene indicates a role in metabolization of sucrose both during phloem loading and in sink organs. *Plant J.* 4, 367-377

Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W (2001) Unravelling cell wall formation in the woody dicot stem. *Plant Mol. Biol.* 47, 239-274

Nakai T, Tonouchi N, Konishi T, Kojima Y, Tsuchida T, Yoshinaga F, Sakai F, Hayashi T (1999) Enhancement of cellulose production by expression of sucrose synthase in *Acetobacter xylinum*. *PNAS* 96, 14-18

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* 26, 1628-1635

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

Ruan YL, Chourey PS, Delmer DP, Perez-Grau L (1997) The differential expression of sucrose synthase in relation to diverse patterns of carbon partitioning in developing cotton seed. *Plant Physiol.* 115, 375-385

Ruan YL, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *The Plant Cell* 15, 952-964

Salnikov VV, Grimson MJ, Delmer DP, Haigler CH (2001) Sucrose synthase localizes to cellulose synthesis sites in tracheary elements. *Phytochemistry* 57, 823-833

Tang GQ, Sturm A (1999) Antisense repression of sucrose synthase in carrot (*Daucus carota* L.) affects growth rather than sucrose partitioning. *Plant Mol. Biol.* 41, 465-479

Triglia T, Peterson MG, Kemp DJ (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* 16, 8186

Zhang XQ, Lund AA, Sarath G, Cerny RL, Roberts DM, Chollet Raymond (1999) Soybean nodule sucrose synthase (nodulin-100): further analysis of its phosphorylation using recombinant and authentic root-nodule enzymes. *Archives of Biochemistry and Biophysics* 371, 70-82

SUMMARY

Isolation and functional genetic analysis of *Eucalyptus* wood formation genes

Honghai Zhou

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

Co-supervised by ***Prof. David K. Berger***

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

Department of Genetics

University of Pretoria

Eucalyptus trees are an important source of wood and fibre. The wood (secondary xylem) of this genus is widely used for pulp and papermaking. However, our understanding of the mechanisms which control the wood formation process (xylogenesis) in *Eucalyptus* and other woody species is far from complete. One reason is that xylogenesis is a very complex developmental process. The major components of wood are lignin and cellulose. Many genes involved in lignin and cellulose biosynthesis have been characterized. For example, *Cinnamoyl CoA reductase* (*CCR*) and *cinnamyl alcohol dehydrogenase* (*CAD*) are two important lignin biosynthesis genes. Plant cellulose is synthesized by cellulose synthase enzymes with the aid of some other proteins, such as sucrose synthase (SuSy) and sucrose phosphate synthase (SPS). Another factor which makes it difficult to analyze the function of *Eucalyptus* wood formation genes *in vivo*, is the long generation times of *Eucalyptus* trees and the difficulty to obtain transgenic *Eucalyptus* plants. Therefore, in this study, we

investigated the use of *Arabidopsis thaliana* as a model system for functional analysis of wood formation genes..

We transformed a lignin and a cellulose biosynthesis gene isolated from *Eucalyptus* to wild-type and mutant genetic backgrounds of *Arabidopsis* in order to test our ability to modify the cell wall chemistry of *Arabidopsis thaliana* using tree genes. The *Eucalyptus CCR* (*EUCCR*) gene was transformed into wild-type *Arabidopsis* (Col-0) and *irregular xylem 4* (*irx4*) mutant plants, in which the homolog of *EUCCR* is mutated. A *Eucalyptus* cellulose synthase gene (*EgCesA1*) was also transformed into *irregular xylem 1* (*irx1*) mutant plants, in which the homolog of *EgCesA1* is mutated. Transgenics were only obtained from wild-type Col-0 transformed with *EUCCR* and from *irx1* transformed with *EgCesA1*. We studied the cell wall chemistry of wild-type *Arabidopsis* plants overexpressing the *Eucalyptus* CCR gene.

Chemical analysis of inflorescence stems revealed the modification of lignin and cellulose content in transgenic plants. Total lignin content was increased in T2 (5%) and T3 (12%) lines as revealed by micro-Klason lignin and thioglycolic acid quantification methods, respectively. High Pressure Liquid Chromatography (HPLC) analysis revealed that cellulose content was significantly decreased (10%) in T2 transgenic plants. This suggested the reallocation of carbon from cellulose to lignin as a result of overexpression of EUCCR in transgenic plants. Interestingly, thioacidolysis analysis revealed that in T2 plants, monomethoxylated guaiacyl (G) monomer was increased (16%) and bimethoxylated syringyl (S) monomer was decreased (21%). Therefore, the S/G lignin monomer ratio was significant decreased (32%). This implied that *EUCCR* might be specific to G monomer biosynthesis. The results described above confirmed that *Arabidopsis thaliana* can be used to model the function of wood formation genes

isolated from *Eucalyptus*.

Two novel full-length *Eucalyptus* sucrose synthase (*SuSy*) genes, *EgSuSy1* and *EgSuSy3*, and one putative pseudogene, *EgSuSy2*, were also isolated in this study. Degenerate PCR was used to amplify *Eucalyptus SuSy* fragments from cDNA and genomic DNA. 3'RACE was used to amplify the 3' ends of two *Eucalyptus SuSy* genes. Genome walking was performed to obtain the 5' ends of *EgSuSy1* and *EgSuSy2*, whereas 5'RACE technology was used to isolate the 5' end of *EgSuSy3*. However, 3'RACE analysis failed when we tried to identify the 3' end of *EgSuSy2*. Sequencing results from the genome walking product of *EgSuSy2* further revealed that the start codon of this gene was missing, and we hypothesize that this is a psuedogene in the *Eucalyptus* genome. The *EgSuSy1* cDNA was 2498 bp in length with an open reading frame of 2418 bp encoding 805 amino acids with a predicted molecular mass of 92.3 kDa. The 2528 bp full-length *EgSuSy3* cDNA contained the same length of open reading frame as *EgSuSy1,* but encoded a polypeptide with a predicted molecular mass of 92.8 kDa. The results of quantitative real-time RT-PCR, phylogenetic analysis and gene structure of the two genes revealed that both genes might be involved in cellulose biosynthesis in primary and secondary cell walls of *Eucalyptus*. These two genes, *EgSuSy1* and *EgSuSy3*, could therefore be useful targets for genetic engineering of wood properties in *Eucalyptus*.

APPENDIX

A: DNA sequence data

EgSuSy1

```
LOCUS       DQ227993            2498 bp    mRNA    linear   PLN 29-SEP-2005
DEFINITION  Eucalyptus grandis sucrose synthase 1 (EgSuSy1) mRNA, complete cds.
ACCESSION   DQ227993
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 2498)
  AUTHORS   Zhou,H. and Myburg,A.A.
  TITLE     Cloning, characterization and expression profiling of two
            Eucalyptus sucrose synthase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2498)
  AUTHORS   Zhou,H. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2005) Genetics Department, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..2498
                     /organism="Eucalyptus grandis"
                     /db_xref="taxon:71139"
                     /note="Organelle: nucleomorph"
     gene            1..2498
                     /gene="EgSuSy1"
     5'UTR           1..41
                     /gene="EgSuSy1"
     CDS             42..2459
                     /gene="EgSuSy1"
                     /note="EC 2.4.1.13"
                     /codon_start=1
                     /product="sucrose synthase (SuSy)"
                     /translation="MADRMLTRSHSLRERLDETLSAHRNDIVAFLSRVEAKGKGILQR
                     HQIFAEFEAISEESRAKLLDGAFGEVLKSTQEAIVSPPWVALAVRPRPGVWEHIRVNV
                     HALVLEQLEVAEYLHFKEELADGSLNGNFVLELDFEPFTASFPRPTLSKSIGNGVEFL
                     NRHLSAKLFHDKESLHPLLEFLQVHCYKGKNMMVNARIQNVFSLQHVLRKAEEYLTSL
                     KPETPYSQFEHKFQEIGLERGWGDTAERVLEMIQLLLDLLEAPDPCTLEKFLDRVPMV
                     FNVVIMSPHGYFAQDDVLGYPDTGGQVVYILDQVRALEEEMLHRIKQQGLDITPRILI
                     ITRLLPDAVGTTCGQRLEKVFGTEYSHILRVPFRNEKGVVRKWISRFEVWPYLERYTE
                     DVASELAGELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYW
                     KKFEEKYHFSCQFTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHMNFTLPGLYRV
```

```
                     VHGIDVFDPKFNIVSPGADMSIYFAYTEQERRLKSFHPEIEELLFSDVENKEHLCVLK
                     DKKKPIIFTMARLDRVKNLTGLVEWYGKNSKLRELANLVVVGGDRRKDSKDLEEQSEM
                     KKMYDLIEKYKLNGQFRWISSQMNRVRNGELYRYICDTKGVFVQPAIYEAFGLTVVEA
                     MTCGLPTFATCNGGPAEIIVHGKSGYHIDPYHGDQAAELLVDFFNKCKIDQSHWDEIS
                     KGAMQRIEEKYTWKIYSERLLNLTAVYGFWKHVTNLDRRESRRYLEMFYALKYRPLAQ
                     SVPPAVE"
     3'UTR           2460..2498
                     /gene="EgSuSyl"
BASE COUNT      606 a    605 c    662 g    625 t
ORIGIN
        1 ggtggttgtt gcagttcttt ttctgagaga agaatttaga catggctgat cgcatgttga
       61 ctcgaagcca cagccttcgc gagcgtttgg acgagaccct ctctgctcac cgcaacgata
      121 ttgtggcctt cctttcaagg gttgaagcca agggcaaagg catcttgcag cgccaccaga
      181 tttttgctga gtttgaggcc atctctgagg agagcagagc aaagcttctt gatggggcct
      241 ttggtgaagt cctcaaatcc actcaggaag cgattgtgtc gcctccatgg gttgctcttg
      301 ctgttcgtcc aaggccgggc gtgtgggagc acatccgtgt gaacgtccat gcgcttgttc
      361 ttgagcaatt ggaggttgct gagtatctgc acttcaaaga agagcttgct gatggaagct
      421 tgaatggtaa ctttgtgctt gagcttgact ttgagccatt cactgcctct tttccgcgcc
      481 cgactctttc caagtctatt ggcaatggcg tcgagtttct caatcgccat ctctccgcta
      541 agctcttcca tgacaaggaa agcttgcacc ctctgcttga attcctccaa gtccactgct
      601 acaaggggaa gaacatgatg gtgaatgcca gaatccagaa tgtgttctcc ctccaacatg
      661 tcctgaggaa ggcggaggag tatctgacct cgctcaaacc cgagaccccg tactcccagt
      721 tcgagcacaa gttccaggag atcgggctcg agcgggggtg gggtgacacg gctgagcgcg
      781 tcctcgagat gatccagctc ctgttggatc tccttgaggc tcccgacccg tgcactctcg
      841 agaagttctt ggatagggtt cccatggtct tcaacgtcgt gatcatgtct ccccacggat
      901 actttgctca ggacgacgtc cttggttatc cggataccgg tggccaggtt gtttacatcc
      961 tggatcaagt tcgtgcccta gaggaagaaa tgcttcaccg cattaagcaa caaggactgg
     1021 atattactcc tcggattctc attatcactc ggcttcttcc agacgcggtt ggaaccacct
     1081 gtggccagcg ccttgagaaa gtttttggga ccgagtactc ccacattctt cgcgtcccct
     1141 tcagaaatga gaagggagtc gtccgcaagt ggatttcccg gttcgaggtg tggccctatt
     1201 tggaaagata cactgaggat gtcgcgagcg aacttgctgg agagttgcag ggcaagcctg
     1261 atctgatcat cggaaactac agtgatggaa acattgttgc ttccttgtta gcacataaat
     1321 taggtgttac acagtgtaca atagcccatg cccttgagaa gacgaagtac ccagagtcag
     1381 acatatactg gaagaaattt gaggaaaagt accacttctc ttgccagttc actgctgatc
     1441 tcatcgccat gaaccacacc gacttcatta tcaccagcac cttccaagaa attgctggaa
     1501 gcaaggatac agtggggcag tatgagagtc acatgaactt cactcttcct ggactctacc
     1561 gagttgtcca cgggatcgac gtcttcgacc cgaagttcaa cattgtttca ccaggtgctg
     1621 acatgagcat ctactttgct tacaccgaac aggagcggcg gttgaaatcc ttccaccctg
     1681 agatcgagga actcctcttc agcgatgttg agaacaagga acacttgtgt gtgttgaaag
     1741 ataagaagaa gcctattatt ttcaccatgg caaggctgga ccgtgtcaag aacttgacag
     1801 ggcttgttga gtggtatggc aagaactcca agttgaggga actcgccaac ttggtcgtgg
     1861 ttggaggtga caggaggaag gattcgaagg acttggaaga gcagtctgag atgaagaaaa
     1921 tgtacgacct catcgaaaag tacaagctga atggccagtt caggtggatt tcctcccaga
     1981 tgaaccgggt gaggaatgga gagctctacc gctacatctg tgacacgaag ggagtcttcg
     2041 ttcaaccggc tatctatgaa gctttcgggt tgaccgtggt tgaggccatg acttgtggat
     2101 tgccaacctt tgccacttgc aatggtggac cagctgagat cattgtgcat ggcaaatcgg
     2161 gctaccacat tgatccttac catggtgacc aggcggccga gcttcttgta gacttcttca
     2221 acaagtgcaa gattgaccag tcccactggg acgagatctc aaagggtgcc atgcagagaa
     2281 ttgaagagaa gtatacatgg aaaatatatt ctgagaggct gttgaacctg actgccgtgt
     2341 atggcttctg gaagcatgtg actaaccttg atcggcgcga gagtcgccgg taccttgaaa
     2401 tgttctatgc cctcaagtat cgcccactgg cacagtctgt tcctccggct gtcgagtaaa
     2461 caaagagaca gattgttacc agaagacgga agcattgg
//
```

EgSuSy3

```
LOCUS       DQ227994            2508 bp    mRNA    linear   PLN 29-SEP-2005
DEFINITION  Eucalyptus grandis sucrose synthase 3 (EgSuSy3) mRNA, complete cds.
ACCESSION   DQ227994
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 2508)
  AUTHORS   Zhou,H. and Myburg,A.A.
  TITLE     Cloning, characterization and expression profiling of two
            Eucalyptus sucrose synthase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2508)
  AUTHORS   Zhou,H. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2005) Genetics Department, University of
            Pretoria, Pretoria 0002, South Africa
FEATURES             Location/Qualifiers
     source          1..2508
                     /organism="Eucalyptus grandis"
                     /db_xref="taxon:71139"
                     /note="Organelle: nucleomorph"
     gene            1..2508
                     /gene="EgSuSy3"
     5'UTR           1..50
                     /gene="EgSuSy3"
     CDS             51..2468
                     /gene="EgSuSy3"
                     /note="EC 2.4.1.13"
                     /codon_start=1
                     /product="sucrose synthase (SuSy)"
                     /translation="MPERLLTRVHSLRERLDETLLAHRNDILAFLTRIEAKGKGILQH
                     HQLIAEFEAISEEHRKKLSEGAFGEILRSSQEAIVLPPWIALAVRPRPGVWEYIRVNI
                     HALVVEELQVTEFLHFKEELVNGNLNGNFVLELDFEPFTAQFPRPTLSKSIGNGVEFL
                     NRHLSAKLFHDKESLHPLLEFLQVHCYKGKNMMVNTRIQNVFSLQHVLRKAEEYLSAL
                     KPETPYSQFEHKFQEIGLERGWGDTAERVLEMIRLLLDLLEAPDPCTLENFLGRIPMV
                     FNVVIMSPHGYFAQDDVLGYPDTGGQVVYILDQVRALESEMLHRIKQQGLDITPRILI
                     VTRLLPDAVGTTCNQRLEKVFGTEYSHILRVPFRTEKGMVRKWISRFEVWPYLETYTE
                     DVANEIAGELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYW
                     KKFEEKYHFSCQFTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRV
                     VHGIDVFDPKFNIVSPGADMSIYFSYTEEKLRLKSFHAEIEELLFSDVENKEHLCVLK
                     DRNKPILFTMARLDRVKNLTGLVEWYGKNTRLRELVNLVVVGGDRRKESKDLEEQAEM
                     KKMYGLIETYNLNGQFRWISSQMNRVRNGELYRYICDMRGAFVQPALYEAFGLTVVEA
                     MTCGLPTFATCKGGPAEIIVHGKSGFHIDPYHGDQAAETLANFFEKCKVDPSHWDKIS
                     QGAMQRIKDKYTWQIYSERLLNLTAVYGFWKHVSNLDRLESRRYLEMFYALKYRKLAE
                     SVPLAVE"
     3'UTR           2469..2508
                     /gene="EgSuSy3"
BASE COUNT      649 a    581 c    648 g    630 t
```

```
ORIGIN
        1 tttcttctct tctgctttgc aagtgttaag gttcttgaga ggattcagac atgccggaac
       61 gcttgcttac gcgtgttcac agcctccgtg aacgtttgga tgagaccctc ttggctcatc
      121 gtaatgacat tttggctttc ctcaccagga tcgaagccaa aggcaagggt atcttgcagc
      181 atcatcaact tattgccgag ttcgaggcga tctctgagga acataggaaa aagctcagtg
      241 aaggagcatt tggagaaatc ttgagatcca gtcaggaagc aatagtcttg cctccatgga
      301 ttgcccttgc tgtgcgtccc aggcccggtg tgtgggagta cattagggta aatattcacg
      361 cccttgttgt tgaggagttg caagtgaccg agtttcttca tttcaaggaa gaacttgtca
      421 atggaaattt gaatggaaac tttgtgcttg aattggactt tgagccattc actgctcaat
      481 ttccccgccc aactctttca aagtccattg gaaatggagt ggagttcctt aaccgccacc
      541 tttcggcaaa gctcttccat gacaaggaga gcttgcaccc tttgcttgag ttccttcaag
      601 ttcattgcta caaaggaaag aacatgatgg tgaacacaag aattcaaaat gtcttctccc
      661 tccaacatgt cctgaggaag gcagaggagt acctctctgc cctcaaacct gagacacctt
      721 actctcagtt cgagcacaag tttcaggaaa ttgggctaga gaggggatgg ggcgacacgg
      781 ccgagcgtgt gcttgaaatg attcggcttc tgctggatct acttgaggct cccgatcctt
      841 gcaccctcga gaatttcttg ggaagaatcc ctatggtctt taacgttgtt atcatgtcgc
      901 ctcacggtta ctttgcccaa gacgatgtcc tcggatatcc tgataccggt ggccaggttg
      961 tgtacatctt ggaccaagtt cgtgcattgg agagtgagat gcttcaccgt atcaaacaac
     1021 aagggctgga tattactcct cggatcctca tcgtcactcg tcttcttcct gatgcagtag
     1081 gaaccacgtg caaccagcgt cttgaaaagg tttttggaac agaatactcc cacattcttc
     1141 gagttccctt tagaacggag aaagggatgg tccgcaaatg gatttcacga tttgaagtgt
     1201 ggccttactt ggaaacatac accgaggatg tcgccaatga aattgcagga gaactacaag
     1261 gcaagcccga tctgatcatt ggcaattaca gtgatggtaa cattgttgct tctttgctgg
     1321 cccataagtt gggcgttacg cagtgcacga tagctcatgc ccttgagaag accaaatacc
     1381 cggaatctga catctactgg aaaaagttcg aagaaaaata tcatttctct tgccagttta
     1441 cagcagatct catagccatg aaccacaccg actttataat caccagcacc ttccaggaga
     1501 ttgccggaag caaagacacg gttgggcaat atgagagtca caccgctttt acacttcctg
     1561 gactgtaccg agtcgtccat ggcattgatg ttttcgatcc caaattcaac attgtgtcac
     1621 caggagctga tatgagcatt tatttctcct acacggagga gaagctgagg ttgaaatctt
     1681 tccatgccga gatcgaagaa cttctgttca gcgatgtgga gaacaaggaa cacctatgcg
     1741 tattgaagga ccggaacaag ccaatcttgt tcaccatggc aagactggac cgtgtcaaga
     1801 acttgacggg gctggtcgag tggtacggaa agaacacccg cttgagagag ctggtcaact
     1861 tggtagtggt cggtggtgac cggaggaagg aatctaaaga cttggaagag caagccgaga
     1921 tgaagaagat gtacggactg attgagacgt acaatttgaa tggccagttc aggtggatct
     1981 cctcccagat gaaccgggtg aggaacggag agctttaccg gtacatctgt gacatgaggg
     2041 gagcatttgt gcagcccgct ctctacgagg cttttggatt gaccgtggtg gaggccatga
     2101 catgtggatt gccgaccttt gcgacttgca aaggcggtcc tgctgagatc attgtgcacg
     2161 gcaagtccgg attccacatt gatccttatc acggggatca ggcagcggag accctcgcca
     2221 acttcttcga gaagtgcaag gtggacccaa gccactggga caagatctca caggggcga
     2281 tgcaaagaat caaggataag tatacatggc aaatttactc cgagaggttg ctcaacttga
     2341 cggcggtgta cggcttctgg aagcacgtct cgaatctcga ccgtctcgag agccgccgct
     2401 acctggaaat gttctatgcc ctcaagtatc ggaaattggc cgagtcggtt ccccttgccg
     2461 ttgagtgagt tagaggtgaa caaaagggca gatctcaaga ggcagaca
//
```

B: **Composition of buffers and solutions**

5× reverse transcription buffer (Roche)

37.5 mM KCl,
1.5 mM MgCl2,
2.5 mM DTT,
25 mM Tris pH 8.3.

10× reaction buffer for Expand Taq polymerase and ExSel polymerase

500 mM KCl,
20 mM MgCl2,
100 mM Tris pH 8.3.

10× reaction buffer for Supertherm polymerase

500 mM KCl,
15 mM MgCl2,
100 mM Tris pH 8.3.

10× restriction buffer B (Roche)

1 M NaCI,
50 mM MgCl2,
10 mM 2-Mercaptoethanol,
100 mM Tris-HCl pH 8.0 at 37°C.

10× restriction buffer O (Fermentas)

10 mM MgCl2,
100 mM NaCl,
0.1 mg/ml BSA,
50 mM Tris-HCl pH 7.5 at 37°C.

LB medium

composition per litre:
10g tryptone,
5g yeast extract,
5g NaCl,
adjust pH to 7.5,
autoclave,
for LB plates, include 15g agar prior to autoclaving.

YEP medium

composition per litre:
10g peptone,
10g yeast extract,
5g NaCl,
adjust pH to 7.0-7.5,
autoclave,
for YEP plates, include 15g agar prior to autoclaving.