Integration of Arabidopsis and Poplar Model Systems to Elucidate Gene Function during Wood Formation

Ellinor Edvardsson
Faculty of Forest Sciences
Department of Forest Genetics and Plant Physiology
Umeå

Doctoral Thesis
Swedish University of Agricultural Sciences
Umeå 2010

Acta Universitatis agriculturae Sueciae

2010:31

Cover: *Wood anemone forest,* copper etching by Eva Olsson. Published with the kind permission from the artist.

ISSN 1652-6880
ISBN 978-91-576-7508-8

Tryck: Arkitektkopia, Umeå, Sweden 2010

Integration of Arabidopsis and Poplar Model Systems to Elucidate Gene Function during Wood Formation

Abstract

Wood is an essential raw material being used for a number of different applications including construction, pulp and paper production and as a source of biomass for bioenergy generation. The increasing number of sequenced plant genomes provides powerful tools to evaluate plant gene function. Cross-species integration is increasingly being used as a way to facilitate gene discovery and functional characterisation. In this thesis poplar and Arabidopsis model systems were integrated with the aim of identifying and characterising cell wall biosynthesis or modifying genes. Wood forming tissues of poplar represent an ideal model system to study developmental processes from cell division through to cell death. However, the relatively large size and slow growth rate of poplar make it practically challenging to use as an experimental system. An alternative is to identify closely related genes in Arabidopsis where their function can be analysed using T-DNA knockout lines and then transfer this knowledge back to tree species.

Transcriptomics data for the wood forming tissues of poplar was used to identify genes which were specifically upregulated in the zone of secondary wall formation. Several of these poplar genes were used to identify the closest Arabidopsis homologs in a 3-step screening filter. One of these genes was selected for further analysis in Arabidopsis resulting in the identification of a small family of 4 related sequences which were named *DUF DOMAIN PROTEINs* (*DDPs*). T-DNA knockouts were generated for each of the genes and all possible mutant combinations made. *ddp* mutants affected root elongation, lateral root formation and seed development. Analysis of the biochemical composition of the *ddp* mutant combinations revealed possible changes in the relative amount of hemicellulose, pectin and/or lignin, compounds important in the secondary cell wall.

The Arabidopsis *epc1* mutant was identified previously as having severe growth defects, with reduced cell-cell adhesion and increased secondary growth. Two poplar EPC1 homologs were identified and complementation studies show that they are able to rescue the Arabidopsis *epc1* mutant. RNAi downregulation of *PtEPC1* resulted in only minor effects on growth which may be due to residual message producing sufficient functional protein.

Keywords: Arabidopsis, poplar, model systems, wood formation, RNAi downregulation, T-DNA mutants.

Author's address: Ellinor Edvardsson, SLU, Department of Forest Genetics and Plant Physiology, SE-901 83 Umeå, Sweden
E-mail: ellinor.edvardsson@genfys.slu.se

Foooocuuuus!
-Peanut
(Jeff Dunham)

Rofl…

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I **Ubeda-Tomas, S., Edvardsson, E., Eland, C., Singh, S.K., Zadik, D., Aspeborg, H., Gorzsas, A., Teeri, T.T., Sundberg, B., Persson, P., Bennett, M. and Marchant, A.** (2007) Genomic-assisted identification of genes involved in secondary growth in Arabidopsis utilising transcript profiling of poplar wood-forming tissues. *Physiologia Plantarum*, **129**, 415-428.

II **Edvardsson E. and Marchant A**. Characterisation of the developmental role of the small Arabidopsis *DDP* gene family with no previously assigned molecular function. (Manuscript)

III **Edvardsson E. and Marchant A.** Analysis of the developmental function of poplar homologs of the Arabidopsis EPC1. (Manuscript)

Paper I reproduced with the permission of the publishers.

Abbreviations

ABA	Abscisic acid
CaMV	Cauliflower mosaic virus
CAZyme	Carbohydrate-active enzymes
CBM	Carbohydrate-binding module
CE	Carbohydrate esterase
CesA	Cellulose synthase
EPC1	Ectopically Parting Cells 1
ER	Endoplasmic reticulum
EST	Expressed sequence tag
EXT	Exostosin
FT-IR	Fourier transform infrared spectroscopy
GH	Glycoside hydrolase
GT	Glycosyltransferase
HGA	Homogalacturonan
MS	Murashige and Skoog agar
OPLS	Orthogonal projections of latent structures
PCA	Principal component analysis
PCR	Polymerase chain reaction
PL	Polysaccharide lyase
PLS-DA	Partial least square discriminant analysis
PME	Pectin methylesterase
qPCR	Quantitative real-time PCR
RG-I	Rhamnogalacturonan I
RG-II	Rhamnogalacturonan II
RT-PCR	Reverse transcription PCR
XEH	Xyloglucan hydrolase
XET	Xyloglucan endotransglycosylase

1 Introduction

The completed and ongoing sequencing of different plant genomes gives an ever increasing opportunity to study the function and development of plants in more detail. This has provided the tools to start to gain better understanding about wood development which is an essential process, both for the tree to provide rigidity, transport of nutrients and protection as well for industries that utilize trees for pulp and paper manufacture, construction and increasingly as a source of biomass and bioenergy production. The complexity of fiber and cell wall formation and regulation makes it a challenging task to unravel the biosynthesis pathways and identify the key enzymes.

Sequencing of the poplar genome together with detailed transcript profiling of poplar wood forming tissues allows novel potential genetic regulators of wood formation to be identified. However, it is time consuming to analyze the function of a large number of genes using RNAi or antisense-based knockdown techniques in poplar. An alternative approach is to first study the function of closely related putative orthologs in Arabidopsis to identify genes with apparent important function. This then allows a subset of genes to be identified which can be studied in poplar whose putative function is indicated by results obtained in Arabidopsis.

In Paper I transcript profiling data obtained for the poplar wood forming tissues was used to identify genes specifically upregulated in zones correlated with late wood development. These genes were then used to identify the closest Arabidopsis homologs. Of the different genes selected and found to be associated with secondary growth, one was *At4g27435*, a gene with no previously assigned function. Despite the efforts made in characterizing the function of all genes in Arabidopsis since its full sequence was established, approximately 30 % of the Arabidopsis genes have either only a hypothetical or unknown function (Brown et al., 2005; Clare et al., 2006). It is important

to develop ways to determine the function of these genes in order to gain a full understanding of all aspects of plant function and development. The protein encoded by the *At4g27435* gene shows homology to proteins encoded by 3 other genes (*At1g52910*, *At1g61065* and *At3g15480*), all of which have no known function. The work reported in Paper II aimed to identify the function of this small family of related Arabidopsis genes and to establish whether there is functional redundancy. In Paper III we describe the two poplar homologs of the Arabidopsis *ECTOPICALLY PARTING CELLS 1* (*EPC1*) gene. The poplar EPC1 homologs have approximately 70 % identity to the Arabidopsis EPC1 at the protein level. Arabidopsis *epc1* knockout mutants are severely affected in growth and exhibit an increase in secondary growth in hypocotyl tissues (Singh et al., 2005). We have shown that the poplar *EPC1* genes are functional orthologs of the Arabidopsis *EPC1* using complementation and experiments were carried out to elucidate the role of the *EPC1* family in *Populus*.

1.1 Wood formation

Plants grow as a result of cell divisions that occur at the shoot and root apical meristems. This process of divisions of cells results in the formation of new organs or an elaboration of the root system (primary growth) (Northington and Schneider, 1996). In woody species a further two lateral meristems, the vascular and cork cambiums function to increase the woody tissues and outer bark respectively. The activity of the vascular and cork cambium allows the trees to increase in girth (secondary growth). The activity of these cambiums varies with the season for trees that undergo a dormant phase resulting in the characteristic growth rings seen in wood. Of the two lateral meristems it is the vascular cambium that produces the secondary phloem and xylem cells and it is the secondary xylem that constitutes wood (Fig. 1). Wood provides the tree with water and nutrient transport as well as overall support. As the tree continues to grow cells begin to die resulting in sapwood. Sapwood consists of predominantly dead cells and it is this part of the wood that conducts the water and nutrient transport. During the lifespan of a tree it accumulates different compounds such as oils, gums, tannins, resins and aromatic substances that infiltrate cells in the central wood region eventually giving rise to the inner heartwood. (Northington and Schneider, 1996).

Gaining a full understanding about the differences in wood formation processes between different tree species as well as the regulation of the formation of wood fibers and cell walls would be of considerable potential benefit to the forestry industry. The possibility to tailor wood fibers could potentially lead to trees with specific end uses, either pulp and paper manufacturing or construction material, businesses with different demands on the wood properties.

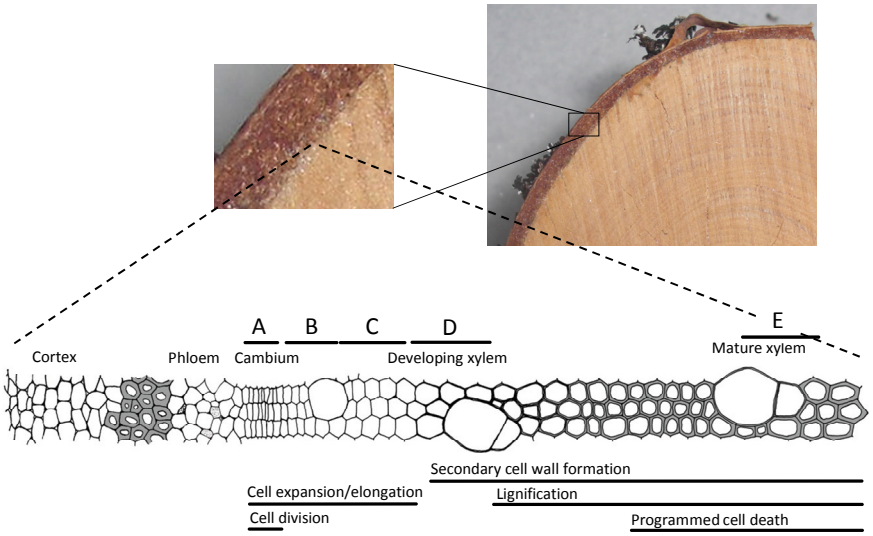

Fig. 1. Photograph showing a cross section of a birch tree trunk with the cambium region highlighted and shown in more detail. The cellular developmental stages are indicated, and A-E denotes the sample locations of microarrays conducted and referred to in this thesis. Figure modified from thesis by Takahashi Schmidt (Takahashi Schmidt, 2008). Photograph taken by E. Edvardsson.

1.2 The cell wall - glycosyltransferases

Cell walls perform a vital structural function providing strength to the plant as well as performing a protective function (Shedletzky et al., 1992; Braam, 1999; Vorwerk et al., 2004), regulating cell volume and shape (Delmer and Amor, 1995) and mediating cell-cell interactions (Karp, 1999). The primary cell wall is comprised predominantly of cellulose microfibrils which are crosslinked via hemicellulose and embedded in a gel-like matrix made of pectin (Fig. 2). Cellulose constitutes roughly 25 % of the dry weight of the primary cell wall. The cellulose microfibrils are made exclusively of glucose monomers joined together via β(1→4) linkages into linear molecules which provide resistance to tensile or pulling forces. In most cases the microfibrils of one layer are oriented at approximately 90° to those of adjacent layers (Zablackis et al., 1995; Northington and Schneider, 1996; Reiter, 1998; Karp, 1999).

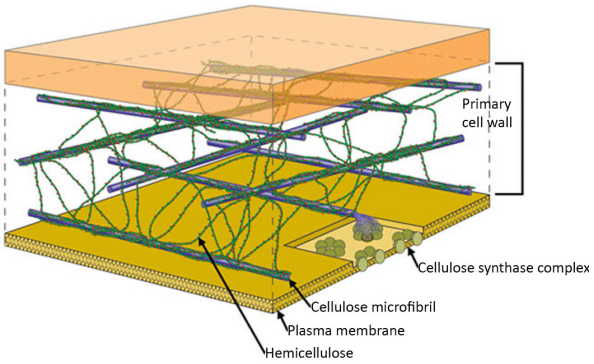

Fig. 2. Schematic overview of the primary cell wall, depicting the cellulose synthase complex generating cellulose microfibrils which gets crosslinked into a complex network together with hemicelluloses surrounded by pectins (not shown). Figure credit United States Department of Energy Genome Programs (http://genomics.energy.gov)

The hemicelluloses are polysaccharides that bind to the surface of cellulose microfibrils, cross-linking them into a complex structural network. There are many classes of hemicellulose and they all have a long linear backbone of only one type of sugar (glucose, xylose, mannose or galactose) from which shorter side chains of other sugars protrude (Karp, 1999).

Xyloglucans are one member of the hemicelluloses and it may account for up to 20 % of the dry weight of the primary wall. Xyloglucan has been hypothesized to be a major factor in controlling cell wall expansion rate, and thereby regulating plant growth (Zablackis et al., 1995; Takeda et al., 2002).

The pectins are negatively charged polysaccharides containing galacturonic acid forming a hydrated gel filling in the space between the fibrous elements. There are three main forms of pectin found in the cell wall, namely homogalacturonan (HGA), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II). HGA is a linear chain of galacturonic acids (GalA) that may be methyl esterified and/or acetylated and under the right conditions two homogalacturonans can bind Ca^{2+} resulting in gelformation. The RG-I backbone consists of alternating GalA and rhamnose residues. The rhamnose residues can be substituted with galactan and arabinan rich side chains which in turn can be methylated or acetylated. RG-II is the most complex plant polysaccharide identified to date and consists of a galacturonic acid backbone which is substituted by four different side chains comprising 20 different monosaccharides, some of which are only found in RG-II. One of the RG-II sidechains contains an apiose residue which is involved in forming a borate-diol ester crosslink between two RG-II polymers which influences the structure and properties of the pectin. Pectin is believed to account for almost a third of the primary cell wall macromolecules and is also involved in the plant defense through the release of signaling molecules or oligosaccharins from the wall when a cell is attacked by pathogens (Shibuya and Minami, 2001; Willats et al., 2001; Osorio et al., 2008).

The structural proteins found in the cell wall make up between 1-10 % of the dry weight of primary walls. They do not appear to have any enzymatic function and become immobilized by covalent cross-linking to other cell wall proteins. They are classified on their bases as glycine-rich proteins (GRPs), proline-rich proteins (PRPs)s and hydroxyproline-rich proteins (HRGPs) (Reiter, 1998; Karp, 1999).

Newly divided cells have primary walls that are extensible allowing for expansion. Cell wall loosening is a requirement for expansion and several enzymes have been implicated in this process. One group of enzymes are the pectin methylesterases (PMEs). The PMEs demethylesterify HGA chains and thereby initiate the binding of the exposed carboxylate ions to Ca^{2+} resulting in crosslinking and stiffening of the pectate network (Fig. 3b). The result of PME action can in some cases stimulate the activity of cell wall hydrolases such as polygalacturonases and pectate lyases resulting in cell wall loosening (Fig. 3a) (Bosh and Hepler, 2005).

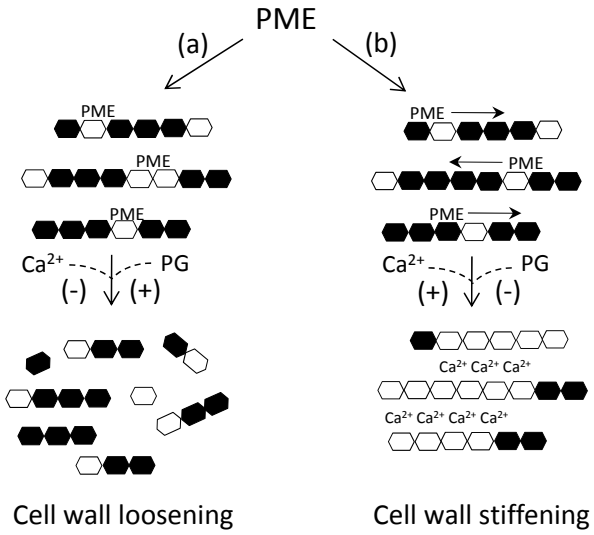

Fig. 3. PME actions in the cell wall; depending on the wall properties it can act in a random fashion (a) promoting hydrolases and resulting in cell wall loosening or act in a blockwise manner (b), resulting in stretches of carboxyl groups that can react with Ca^{2+} and hence cause stiffening of the wall. PME: pectin methylesterase, PG: polygalacturonase, Ca^{2+}: calcium, black blocks represents methylesterified galacturonic acid and white blocks are demethylesterified GalA. Adapted from (Micheli, 2001).

Another important group of cell wall modification enzymes is the xyloglucan endotransglucosylase/hydrolase (XTH) family consisting of xyloglucan endotransglucosylases (XETs) and xyloglucan hydrolases (XEHs) which can cut and rejoin xyloglucan chains, a feature necessary for the expansion of cell walls (Rose et al., 2002).

Most mature cells lack the extension capacity due to the formation of a inner rigid secondary cell wall (Turner et al., 2001). The secondary wall is important due to its strengthening function as well as being involved in water conduction for certain cell types. This transformation from having only a primary wall to consisting also of a secondary wall occurs when the cell stops growing and increases its content of aligned and crystalline cellulose and different hemicelluloses (Ebringerova and Heinze, 2000) and usually it incorporates a phenol-containing polymer called lignin. Secondary walls contain more cellulose than primary walls and pectin is not always present making cells with secondary walls rigid. Secondary walls are usually deposited in three distinct layers (S_1, S_2 and S_3) and each layer differs in the

orientation of their cellulose microfibrils. This contributes to the increased strength of the cell wall. Lignin provides the cell with structural support by cross-linking cellulose and hemicellulose and it is also the second major component of wood. It is the hydrophobic property of lignin that makes it possible for certain cell types to play a role in the water transport of plants. It is believed that lignin biosynthesis was a crucial development step in order for plants to live on land and the complete lignin biosyhtesis pathway seems to have appeared in mosses first (Xu et al., 2009).

It becomes obvious that the cell wall is important when looking at the number of genes involved in the synthesis and destruction of components from the wall. In Arabidopsis it is predicted that over 700 genes, or 2.8 % of the total coding regions, encode carbohydrate-active enzymes (CAZyme) and that is 2x more than in *C. elegans* and *D. melanogaster* (Henrissat et al., 2001). In a woody species such as poplar the number of carbohydrate-active enzymes is even greater having almost the double compared to Arabidopsis reflecting the differences between herbaceous and woody plant species (Geisler-Lee et al., 2006). There are different classes of carbohydrate-active enzymes such as glycoside hydrolases (GHs), the polysaccharide lyases (PLs), carbohydrate esterases (CEs) and glycosyltransferases (GTs) (http://www.cazy.org/; Geisler-Lee et al., 2006; Cantarel et al., 2009). GTs have a central role in polysaccharide synthesis by catalyzing the transfer of a sugar moiety from an activated donor sugar onto saccharide and nonsaccharide acceptors creating a glycosidic bond. Based on their predicted protein sequence the GTs have been divided into 90 families (Campbell et al., 1997; Henrissat et al., 2001; Coutinho et al., 2003) and the GTs either catalyse a group transfer with a net inversion, forming a β glycosidic linkage, or a net retention, forming an α linkage (Coutinho et al., 2003; Lairson et al., 2008). The number of GT mutants with cell wall defects highlights the importance of GTs in the function and maintenance of specific cell wall components. For example, a number of different CesA mutants, belonging to the GT family 2, have been found to affect cellulose synthesis in either the primary or secondary cell walls (Taylor, 2008). It has been demonstrated that AtCESA1, 3 and 6 are required for primary cell wall synthesis and mutations in these genes can range from embryo-lethal to less severe growth defects (Fagard et al., 2000; Persson et al., 2007). Several AtCESAs have been proven to show redundancy, at different growth stages AtCESA2, 5 and 9 are partially redundant with AtCES6 (Desprez et al., 2007; Persson et al., 2007). Mutations affecting the secondary cell wall synthesis due to collapsed xylem has been indentified and were named irregular xylem mutants (irx) (Turner and Somerville, 1997) and *irx1*, *irx3* and *irx5* are caused by defects in AtCESA8, 7 and 4 respectively (Taylor et al., 2003).

Other mutations in AtCESA 7 and 8 result in mutants with reduced strength in their interfascicular fibers, named fragile fiber 5 and 6 (Zhong et al., 2003). Other Arabidopsis GT mutants such as *quasimodo 1* (*qua1*) mutant belonging to the GT8 family and *qua2* which is a putative methyl transferase, have also been identified which are defective in synthesis or modification of homogalacturonan resulting in a defect in cell adhesion (Bouton et al., 2002; Mouille et al., 2007). Another GT8 mutant, in the PARVUS gene, displays effects in the assembly of hemicellulosic xylan. A xylan decrease can also been seen in mutants from other GT families namely *irx9* and *irx14* (GT43) and *irx10/irx10-L* double mutant and *fra8* (GT47). The *irx9*, *irx10* and *irx14* mutations cause a disruption in the xylan backbone chain elongation while *fra8* affects synthesis of the reducing end tetrasaccharide of glucuronoxylan (Lee et al., 2007; Brown et al., 2009; Lee et al., 2009; Wu et al., 2009).

1.3 Model systems

Arabidopsis was the first plant to have its genome fully sequenced in 2000 which confirmed its status as an excellent model plant system. Since then a number of other plant models have emerged, including poplar, tobacco, rice and moss thus providing researchers with a range of diverse species with different developmental characteristics. Attributes that are useful for a model plant are ease of cultivation, ability to be transformed, a sequenced genome and preferably a short generation time. Although Arabidopsis represents an excellent model system for most aspects of plant development it is not ideal for studies into processes such as fruit ripening or wood formation. Therefore it may be necessary to adopt alternative model systems or to integrate the use of two or more model species in a single study. The work in this thesis made use of both the poplar and Arabidopsis model systems to study the genetic regulation of wood development.

1.3.1 Arabidopsis thaliana

Arabidopsis thaliana is a small annual plant with a relative short generation time of about 6-8 weeks and is easy to transform. Arabidopsis was the first plant genome to be fully sequenced revealing approximately 25 500 genes distributed on 5 chromosomes (The Arabidopsis Genome Initiative (2000)). A major advantage of Arabidopsis is the availability of a large number of

transfer DNA (T-DNA) mutant insertion lines which allow for knockout lines to be obtained for the majority of genes. In Papers I and II both SALK and SAIL T-DNA lines were utilised (Sessions et al., 2002; Alonso et al., 2003). Thus Arabidopsis is an ideal model organism to screen for the function of large numbers of genes of interest. Under normal growth conditions Arabidopsis does not form extensive secondary growth, although it is still possible to induce this feature when grown under short day conditions. When grown under these more special conditions the hypocotyl undergoes increased secondary growth and develops both vascular and cork cambiums, making it resemble woody species such as poplar (Chaffey et al., 2002). However, Arabidopsis does not form parenchyma rays therefore limiting its usefulness in studying genes involved in wood formation (Nieminen et al., 2004). Thus in order to fully evaluate gene function in wood development it is important to conduct studies in a true wood species such as poplar.

1.3.2 Populus

Poplar has been shown to represent a good model system to study the formation of wood in part due to the distinct ontology of cellular development within the cambium from the meristem through to fully lignified cells. Transcript profiling of the different developmental zones has identified genes which are up or downregulated during different specific developmental stages of wood development (Hertzberg et al., 2001; Schrader et al., 2004; Moreau et al., 2009). The findings of these transcript profiling studies have been collated in the PopGenIE database (Sjodin et al., 2009). Sequencing of the poplar genome has further increased its potential as a model system for wood development. Compared to other tree species, poplar has a relatively small genome size of around 480 Mbp coding for approximately 45 500 genes distributed on 19 chromosomes (Tuskan et al., 2006). The development of a transformation system for poplar (Nilsson, 1992) made it possible to introduce constructs to overexpress or downregulate expression of specific genes allowing stable transgenic lines to be generated to study various aspects of development. Compared to other tree species such as pine or spruce, *Populus* undergoes rapid growth often making it the tree of choice for studies into wood formation.

2 Main objectives

The first part of my work has been on a set of genes with no previously designated function. In Paper I (Ubeda-Tomas et al., 2007) poplar transcription data combined with different selection filters were used with the aim to more rapidly gain insight to poplar gene function by the use of Arabidopsis homologs. One of these Arabidopsis genes was found to be part of a small four member family which has not previously been assigned a putative function. These genes were investigated further in Paper II with the focus on understanding their function and potential role in plant development and growth.

The second part of my work focused on characterizing poplar homologs of the Arabidopsis *EPC1* gene. Arabidopsis *epc1* knockout mutants have reduced cell-cell adhesion and this is hypothesised to be a contributing factor in the increased secondary growth observed in hypocotyl tissues (Singh *et al.* 2005). This observation indicates that tree homologs of the Arabidopsis *EPC1* gene may provide a useful genetic tool to modulate wood biomass formation. The possibility to influence the secondary growth or wood production in trees would have a significant impact for the forestry industry, potentially generating trees with an increased wood formation rate. The aim of the project reported in Paper III was to elucidate the role of PtEPC1 by the use of RNAi and overexpression lines.

3 Overview of methods used

3.1 Gene identification

In Paper I transcript profiling was utilized to select genes expressed during the secondary growth phase in poplar. Microarray data sets generated from previous studies (Hertzberg et al., 2001; Schrader et al., 2004; Aspeborg et al., 2005) were analysed to identify ESTs highly expressed in the D zone representing the region of secondary cell wall formation. In total 150 EST sequences were identified and used to find full length sequences from the poplar genomic database. Due to its relative large physical size and slow growth, poplar presents some practical problems when it comes to analyzing a large set of genes in detail. Therefore Arabidopsis was used to identify the function of the closest homologs of the poplar genes of interest with the aim of transferring findings back to poplar for a selected subset of the genes. In order to ensure the highest likelihood of this approach being successful it was important to place stringent criteria on the selection of Arabidopsis homologs of the poplar genes by use of three selection filters. The first filter identified putative Arabidopsis homologs by tBlastX searches against the Arabidopsis genome, using the predicted full length poplar cDNAs. Since the aim was to identify genes involved in secondary cell wall formation the second filter selected Arabidopsis genes expressed in secondary thickened hypocotyls due to its ability to undergo secondary growth and hence resemble wood formation to some extent (Chaffey et al., 2002). Despite these two filters in some cases there was still two or more related genes which passed the first selection filters. In these cases the expression levels in other tissues were also taken into account and genes were selected which were highly expressed in the secondary thickened hypocotyl and had a low expression level in other tissues. It was reasoned that these genes were most

likely to have a function which was specifically related to secondary cell wall formation. Three poplar ESTs were selected in order to validate the selection design developed and potential orthologs in Arabidopsis were found and investigated in more detail. One of these Arabidopsis genes encodes a predicted protein of 173 amino acids with an N-terminal signal peptide sequence and three transmembrane spanning regions. The protein is annotated as having no known function. Blast searches of the protein sequence against the Arabidopsis database revealed three further related proteins also annotated as having no known function. Paper II reports the work done to characterize this small family of four Arabidopsis genes in more detail.

In Paper III we examined the poplar EPC family with the focus on the closest homologs to the Arabidopsis *EPC1* gene. The distinct developmental ontology found in the cambium of poplar represents an ideal tissue to identify important genes involved in different stages of wood formation. Previous work utilized the transcription profiling data set generated for specific developmental zones (Hertzberg et al., 2001) to identify an EST for a glycosyltransferase from family GT64 (Singh et al., 2005). Blast searches against the Arabidopsis database using the poplar GT64 EST revealed high homology to *At3g55830* which was subsequently named *Ectopically Parting Cells 1* (*EPC1*). The Arabidopsis *epc1* mutant has been reported to display a severe growth phenotype, cell adhesion loss (Singh et al., 2005), hypersensitivity to ABA and a reduction in the level of galactan (Bown et al., 2007). It was found that there was an increase in secondary growth associated with the reduction in cortical cell adhesion in the *epc1* hypocotyls and this resulted from an increase in the number of cells rather than cell size. This link between EPC1 and secondary growth in Arabidopsis indicated that the poplar homolog(s) may also play a critical function in regulating the development of woody tissue prompting the work described in Paper III looking at the effect of *PtEPC1* downregulation on development.

3.2 Forward and reverse genetics

There exist different ways in order to study gene function. One way is to generate mutants which have either a loss or reduced gene activity or specific genes can be induced to be highly expressed throughout the plant or in certain tissues. Forward genetics is the generation of randomly generated mutants most commonly created by insertion of T-DNA (transfer DNA) or

EMS (ethyl methanesulfonate) mutagenesis allowing any resulting phenotypes to eventually be associated with a specific gene defect. The T-DNA comes from part of the Ti (tumor-inducing) plasmid of the bacterium *Agrobacterium tumafaciens*. The T-DNA sequence can be transferred from the bacterium and stably inserted into the plant genome, a property which is exploited by researchers to introduce specific transgenes. Inclusion of an antibiotic resistance marker allows for simple selection of the transformed plants (Azpiroz-Leehan and Feldmann, 1997). EMS mutagenesis typically results in single base pair changes, mispairing, which can give rise to missense (amino acid change), or nonsense (premature stop codons) mutations as well as splicing defects. In Paper I and II, specific Arabidopsis T-DNA insertion lines from the SALK and SAIL populations mutants were utilized (Sessions et al., 2002; Alonso et al., 2003).

With reverse genetics the aim is to identify phenotypes arising from mutation or manipulation of the expression of specific targeted genes. In Paper III reverse genetics in the form of RNAi (interference) and gene overexpression was used in order to change the *EPC1* expression level. The method of silencing genes using RNAi was developed in the late 1990's for *Caenorhabditis elegans* (Fire et al., 1998) and has since been used successfully in a variety of different organisms including plants. RNAi makes use of double stranded hairpin that targets specific mRNAs for degradation. The RNAi vector pK7GWIWG2(I) compatible with the GATEWAY system was used in Paper III for a rapid and easy way to generate the necessary constructs (Karimi et al., 2002).

Another way to study the impact of alterations in gene expression is to overexpress genes typically by placing them under the control of the 35S cauliflower mosaic virus promoter (35S CaMV), an approach that was used in Paper III. The CaMV genome was sequenced 30 years ago (Franck et al., 1980) and the 35S and 19S promoters were the first promoters identified (Covey et al., 1981; Odell et al., 1985). The possibility to constitutively overexpress genes has since then been widely used in plant molecular genetics.

3.3 Plant transformation

Genes can be introduced into a plant's genome in different ways. The most commonly used method for Arabidopsis transformation is floral dip whereby the floral tissues of the plant are immersed into an *Agrobacterium* suspension transformed with the construct of interest (Clough and Bent, 1998). After transformation, seeds need to be screened typically using a selective agent such as an antibiotic, to find plants that have integrated the desired construct in their genome.

The introduction of genes into poplar involves a different approach which takes longer compared to Arabidopsis. Small pieces of petiole or stem tissue are soaked in an *Agrobacterium* solution containing the desired construct and then placed on plates with growth hormones (indole-3-butyric acid (IBA), 6-benzylaminopurine (BAP) and thidiazuron (TDZ)) and selective antibiotics (claforan and kanamycin or hygromycin) (Nilsson, 1992). After callus has formed it is transferred to media promoting shoot elongation, containing IBA and BAP. Elongated shoots are cut and transferred to media not supplemented with any hormones to allow for root initiation. When enough rooted material has been generated the plants can be transferred to soil and propagated in a controlled environment.

3.4 Polymerase Chain Reactions – RT-PCR and qPCR

In order to investigate the gene expression level in transgenic lines both reverse transcription PCR and real-time PCR were used. In this thesis I have adopted the nomenclature suggested by Bustin *et al.* where reverse transcription PCR is denoted RT-PCR and quantitative real-time PCR is named qPCR (Bustin et al., 2009). The RT-PCR technique can be used as a way to semi-quantitatively determine gene expression and was utilized in both Papers I and II. Total RNA or mRNA is extracted and reverse transcribed into cDNA using reverse transcriptase. The cDNA is then used as template in the following PCR reaction and the products are separated on an agarose gel for analysis. RT-PCR is a useful method to determine whether a gene is transcribed in a particular tissue or at a specific timepoint but does not provide reliable quantitative information. The qPCR method allows detection of the amplified DNA in real time by monitoring the

incorporation of fluorophores (SYBRgreen was used in Paper I and III). It is then possible to calculate the relative gene expression more accurately even of genes with a weak expression and also to discern between closely related genes (Czechowski et al., 2004).

In order to obtain reliable qPCR data it is important to choose reference genes that are stably expressed in the tissue tested. This is an important issue that has been extensively highlighted in animal, yeast and bacterial research but has not been so widely reported to date in plant systems (Gutierrez et al., 2008). The possibility to choose a good reference gene for any particular experiment has recently been aided by the development of the geNorm software (Vandesompele et al., 2002) (http://medgen.ugent.be/~jvdesomp/genorm/). geNorm is an algorithm based program designed to determine the most stable reference gene from a group of chosen genes. The geNorm software was utilised in Paper II and of the tested reference genes UBQ-L and GADPH both turned out to be stably expressed and were used in the study. In Paper I the reference gene used was GADPH and in Paper III both GADPH and 18S were used.

3.5 Fourier Transform Infrared Spectroscopy

FT-IR is a rapid and non-invasive method that can detect a range of functional groups in complex samples. Almost no compounds have identical Infra-Red (IR)-signatures apart from stereoisomers. Different compounds absorb at different wavenumbers and can therefore be identified by IR spectroscopy. FT-IR can be a useful tool for identification and classification of unknown samples and can assist when looking for small biochemical changes in a range of different materials (Morikawa et al., 1978; Morikawa and Senda, 1978; Williams and Fleming, 1980). In the IR spectrometer, based on the optical design of a Michelson interferometer, an infrared source emits radiation with a range of frequencies of 400 to 4000 wavenumbers measured in units of cm^{-1}, which are passed through the sample. Specific chemical bonds in the sample will thus absorb radiation of specific frequencies from the beam, which passes to a detector (Williams and Fleming, 1980). Spectra collected in Papers I-III were loaded into OPUS (version 5, 2003, Bruker Optik GmbH) and baseline corrected and normalized before being exported to Excel. The raw data was processed in SIMCA-P multivariate analysis software version 10.5 or 12.0 (Umetrics AB, Umeå, Sweden).

Depending on the aim of the particular experiment, there are different ways to prepare materials for analysis. In both Papers I and II cryosectioned Arabidopsis hypocotyls were analyzed in order to obtain information about the composition of the intact secondary growth tissue. The poplar stem tissues analyzed in Paper III were freeze dried and ball-milled prior to FT-IR measurements meaning that a homogenous mix of all different cell types was analysed rather than just the secondary xylem cells. It has been reported that the physical effect of ball-milling can give rise to artifactual spectral differences (Schwanninger et al., 2004). In Paper III the ball-milling time was limited to 70 seconds to try and reduce these artifacts.

3.6 Statistical analyses

In the 1970's Wold developed a method of statistical analysis that is now known as chemometrics, namely the field of applying mathematical information to interpret both chemical and biological data. The vast data generated through methods such as FT-IR can be analyzed with different multivariate methods including Principle Component Analysis (PCA) (Wold, 1976), Partial Least Square-Discriminant Analysis (PLS-DA) and Orthogonal Projections to Latent Structures (OPLS) (Trygg and Wold, 2002). The multivariate methods are used to interpret data with many variables and enabling the essential data to be presented in plots. The SIMCA software (Soft Independent Modelling of Class Analogies) can be used to visualize information obtained from multivariate analysis methods. The SIMCA-P (Umetrics AB, Umeå, Sweden) versions used in this thesis were 10.5 (Paper I) and 12.0 (Paper II and III).

In Paper II a vast number of root measurements were made and in order to test the hypothesis if any significant difference in growth pattern exists between mutants and wildtype, the data was analyzed using a linear model in the software called R.

4 Results and discussion

4.1 Screening methods to identify potentially important genes involved in secondary growth (paper I)

The sequencing of an increasing number of plant genomes is providing a wealth of information which can be utilised to identify key regulators of cell wall formation and wood development. During the course of evolution important key genes have been maintained and this is evidenced by the high degree of conservation seen across species between certain genes. The process of secondary growth in plants is an important feature, yet despite this, many essential genetic components remain uncharacterised. In Paper I the focus was on poplar genes identified using transcript profiling data as being differentially expressed in the secondary cell wall forming zone (D zone) and by the use of Arabidopsis T-DNA mutants evaluate possible functions for the poplar genes.

By adapting three selection filters potential Arabidopsis orthologs to three selected poplar cDNAs, *PttGT2A*, *PttGT47A* and *PttUnkA*, were found. In the first selection potential Arabidopsis homologs to poplar genes with an expression in the D zone were identified through BlastX searches. The second filter was based on transcription profiling data from Arabidopsis hypocotyls grown under short day conditions to further narrow down potential candidates to genes associated with secondary growth in Arabidopsis. In cases where more than one candidate Arabidopsis gene was found a third selection filter was applied in which expression pattern from other tissues were also analysed in order to select for genes specifically associated with the secondary growth.

4.1.1 Integrating poplar and Arabidopsis model systems – a way to identify and characterise key genes?

The three poplar cDNAs were selected in order to validate this three stage selection filter design and potential Arabidopsis orthologs were found by the use of the selection filters and corresponding T-DNA lines were obtained. The first of the three genes was *AtCSLA2*, a potential ortholog of *PttGT2A* belonging to the cellulose-synthase-like family of genes. The second poplar gene selected was homologous to *At1g27440* (now known as *IRX10*) and the third was related to *At4g27435*, a gene encoding a protein with no known function but shown to be co-expressed with secondary cell wall-associated CESA genes (Brown et al., 2005; Persson et al., 2005). T-DNA lines were obtained for the candidate genes and the initial tests revealed no visible morphological defect. Despite the lack of clear visible phenotypes in the single mutants it was possible that there were more subtle biochemical differences. To investigate this further hypocotyl sections were subjected to FT-IR analysis and compared to the wild type. Although the homozygous insertion lines for *At4g27435* and *Atcsla2* did not show any difference to wild type samples in the secondary thickened hypocotyls it does not exclude the possibility of alterations in other tissues or in compounds not detected by FT-IR. In contrast, the FT-IR result for the *At1g27440* (*Atirx10*) homozygous line did show a clear separation from wild type samples, with alterations suggesting an increase in the amount of cellulose and/or hemicellulose.

Among the plausible explanations for a lack of a visible phenotype in the *At4g27435* and *AtcslA2* single mutants is the possibility of genetic redundancy where other genes are able to perform the same or similar functions. It is also possible that the genes play only minor roles in plant development. For the *At4g27435* gene, the possibility of redundancy was investigated further by identifying 3 additional related Arabidopsis genes and generating all possible single and combinatorial mutants in Paper II.

4.2 Combinatorial knockouts of a small gene family of unknown function (paper II)

The *At4g27435* gene encoding a protein with no known function and shown to be associated with secondary growth in Paper I was analysed further in Paper II together with three additional related genes all with no known function.

Specific signal peptides mostly located in the N-terminal domain can direct proteins to different compartments and are transported via the endoplasmic reticulum and Golgi. Some proteins are incorporated into the plasma membrane through transmembrane spanning regions that function as an anchor while other proteins lack an anchor and function elsewhere (Karp, 1999). All four Arabidopsis genes investigated in Paper II encode proteins with three predicted transmembrane spanning regions and an N-terminal signal peptide. This indicates that the proteins may be transported through the secretory pathway and potentially be incorporated into the plasmamembrane. All of the four selected Arabidopsis genes lack any assigned function and they all also contain a Domain of Unknown Function (DUF) number 1218, and so were collectively named DUF1218 Domain Proteins (DDPs).

4.2.1 Knockout of the *DDPs* reveal only minor growth alterations

Homozygous T-DNA single knockout lines (*ddp1, ddp2, ddp3* and *ddp4*) were obtained and grown on soil for total height to be measured and overall development to be observed. The knockout lines were also grown on MS agar plates in order to look at root and hypocotyl length and lateral root development. The single mutant lines did not reveal any severe effects and therefore all possible knockout combinations were made to establish whether there is redundancy between the members of the family. However, only minor phenotypes were observed in the combinatorial mutants including the quadruple where all four genes were mutated. For soil grown plants a shorter inflorescence was recorded for the quadruple mutant. When grown on MS agar plates a shorter root length was recorded for two single mutants (*ddp1* and *ddp2*), two double mutants (*ddp1,3, ddp2,3*), three triple mutants (*ddp1,2,3, ddp1,2,4,* and *ddp1,3,4*) as well as the quadruple mutant (*ddp1,2,3,4*). In addition more lateral roots were formed by one single mutant (*ddp2*), four double mutants (*ddp1,2, ddp1,3, ddp1,4* and *ddp2,3*), three triples (*ddp1,2,3, ddp1,2,4* and *ddp2,3,4*) and the quadruple combination (*ddp1,2,3,4*). These small changes in height and root development did not reveal any clear indications about the function of the genes. In order to determine if any biochemical differences had occurred FT-IR was used on short day grown hypocotyls. A gradation of differences could be seen with the FT-IR and the differences are more pronounced the more genes that were knocked out with the clearest effects in the triple and quadruple mutants. The FT-IR differences are indicative of changes in the

relative amount of hemicellulose, pectin and/or lignin though further work is required to validate these observations.

The work conducted to date is not conclusive as to the precise role of the DDP family and further investigations are needed in order to gain a complete understanding of their developmental role in Arabidopsis.

4.3 Function of PtEPC1 during wood development (paper III)

This Paper addressed the question of the role of EPC1 in poplar plants as a continuation of the work conducted in Arabidopsis (Singh et al., 2005). The cambium of poplar represents an ideal tissue for studying GTs important for development due to its large size and the vast information that has been generated from transcription studies of the wood forming tissues. Singh and coworkers searched these transcription data sets for GTs upregulated in the A, B and C zones correlating to genes potentially involved in the primary cell wall biosynthesis. The work identified a poplar EST (AI162408) and comparisons against the Arabidopsis database indentified *At3g55830 (EPC1)* as the gene with the highest homology. *EPC1*-related genes are found in a wide range of species, both plant and non-plant, such as rice and moss as well as frog, rodents and humans. Of the five related human homologs, four are encoded by bi-modular proteins containing both a GT47 and a GT64 domain, and defects in these proteins cause hereditary multiple exostosis (HME) an autosomal dominant disorder (Lind et al., 1998; McCormick et al., 1998). In contrast, the plant EPC1 consists of only a GT64 domain. This conservation of EPC1 across such a wide range of different species and the severe effects seen in Arabidopsis indicates a highly conserved and important function for this gene. The mutant *Atepc1* displayed severe growth defects in the formation of the inflorescence, leaves, and roots and is virtually infertile only rarely forming 3 or 4 seeds per plant. It was also established in the Singh et al. study that the mutant plants displayed a reduced cell adhesion within the hypocotyl and leaf tissues but this was not seen in a later study (Bown et al., 2007). In Paper III we chose to investigate the role of EPC1 directly in poplar utilising overexpression and RNAi knock down approaches.

4.3.1 *Ptepc1* knockdown lines do not display the same severe developmental defects as *Atepc1*

Transcript profiling data from 2004 (Schrader et al., 2004) and 2009 (Moreau et al., 2009) revealed that PtEPC1 was associated with the later developing zones corresponding to the secondary cell wall formation zone in contrast to the earlier findings of Hertzberg and coworkers (Hertzberg et al., 2001). The fact that *PtEPC1* is able to fully complement the Arabidopsis *epc1* mutant demonstrates that both proteins are likely to perform the same function. However, despite having confirmed altered expression levels in transformed hybrid aspen trees, no severe phenotype was observed in any of the overexpression or RNAi lines.

The overexpression lines were grown for 8 weeks in a greenhouse with weekly recordings of height, internode diameter and leaf development. No severe growth defects were detected and there was no separation of the lines compared to wild type samples after FT-IR screening.

The RNAi lines were also grown on soil for 8 weeks with measurements made of total height of the plants and the diameter of internode 19, counted from the apex. Two lines displayed an increase in overall height compared to the wild type controls but no difference in respect to internode diameter or leaf development was found. The length and width of fibers were measured for samples taken from the basal part of the stem, approximately 30 cm above the soil, in order to establish whether there were any developmental effects at the single cell level. One line showed shorter fibers while two other lines displayed an increase in width of individual fibers.

No striking difference was found in respect to total height, diameter or leaf morphology in the RNAi lines, however this does not rule out the possibility that smaller chemical differences still have occurred. FT-IR is a tool that can provide information about small biochemical changes occurring in different samples and this approach was chosen for the greenhouse grown plants. Samples were collected from the basal part of the stem and stored in a freezer. Samples were ultimately freeze-dried and ball-milled for 70 sec before FT-IR measurements. The FT-IR analysis was done in SIMCA 12.0 and when including all lines revealed a Q2 value of -0.645 making it impossible to draw any conclusions since the Q2 value is negative. The predictive ability (Q2) of a model is based on cross-validation and a Q2 value of 1 is a "perfect" model and a model with Q2=0 means there is no predictive ability. However, looking more closely at the lines it is clear that two of the lines cluster closer to the wild type T89 clones and only one of these showed a small increase in fiber width while the other had no recorded difference. By excluding both of these 'wild type' appearing lines

more meaningful data could be gained resulting in a Q2 value of 0.331. This is still a low Q2 value making a prediction of the data suboptimal but more reliable. Thus, analysis of the RNAi lines can only be interpreted as indicative for two of the four lines due to this low predictability. The FT-IR indicates that RNAi lines 14 and 15 may have an elevated lignin and absorbed water content compared to the wild type T89 lines which have proportionally more cellulose, hemicellulose and/or pectin.

For the transformed hybrid aspen RNAi lines there was still a low level of residual message present (2-17 %) making it possible that this is sufficient to produce enough functional protein to give a wild type phenotype. In contrast, the use of T-DNA insertion in Arabidopsis results in a total knockdown of gene expression. This renders the plant unable to transcribe even the smallest amount of protein circumventing the possibility of residual message. It has not yet been possible to test if the degree of poplar *EPC1* gene downregulation correlates with a comparable decrease in protein level due to the lack of EPC1 antibodies. The relatively short growth time of transformed plants could also be a reason for not seeing a phenotype or it might be possible that growing the plants in suboptimal conditions or field trials could uncover phenotypes that are not evident when grown under near optimal greenhouse conditions. The findings in Arabidopsis of different levels of glucose and β-(1-4) galactan and indications from FT-IR analysis of changes in sugar containing polymers in RNAi hybrid aspen lines suggests changes in the composition or synthesis of cell wall components. Possible affected components dependent on galactose are extensins, arabinogalactan proteins and Rhamnogalacturonan I sidechains.

5 Summary and future perspectives

The sequencing of genomes of important plants provides essential and valuable tools in the quest to establish the function of genes and to understand plant development. The possibility to tailor wood fibers to specific end uses is something with great appeal to the forest industry but in order to investigate all aspects of wood development a vast number of genes needs to be characterized, both ones with proposed functions as well as genes with previously unknown functions.

The techniques available to alter gene expression, such as RNAi downregulation, 35S overexpression and T-DNA knockouts provide different ways to study the functions of specific genes. Despite these advantages a distinct phenotype is not always guaranteed making it sometimes hard to assign functions to specific genes. A lack of visible phenotype could have different reasons. There could be a redundancy issue where related genes can take over or mask the effect of the targeted gene. Functional redundancy in Arabidopsis has been characterized, reviewed in 2006 (Briggs et al., 2006) and with wood formation being crucial for trees it is likely a similar regulatory machinery exist in poplar. Recent work has indicated possible redundant poplar genes (Zhang et al., 2010; Zhong et al., 2010). It is also possible that knockout of certain genes does not give rise to a viable plant and hence only plants with enough transcribed protein for normal development are recovered from the transformation procedure. A low level of residual message in transformed RNAi lines, seen in Paper III, could be sufficient to allow for normal development. It is not always guaranteed that the degree of message downregulation correlates with a similar decrease in protein levels and for EPC1 it is not currently possible to test the level of EPC1 protein in poplar due to a lack of antibodies. The findings in Arabidopsis of different levels of glucose and β-(1-4) galactan and the FT-IR based indication of changes in sugar containing polymers in the

RNAi lines suggests changes in the composition or synthesis of cell wall components. The lack of phenotype in the transformed hybrid aspen lines could also be due to compensatory changes occurring. The possibility by plants to sense changes and react by triggering multiple signaling pathways is an interesting feature (Pilling and Hofte, 2003). Studies have indicated the possibility to compensate a loss of one polymer with an increase in another. In the Arabidopsis *mur3* mutant there is an increase in specific xyloglucan derived oligomers possibly to compensate for a loss of fucose containing side-chain (Madson et al., 2003).

More extensive work (analysis of cell wall sugars, microscopic analysis, PtEPC1 protein levels, inducible constructs and also a prolonged growth period) is required in order to find out what part PtEPC1 plays in the intricate machinery that regulates wood formation.

Combining Arabidopsis and poplar model systems has been used extensively in this thesis as a way to identify and rapidly screen for the function of genes. The fact that Arabidopsis when grown under short day conditions can resemble poplar wood forming tissues to some extent and the high degree of gene conservation facilitates the integration of the two species in studying wood development. The three stage selection approach used in Paper I, is a good method to identifying additional genetic regulators not found in alternative studies since the three genes identified in this way were not picked up in a separate study of the transcriptome of secondary growth in Arabidopsis (Oh et al., 2003). It is of great importance to understand function and interaction of all genes if we are to fully understand plant development. Despite the efforts made to characterising the function of all genes in Arabidopsis still approximately 30 % of the genes have either only a hypothetical or even unknown function. The DDP genes studied in Paper II were depicted as having no known function and our efforts in characterising these, despite a lack of severe growth effects, revealed differences in FT-IR that points to roles in the secondary growth. The exact role for DDPs remains unknown but it is clear they are needed to maintain correct levels of hemicellulose, pectin and lignin.

It is important to remember that all experiments in this thesis have been conducted under ideal greenhouse conditions and the possibility that a phenotype would manifest itself if plants were grown under suboptimal conditions or even outside in a field trial has not been evaluated. The next steps to be taken are to pin point the precise sugar alterations seen for both the DDPs as well as PtEPC1 and continue to integrate both Arabidopsis and

poplar model systems. To be able to fully facilitate in the production processes for the forest industries all plant genes need to be identified and their interaction determined. With the ever increasing insight gained and the constant development of better scientific techniques, pathways will be more easily unraveled and this holds the promise of new and interesting findings in the future.

6 Acknowledgements

Alan, thanks for providing me with an umbrella when it "sometime rains and sometimes pours"... I have learned alot during these past years and even though it perhaps was not optimal with your move to South Hampton I still think we made it work!

All the best to the people (both new and old) who has formed the Marchant group over the years in Umeå, with help, assistance and input: Cat, Sunil, Ai-Min, Min-Soo and of course Emma (som har givit mycket och uppskattad input, stöd och trevligt resesällskap =)!

Ett uppskattande riktat till Forskarskolan med alla som är associerade till den och till minne av alla (sjukt?) lysande resor.

Tack till alla er som får UPSC att fungera och vara en underhållande plats att vara på! Särskilt tack till Maria L, Inga-Lis, Siv, Kjell O, Inga-Britt, Marie, Britt-Marie, Stefan och "dataKjell".

"Kramis" to all of you wonderful PhD students and postdocs that have provided interesting discussions, lovely laboratory company and of course fun parties! Extra hugs to all my office buddies and to Maria K, Emma, Anna K, Madde, Sara A, Jon, Sara V.P, Annika and Lasse.
and András....you know...I know...let's leave it at that.... ;-)

Ida! =) Stödet i stormen, tack för underhållande (o vindränkta) kvällar o kletande med drejerigrejer.

Cat, Dan, Amaya and Celiné, thanks for all the fun times and valuable support you all are! Hope to see you all soon again! xoxo

Kramar till alla mina vänner i både Sthlm/Uppsala, Ö-r-n-s-k-ö-l-d-s-vik o Umeå. Underbart att få glömma bort bekymmer o besvär (både med o utan rom o cola...=)

Tack till mina underbara päron för att ni stöttat mig (trots att jag flyttade så långt bort) och gjort ert bästa för att försöka förstå vad jag gör, det där med att ”titta på träd” ni vet. Kram till mina bröder (med alla yngel) för att ni lyser upp min tillvaro och för att vi tillsammans försöker göda våra föräldrar varje jul!

Stor saknad känns efter den äldre generationen, Mormor Ellen och Farmor Mia, vars enda kommentar var ”Det där fixar du!” så fort tvivlet över doktorerandet satte in.

Erik, puzzuz, för att du är min filbunke till pratkvarn o ser så snygg ut på en skoter...OMGeee....mums.... Ogog...

Tack till Elin H., Evan och Parken.... =) mina fysiska psykiska ventiler under denna tid...

References

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

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301,** 653-657.

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

Azpiroz-Leehan, R., and Feldmann, K. (1997). T-DNA insertion mutagenesis in Arabidopsis: going back and forth. Trends in Genetics **13,** 152-156.

Bosh, M., and Hepler, P. (2005). Pectin Methylesterases and Pectin Dynamics in Pollen Tubes. Plant Cell **17,** 3219-3226.

Bouton, S., Leboeuf, E., Mouille, G., Leydecker, M.T., Talbotec, J., Granier, F., Lahaye, M., Hofte, H., and Truong, H.N. (2002). QUASIMODO1 encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in Arabidopsis. Plant Cell **14,** 2577-2590.

Bown, L., Kusaba, S., Goubet, F., Codrai, L., Dale, A.G., Zhang, Z., Yu, X., Morris, K., Ishii, T., Evered, C., Dupree, P., and Jackson, S. (2007). The ectopically parting cells 1-2 (epc1-2) mutant exhibits an exaggerated response to abscisic acid. Journal of Experimental Botany **58,** 1813-1823.

Braam, J. (1999). If walls could talk. Curr Opin Plant Biol **2,** 521-524.

Briggs, G.C., Osmont, K.S., Shindo, C., Sibout, R., and Hardtke, C.S. (2006). Unequal genetic redundancies in Arabidopsis--a neglected phenomenon? Trends Plant Sci **11,** 492-498.

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

Brown, D.M., Zhang, Z., Stephens, E., Dupree, P., and Turner, S.R. (2009). Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J **57,** 732-746.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., and Wittwer, C.T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem **55,** 611-622.

Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J **326 (Pt 3),** 929-939.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res **37,** D233-238.

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

Clare, A., Karwath, A., Ougham, H., and King, R.D. (2006). Functional bioinformatics for Arabidopsis thaliana. Bioinformatics **22,** 1130-1136.

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

Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. J Mol Biol **328,** 307-317.

Covey, S.N., Lomonossoff, G.P., and Hull, R. (1981). Characterisation of cauliflower mosaic virus DNA sequences which encode major polyadenylated transcripts. Nucleic Acids Res **9,** 6735-6747.

Czechowski, T., Bari, R.P., Stitt, M., Scheible, W.R., and Udvardi, M.K. (2004). Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant J **38,** 366-379.

Delmer, D.P., and Amor, Y. (1995). Cellulose biosynthesis. Plant Cell **7,** 987-1000.

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

Ebringerova, A., and Heinze, T. (2000). Naturally occurring xylans structures, isolation procedures and properties. Macromolecular Rapid Communications **21,** 542-556.

Fagard, M., Desnos, T., Desprez, T., Goubet, F., Refregier, G., Mouille, G., McCann, M., Rayon, C., Vernhettes, S., and Hofte, H. (2000). PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of Arabidopsis. Plant Cell **12,** 2409-2424.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature **391,** 806-811.

Franck, A., Guilley, H., Jonard, G., Richards, K., and Hirth, L. (1980). Nucleotide sequence of cauliflower mosaic virus DNA. Cell **21,** 285-294.

Geisler-Lee, J., Geisler, M., Coutinho, P.M., Segerman, B., Nishikubo, N., Takahashi, J., Aspeborg, H., Djerbi, S., Master, E., Andersson-Gunneras, S., Sundberg, B., Karpinski, S., Teeri, T.T., Kleczkowski, L.A., Henrissat, B., and Mellerowicz, E.J. (2006). Poplar carbohydrate-active enzymes. Gene identification and expression analyses. Plant Physiol **140,** 946-962.

Gutierrez, L., Mauriat, M., Guenin, S., Pelloux, J., Lefebvre, J.F., Louvet, R., Rusterucci, C., Moritz, T., Guerineau, F., Bellini, C., and Van Wuytswinkel, O. (2008). The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol J **6,** 609-618.

Henrissat, B., Coutinho, P.M., and Davies, G.J. (2001). A census of carbohydrate-active enzymes in the genome of Arabidopsis thaliana. Plant Mol Biol **47,** 55-72.

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

http://www.cazy.org/. CAZy ~ Carbohydrate-Active Enzymes database.

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

Karp, G. (1999). Cell and molecular biology. (John Wiley and Sons, Inc.).

Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G. (2008). Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem **77,** 521-555.

Lee, C., Teng, Q., Huang, W., Zhong, R., and Ye, Z.H. (2009). The poplar GT8E and GT8F glycosyltransferases are functional orthologs of Arabidopsis PARVUS involved in glucuronoxylan biosynthesis. Plant Cell Physiol.

Lee, C., Zhong, R., Richardson, E.A., Himmelsbach, D.S., McPhail, B.T., and Ye, Z.H. (2007). The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant Cell Physiol **48,** 1659-1672.

Lind, T., Tufaro, F., McCormick, C., Lindahl, U., and Lidholt, K. (1998). The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. J Biol Chem **273,** 26265-26268.

Madson, M., Dunand, C., Li, X., Verma, R., Vanzin, G.F., Caplan, J., Shoue, D.A., Carpita, N.C., and Reiter, W.D. (2003). The MUR3 gene of Arabidopsis encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. Plant Cell **15,** 1662-1670.

McCormick, C., Leduc, Y., Martindale, D., Mattison, K., Esford, L.E., Dyer, A.P., and Tufaro, F. (1998). The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. Nat Genet **19,** 158-161.

Micheli, F. (2001). Pectin methylesterases: cell wall enzymes with important roles in plant physiology. Trends Plant Sci **6,** 414-419.

Moreau, C., Pesquet, E., Sjödin, A., Muniz, L., Bollhönder, B., Kaneda, M., Samuels, L., Jansson, S., and Tuominen, H. (2009). A unique program for cell death in xylem fibers of *Populus* stem. Plant Journal **58,** 260-274.

Morikawa, H., and Senda, M. (1978). Infrared Analysis of Oat Coleoptile Cell Walls and Oriented Structure of Matrix Polysaccharides In the Walls Plant and Cell Physiology **19,** 327-336.

Morikawa, H., Hayashi, R., and Senda, M. (1978). Infrared Analysis of Pea Stem Cell Walls and Oriented Structure of Matrix Polysaccharides In Them. Plant and Cell Physiology **19,** 1151-1159.

Mouille, G., Ralet, M.C., Cavelier, C., Eland, C., Effroy, D., Hematy, K., McCartney, L., Truong, H.N., Gaudon, V., Thibault, J.F., Marchant, A., and Hofte, H. (2007). Homogalacturonan synthesis in Arabidopsis thaliana requires a Golgi-localized protein with a putative methyltransferase domain. Plant J **50,** 605-614.

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

Nilsson, O., Aldén T, Sitbon F, Little A, Chalupa V, Sandberg G and Olsson O. (1992). Spatial pattern of cauliflower mosaic virus 35S promoter-luciferase expression in transgenic hybrid aspen trees monitored by enzymatic assay and non-destructive imaging. Transgenic Research **1,** 209-220.

Northington, D., and Schneider, E. (1996). The Botanical World. (Wm. C. Brown Publishers).

Odell, J.T., Nagy, F., and Chua, N.H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313,** 810-812.

Oh, S., Park, S., and Han, K.H. (2003). Transcriptional regulation of secondary growth in Arabidopsis thaliana. J Exp Bot **54,** 2709-2722.

Osorio, S., Castillejo, C., Quesada, M.A., Medina-Escobar, N., Brownsey, G.J., Suau, R., Heredia, A., Botella, M.A., and Valpuesta, V. (2008). Partial demethylation of oligogalacturonides by pectin methyl esterase 1 is required for eliciting defence responses in wild strawberry (Fragaria vesca). Plant J **54,** 43-55.

Persson, S., Wei, H., Milne, J., Page, G.P., and Somerville, C.R. (2005). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad Sci U S A **102,** 8633-8638.

Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., Khitrov, N., Auer, M., and Somerville, C.R. (2007). Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. Proc Natl Acad Sci U S A **104,** 15566-15571.

Pilling, E., and Hofte, H. (2003). Feedback from the wall. Curr Opin Plant Biol **6,** 611-616.

Reiter, W.-D. (1998). Arabidopsis thaliana as a model system to study synthesis, structure, and function of the plant cell wall. Plant Physiology and Biochemistry **36,** 167-176.

Rose, J.K., Braam, J., Fry, S.C., and Nishitani, K. (2002). The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol **43,** 1421-1435.

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

Schwanninger, M., Rodrigues, J.C., Pereira, H., and Hinterstoisser, B. (2004). Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. Vibrational Spectroscopy **36,** 23-40.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M., and Goff, S.A. (2002). A high-throughput Arabidopsis reverse genetics system. Plant Cell **14,** 2985-2994.

Shedletzky, E., Shmuel, M., Trainin, T., Kalman, S., and Delmer, D. (1992). Cell Wall Structure in Cells Adapted to Growth on the Cellulose-Synthesis Inhibitor 2,6-Dichlorobenzonitrile : A Comparison between Two Dicotyledonous Plants and a Graminaceous Monocot. Plant Physiol **100,** 120-130.

Shibuya, N., and Minami, E. (2001). Oligosaccharide signalling for defence responses in plant. Physiol Mol Plant Pathol **59,** 223-233.

Singh, S.K., Eland, C., Harholt, J., Scheller, H.V., and Marchant, A. (2005). Cell adhesion in Arabidopsis thaliana is mediated by ECTOPICALLY PARTING CELLS 1 - a glycosyltransferase (GT64) related to the animal exostosins. Plant Journal **43,** 384-397.

Sjodin, A., Street, N.R., Sandberg, G., Gustafsson, P., and Jansson, S. (2009). The Populus Genome Integrative Explorer (PopGenIE): a new resource for exploring the Populus genome. New Phytol **182,** 1013-1025.

Takahashi Schmidt, J. (2008). Functional Studies of Selected Extracellular Carbohydrate-Active Hydrolases in Wood Formation. In Forest genetics and plant physiology (*PhD thesis*, Umeå: SLU).

Takeda, T., Furuta, Y., Awano, T., Mizuno, K., Mitsuishi, Y., and Hayashi, T. (2002). Suppression and acceleration of cell elongation by integration of xyloglucans in pea stem segments. Proc Natl Acad Sci U S A **99,** 9055-9060.

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

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

Trygg, J., and Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics **16,** 119-128.

Turner, S.R., and Somerville, C.R. (1997). Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell **9,** 689-701.

Turner, S.R., Taylor, N., and Jones, L. (2001). Mutations of the secondary cell wall. Plant Mol Biol **47,** 209-219.

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

Ubeda-Tomas, S., Edvardsson, E., Eland, C., Singh, S.K., Zadik, D., Aspeborg, H., Gorzsas, A., Teeri, T.T., Sundberg, B., Persson, P., Bennett, M., and Marchant, A. (2007). Genomic-assisted identification of genes involved in secondary growth in Arabidopsis utilising transcript profiling of poplar wood-forming tissues. Physiologia Plantarum **129,** 415-428.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol **3,** RESEARCH0034.

Willats, W.G., McCartney, L., Mackie, W., and Knox, J.P. (2001). Pectin: cell biology and prospects for functional analysis. Plant Mol Biol **47,** 9-27.

Williams, D.H., and Fleming, I. (1980). Spectroscopic Methods in Organic Chemistry. (McGraw-Hill Book Company (UK)).

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. Pattern Recogn. **8,** 127-139.

Vorwerk, S., Somerville, S., and Somerville, C. (2004). The role of plant cell wall polysaccharide composition in disease resistance. Trends Plant Sci **9,** 203-209.

Wu, A.M., Rihouey, C., Seveno, M., Hornblad, E., Singh, S.K., Matsunaga, T., Ishii, T., Lerouge, P., and Marchant, A. (2009). The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. Plant J **57,** 718-731.

Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K.L., Stewart, N.R., Syrenne, R.D., Yang, X., Gao, P., Shi, W., Doeppke, C., Sykes, R.W., Burris, J.N., Bozell, J.J., Cheng, M.Z., Hayes, D.G., Labbe, N., Davis, M., Stewart, C.N., Jr., and Yuan, J.S. (2009). Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics **10 Suppl 11,** S3.

Zablackis, E., Huang, J., Muller, B., Darvill, A.G., and Albersheim, P. (1995). Characterization of the cell-wall polysaccharides of Arabidopsis thaliana leaves. Plant Physiol **107,** 1129-1138.

Zhang, D., Yang, X., Zhang, Z., and Li, B. (2010). Expression and nucleotide diversity of the poplar *COBL* gene. Tree Genetics & Genomes **6,** 331-344.

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

Zhong, R., Morrison, W.H., 3rd, Freshour, G.D., Hahn, M.G., and Ye, Z.H. (2003). Expression of a mutant form of cellulose synthase AtCesA7 causes dominant negative effect on cellulose biosynthesis. Plant Physiol **132,** 786-795.