Deep digital gene expression profiling during early and late tension wood induction in *Eucalyptus* trees

by

Janine Francina Silberbauer

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

April 2011

Under the Supervision of Prof. Alexander A Myburg and Co-supervision of Prof. Dave K Berger and Prof. Fourie Joubert

Declaration

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

_____

Janine Silberbauer

April 2011

DISSERTATION SUMMARY

DEEP DIGITAL GENE EXPRESSION PROFILING DURING EARLY AND LATE TENSION WOOD INDUCTION IN *EUCALYPTUS* TREES

Janine F. Silberbauer

Supervised by ***Prof. A.A. Myburg***

Co-supervised by ***Prof D.K Berger*** *and* ***Prof F. Joubert***

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

Department of Genetics

University of Pretoria

Genetic engineering of superior wood properties and exploiting natural genetic variation found within commercially important trees, such as *Eucalyptus* spp., promise to increase cellulose biomass production. It is therefore essential to understand the molecular genetics of wood formation. Digital Gene Expression (DGE) profiling is adept in not only assessing the expression level of genes transcriptome-wide, but also in characterising alternative splice forms of transcripts and identifying novel transcripts. Tension wood is a specialised type of wood which functions in the response to mechanical stress in trees and is formed on the upper side of a branch or a bent stem. The characteristics of tension wood differ from normal wood by increased cellulose and xyloglucan content and decreased lignin and xylan content. During tension wood formation, transcriptome-wide changes in the expression of genes involved in secondary cell wall formation underlie changes in cell wall composition. Most notably is an increase in fasciclin-like arabinogalactan protein (*FLA*) and xyloglucan endotransglucosylase (*XTH*) and a decrease in lignin biosynthesis gene expression. Differential expression patterns are shown by cellulose synthase (*CesA*) genes, which have been found to be either up- or down-regulated during tension wood formation. No previous study has

profiled gene expression during early as well as late tension wood formation. The aim of this M.Sc study was to identify genes that are differentially expressed during early tension wood induction and late tension wood formation in the immature xylem tissues of *Eucalyptus grandis x Eucalyptus urophylla* hybrid trees.

DGE profiling is a transcriptome-wide expression profiling technique based on ultra-high throughput second generation DNA sequencing technology. The processing, analysis and interpretation of DGE data has not yet been standardised. To address this problem, a case study was performed of DGE data mapping to seven well characterised *Eucalyptus grandis CesA* (*EgCesA*) genes. The DGE data processing guidelines developed based on this case study produced *EgCesA* expression profiles in normal wood that were comparable to the profiles of these genes determined with other technologies. A possible alternative splice variant occurring during tension wood formation was identified for the secondary cell wall gene *EgCesA3*. However future work is needed for the validation of this observation.

Early tension wood induction and late tension wood formation was investigated by sampling differentiating xylem from ramets of a *Eucalyptus grandis x Eucalyptus urophylla* clone induced to form tension wood for 6 hours, 24 hours, 1 week, 2 weeks and 6 months. Up to 2,654 transcripts were found to be significantly differentially expressed during tension wood formation. *FLA* transcripts were the highest expressed transcripts and were, along with *XTH* genes, highly up-regulated in early and late tension wood formation. Genes differentially regulated during early tension wood formation reflected a general stress response and hormone signalling pathways. Late tension wood formation was marked by the differential regulation of secondary cell wall biosynthetic genes, which reflected the chemical composition of tension wood. Two secondary cell wall *CesA* genes were significantly up-regulated, while genes involved in lignin and xylan biosynthesis were significantly down-regulated. Observations suggest that the eucalypt trees used in this study formed tension wood to stabilise the bent stem, while apical dominance was transferred to new side branches which showed signs of extra secondary growth.

PREFACE

Wood produced by trees is the raw material used for a number of products, such as pulp for paper and forms the corner stone of a massive global wood and paper products industry. To reduce the exhaustion of natural forests and to meet growing demands for wood, tree breeders aim to increase wood production from forest plantations. Exploiting the natural genetic diversity within forest tree species is one approach towards improving wood fibre quality. Breeders wish to alter the chemical characteristics of wood fibres in order to reduce chemical and energy consumption during the pulping process. Tension wood is a naturally occurring modified wood which plays an important role in tree growth. The three main components of wood fibres, cellulose, lignin and hemicellulose, are altered in tension wood. Even though tension wood is regarded by the industry as a major defect which reduces the quality of lumber, the chemical composition of this type of wood tends towards the ideal chemical composition of pulp wood. That is, increased cellulose content and decreased lignin content. Furthermore, tension wood induction involves large scale reprogramming of gene expression profiles. Therefore the induction of tension wood may be used as a system to study global underlying genetic changes during wood formation. The molecular functions of genes involved in the biosynthesis of industrially important constituents of wood fibres may be investigated using various expression profiling approaches. Expression profiling has only been performed in early *Eucalyptus* tension wood forming tissues and in mature *Populus* tension wood forming tissues. The main **aim of this M.Sc study** was to identify genes that are differentially expressed during early and late tension wood formation in the immature xylem tissues of a *Eucalyptus grandis x Eucalyptus urophylla* clone. This was achieved by Digital Gene Expression (DGE) profiling, an ultra-high-throughput transcript tagging technique based on second generation DNA sequencing technology.

Literature pertaining to the development of tension wood is reviewed in the first section of **Chapter 1**. The anatomical and chemical characteristics of tension wood are discussed. The underlying gene expression patterns of genes involved in the biosynthesis of lignin, cellulose and hemicellulose as revealed during key tension wood studies are provided. Lastly, the role of three plant hormones in tension wood development as presented in previous studies is discussed. Section

two reviews various tag-based expression profiling methods. In closing, data analysis methods employed in DGE and SAGE are discussed.

The objective to develop a set of guidelines for the processing of raw DGE data is presented in **Chapter 2.** At the onset of this study DGE was a novel technology with few publications to guide processing of the complex data sets generated by the technique. Thus this objective arose as a means to extract gene expression profiles from DGE data. These guidelines are based on an in depth examination of the features of DGE data mapping to the *Eucalyptus* cellulose synthase (*EgCesA*) genes.

The effect of tension wood formation on genes involved in the biosynthesis of secondary cell wall cellulose, lignin and hemicelluloses is investigated in **Chapter 3.** Significantly differentially expressed genes induced by bending stress in *Eucalyptus* stems are identified, and the formation of tension wood is verified by comparing expression profiles obtained in this study to signature tension wood expression profiles from key tension wood studies. Finally genes significantly differentially expressed during early and late *Eucalyptus* tension wood formation are presented in this chapter. The conclusions of and future directions for this study are summarised in a final **Concluding remarks** section of the dissertation.

The outcomes from this study undertaken between February 2008 and December 2010 in the Department of Genetics, University of Pretoria, under the supervision of Prof. AA Myburg are presented in this dissertation. The following posters were presented based on preliminary results obtained in this M.Sc study:

Silberbauer, J.F, McNair, G, Ranik, M, and Myburg A.A 2008. Analysis of the *Eucalyptus* transcriptome during tension wood formation. Congress of the South African Genetics Society (SAGS), University of Pretoria, Pretoria, South Africa (Awarded ‘best poster presentation by a postgraduate student’).

Silberbauer, J.F, Law, P, Joubert, F, and Myburg, A.A. 2009. An Analysis Pipeline for Digital Gene Expression (DGE) Profiling. Second Southern African Bioinformatics Workshop at the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT), University of the Witwatersrand, Johannesburg, South Africa.

Silberbauer, J.F, Mizrachi, E, Hefer, C and Myburg, A.A. 2010. Analysis of Digital Gene Expression (DGE) profiling data from *Eucalyptus* tension wood forming tissues. Congress of the South African Genetics Society (SAGS), University of the Freestate, Bloemfontein, South Africa (Awarded 'best poster presentation by a postgraduate student').

ACKNOWLEDGEMENTS

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

- To Prof. Zander Myburg for his patient supervision of this project, professionally led with creativity and for reviewing the dissertation.
- Prof. Dave Berger for his insights into this project and reviewing of this dissertation.
- Prof. Fourie Joubert for advice on the bioinformatic aspects of the project and for reviewing the dissertation.
- To Eshchar Mizrachi for motivation and guidance on the project and review of the dissertation.
- To Grant McNair, Steven Hussey and Martin Ranik for your support during sample preparation trips at the University of the Western Cape.
- To Tracey Meiring for delivering samples to the University of the Western Cape.
- To FMG research group members Minique de Castro, Therese de Castro, Mmoledi Mphahlele, Eshchar Mizrachi, Nicky Creux, Melissa Reynolds, Anand Kullan and Steven Hussey. I thank you for your assistance on the fieldtrips where we collected the samples for this project. Also to the whole FMG group thank you for your friendship, company and support.
- To the Department of Genetics and the Forestry and Agricultural Biotechnology Institute (FABI) at the University of Pretoria for allowing me to make use of their excellent facilities.
- To the Bioinformatics and Computational Biology Unit at the University of Pretoria for their invaluable support and assistance in the project, especially Philip, Jurgens and Charles.

- To Sappi for supplying the plant materials used in this study and for the assistance provided by Sappi employees during field trips, as well as for funding contributions.

- To Mondi for funding contributions.

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

- Last but not the least, to my family and friends for all their wonderful support and encouragement.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABA	Abscisic acid
AGP	Arabinogalactan protein
AT	*Arabidopsis thaliana*
CAZymes	Carbohydrate active enzymes
cDNA	Complementary deoxyribonucleic acid
CK	Cytokinin
CTAB	Cetyl trimethylammonium bromide
DGE	Digital Gene Expression
EST	Expressed sequence tag
GA	Gibberellic acid
GPI	Glycosylphosphatidylinositol
JA	Jasmonic acid
MAP	Microtubule associated protein
MAQC	MicroArray Quality Control
MIQE	Minimum Information for Publication of Quantitative real time PCR Experiments
MPSS	Massively parallel signature sequencing
PCR	Polymerase chain reaction
PPM	Parts per million
RT-qPCR	Real-time quantitative polymerase chain reaction

SNP	Single nucleotide polymorphism
ROS	Reactive oxygen species
SAGE	Serial Analysis of Gene Expression
TW/C	Tension wood to control ratio

Chapter 1

LITERATURE REVIEW

THE BIOLOGY OF TENSION WOOD FORMATION AND APPLICATION OF TAG BASED EXPRESSION PROFILING APPROACHES FOR THE INVESTIGATION OF TRANSCRIPTOMES

1.1. INTRODUCTION

Global energy consumption keeps increasing at a very rapid rate as the world population continues to grow. The consequence of this is the deterioration of the environment, coupled with an increasing demand for natural resources. Natural forests are just one of the world's resources that are exploited to a great extent (Rudel, 2006). Trees are of immense environmental and industrial importance. In the past few decades, the driving force behind natural forest devastation has been more industrially driven by practices such as logging, mining and large-scale agriculture than other changing agents such as subsistence farming and road building (Butler and Laurance, 2008, Geist and Lambin, 2002, Rudel, 2007). The world's heavy demand for wood cannot be sustained by natural forests at the same continued rate without the severe degradation of these biologically sensitive areas (Fenning and Gershenzon, 2002, Imai et al., 2009). Wood obtained from trees is the raw material used for numerous products such as fuel, pulp, paper and furniture. The process of wood formation is one of the most important carbon sequestration processes on a global scale and therefore crucial to the well-being of the environment (Bala et al., 2007, Fearnside, 2008, Geist and Lambin, 2002, Jana et al., 2009).

Ever growing environmental concern supports movements that aim to curb the destruction of natural forests (Butler and Laurance, 2008, Imai et al., 2009). In accord with being less reliant on natural forests to meet the demand for wood, tree breeders aim to increase wood production from forest plantations. This is in part done by improving tree plantations to grow faster to increase the produced biomass per hectare and to be less demanding in terms of water consumption (Nelson and Johnsen, 2008). Breeders not only wish to increase cellulose production but also alter other chemical characteristics of wood. This will reduce chemical and energy consumption during the pulping process (Pilate et al., 2004). Traditional breeding methods require many years to incorporate the desired characteristics into a specific forest plantation crop. The reason for this is that forest trees do not reach reproductive maturity quickly, resulting in long periods of evaluation before selection for specific traits can be made (Fenning and Gershenzon, 2002). A future objective for the forestry industry is to breed elite clones and grow plantations for a specific end-use, producing wood

specialised either for paper or timber production (Andersson Gunnerås, 2005, Nelson and Johnsen, 2008). The application of biotechnology tools in forestry opens the opportunity to greatly accelerate the breeding cycle by making tree improvement strategies more directed. Studying the genomics of wood development reveals candidate genes and biochemical pathways that underlie phenotypic traits of interest (Shepherd et al., 2003). Existing genetic variation present in plantation trees can then be utilised to genetically improve wood and growth properties by combining specific genotypes into a desired clonal type (Fenning and Gershenzon, 2002, Nelson and Johnsen, 2008).

The formation of different types of wood within a tree stem and between trees is attributed to environmental and genetic variation (reviewed in Plomion et al., 2001). Environmental variation during the growing season, especially in temperate zones, produces the difference between early, late and reaction wood. The intrinsic variation during the lifespan of a tree causes the differences found between juvenile and mature wood. Sapwood comprises the outer living part of a tree and the mostly dead core of a tree consists of heartwood. Further distinctions between different types of wood are attributed to anatomical, physical and chemical differences. Reaction wood is a stress-induced type of wood that forms by gravitational and mechanical stimuli when a tree is bent or twisted (Plomion et al., 2001). Commercially important plantation species such as *Eucalyptus* and *Populus* form reaction wood. This is problematic to the solid wood industry as reaction wood formation is regarded as a major defect. This is because during later lumber drying phases the compositional differences between normal and reaction wood can cause severe problems such as log-end cracks and lumber distortion (Fang et al., 2008).

Understanding the process of wood formation will enable researchers to improve wood biomass production and enhance the properties of wood and wood fibres. The characteristics of wood are directed by the composition and morphology of the xylem cell walls (Mansfield, 2009). Xylogenesis is a highly complex form of cell differentiation with distinctly different developmental phases. The developmental mechanisms of xylogenesis can further be elucidated by studying systems in which the normal process has been altered. Thus even though reaction wood is an undesired trait in plantation trees, comparative studies to normal wood may be useful to refine the model of wood

development (Hamilton et al., 2009, Washusen and Evans, 2001). Comparative gene expression analysis studies could assist in unravelling the underlying genetic mechanisms. This will highlight key genes that influence the chemical composition and architecture of reaction wood cell walls.

Second generation sequencing has changed the face of genomic research. The vast amount of data produced enables whole-genome/ whole-transcriptome investigations which ultimately contribute to a systems biology approach in the quest to understand the biology of an organism (Nelson and Johnsen, 2008). The genomics of wood development and wood quality traits have been greatly enriched by the various applications of second generation sequencing platforms. Knowledge of wood development in poplar, pine and eucalypts have been enhanced by second generation sequencing applications through single nucleotide polymorphism (SNP) discovery projects (Kulheim et al., 2009, Novaes et al., 2008) and chloroplast genome sequencing for evolutionary studies (Cronn et al., 2008). Other applications of second generation sequencing technologies, which combine older techniques with high-throughput sequencing include *de novo* genome assembly (Reinhardt et al., 2009) and re-sequencing (Vasta et al., 2009), small non-coding RNA characterisation and analysis (Darley et al., 2001), methylation pattern analysis (Barski et al., 2007, Wang et al., 2009), protein-nucleic acid interactions (Seo et al., 2009) and transcriptome studies with regards to transcriptome architecture (Mortazavi et al., 2008, Nagalakshmi et al., 2008), gene discovery (Emrich et al., 2007) and expression profiling (Eveland et al., 2008, Mortazavi et al., 2008). Digital Gene Expression (DGE) is a tag-based expression profiling application developed by Illumina (San Diego, California, USA) for ultra-deep, gene expression profiling (Hanriot et al., 2008). DGE and the many similarities that it shares with Serial Analysis of Gene Expression (SAGE) (Hene et al., 2007) forms part of the focus of section 3 of this literature review.

Like most technological advancements, high-throughput genome-wide projects come with a few hurdles. Second generation sequencing techniques and now third generation single molecule sequencing techniques generate almost unmanageable amounts of data when using the current computing power and technology (Gupta, 2008). Chain et al. (2009) also discuss this problem and state that, up to 2009, approximately 1300 draft eukaryotic genomes have been sequenced, but that of

these only about 200 eukaryotic genomes have been successfully completed. This emphasises the fact that the new sequencing techniques have made it very easy to produce immense amounts of data - up to one billion bases in a single day (Pop and Salzberg, 2009), but successful analysis of this data remains problematic. In addition to the problem of managing large amounts of data, there is the problem of dealing with new types of data. Thorough assessment of these different types of new data is crucial to identify any unexpected characteristics which may be hidden in plain sight in the sequence information (Keime et al., 2007). These investigations should be accompanied by the comparison and knowledge of techniques from which inspiration was drawn for the development of these new sequencing applications (Hene et al., 2007).

Section 2 of this literature review focuses on the biology of tension wood formation. This includes changes seen in the anatomy and chemical characteristics of tension wood when compared to normal wood (sections 2.1 - 2.2). Published research describing the most important biosynthetic genes that impact these characteristics of tension wood will be discussed. Section 2.3 will conclude with the hormonal control of tension wood formation and genes found to play a significant role in hormone-related biosynthesis pathways. Section 3 will focus on tag-based methods used for gene discovery and expression analysis. This includes expressed sequence tag (EST) analysis (section 3.1), SAGE (section 3.2) and how the SAGE technique has been applied to second generation sequencing with DGE (section 3.3). Section 3.3.1 will conclude with how existing knowledge on SAGE can guide the analysis of DGE data analysis and how previous studies have dealt with DGE data processing. Finally general concluding remarks are given in section 4.

1.2. TENSION WOOD FORMATION

A type of environmentally variable wood found within trees is reaction wood. Reaction wood is a specialised type of wood which allows a tree movement so that it is able to maintain its correct morphology. It is not only found in a tree's mature branches for additional strength and support, but may also be the mechanism of action by which a tree reacts to stem displacement (Jiang et al., 2008, Qiu et al., 2008). In herbaceous plants the reaction to a bent stem is simply asymmetrical longitudinal growth (Du and Yamamoto, 2007). In the case of large heavy plants like trees, an increase in longitudinal growth is not sufficient to correct organ displacement and the formation of reaction wood is the response to a bent stem (Du and Yamamoto, 2007, Funada et al., 2008).

The purpose of reaction wood could either be to stabilise a branch or to re-orientate a stem towards an upright position when bent due to constant heavy loads such as wind or snow (Bamber, 2001). The formation of reaction wood may be induced by many different stimuli including gravity, compression and tensile stress. Reaction wood formation may also be induced or inhibited artificially by applying hormones and chemicals (Du and Yamamoto, 2007). In some cases, reaction wood formation is the result of a tree's internal control mechanism in the absence of any external stimuli. Some trees produce reaction wood when the apical shoot is removed. This enables the replacement of the lost leader by causing a side branch to bend upwards (Fisher and Stevenson, 1981, Wilson, 2000).

Angiosperm (hardwood) and gymnosperm (softwood) trees form two different types of reaction wood termed tension wood and compression wood respectively. A characteristic by which tension wood and compression wood differ is the anatomical position at which reaction wood formation takes place. In gymnosperms, compression wood forms on the under-side of a stem bent downwards while in angiosperms, tension wood forms on the upper-side of a stem bent downwards (Bamber, 2001). The formation of tension wood is visible in cross sections of tree stems that have been bent for a prolonged period of time (Figure 1.1).

The reaction wood in angiosperms is termed tension wood as this is the area of the tree that is held in tension (Goswami et al., 2008). There is still much debate around the exact mechanism of

action of tension wood. The formation of a crystalline cellulose rich G-layer in the tension wood fibre that replaces most of the S_2 and S_3 layers of secondary cell wall in a normal wood fibre (Figure 1.2) is believed to play the crucial functional role in the tension wood mechanism of action (Clair et al., 2006a). Currently the growth stress hypothesis is the most widely accepted hypothesis on tension wood formation.

Figure 1.1: Cross section of a *Eucalyptus grandis* x *Eucalyptus urophylla* two-year-old stem after prolonged bending showing the visible difference between tension wood (highlighted) and opposite wood (lower-side). (Photo by Forest and Molecular Genetics Programme, University of Pretoria).

Longitudinal tensile mechanical stresses are created during cell growth, differentiation and maturation. Tensile stresses are created when the shrinking G-layer interacts with the surrounding secondary cell wall layers. The shrinking of the G-layer and the mechanism of action of tension wood relies on the altered anatomical, physical and chemical properties of tension wood (Clair et al., 2005, Washusen and Ilic, 2001). The S_2 secondary cell wall layer in particular has a large microfibril angle with low stiffness enabling it to respond to deforming stresses created by the loosely attached G-layer (Clair et al., 2005, Goswami et al., 2008). When these tensile stresses are uneven, as is the case during asymmetrical tension wood formation, a pulling action occurs which facilitates the correction of the stem or branch shape (Clair et al., 2006a, Funada et al., 2008, Yamamoto et al., 2002).

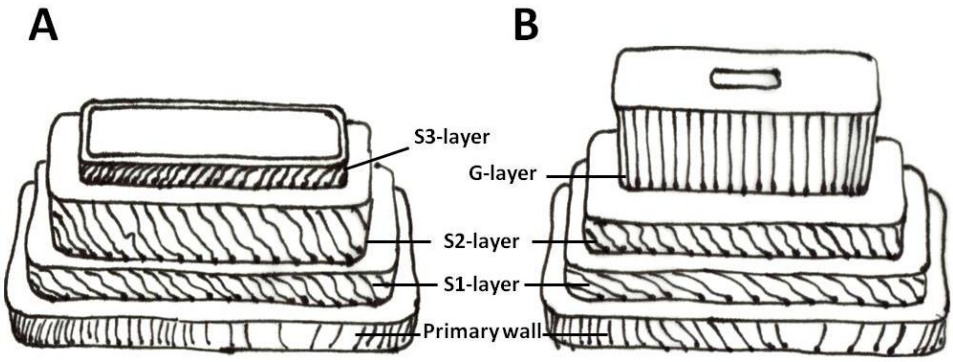


Figure 1.2: Diagram illustrating the anatomical differences between normal wood fibres (A) and G-fibres (B). The characteristic G-layer in a tension wood fibre replaces the S_3 and most of the S_2 layers of a normal wood fibre (redrawn by Michael Silberbauer from Andersson Gunnerås, 2005).

1.2.1 G-layer anatomy

Tension wood formation affects differentiation and maturation of the xylem cells during xylogenesis. In contrast to normal wood formation, differentiation during tension wood formation results in an increased number of fibres and rays and a decreased number of vessels. During the maturation phase of tension wood development, fibres and vessels are increased in length (Jourez et al., 2001a).

In some species a characteristic feature of tension wood is the gelatinous G-layer of the fibres (Bamber, 2001, Washusen, 2002). Different fibre patterns have been observed in different angiosperm species during tension wood formation. In some species it is difficult to define the difference between normal wood and tension wood, based on classical anatomy (Clair et al., 2006b). Fisher and Stevenson (1981) demonstrated G-layer formation in the branches of 46% of the 122 species investigated for tension wood formation. Similarly Clair et al. (2006b) found the formation of G-layers in only 7 of the 21 Angiosperm species studied. *Eucalyptus* is a hardwood species that forms tension wood sometimes lacking G-layers. G-layer formation in eucalypts was confirmed by studies such as Paux et al. (2005) and Washusen et al. (2005) but found to be absent in a tension wood formation study by Qiu et al. (2008). Fang et al. (2008) proposed that the G-layer plays the most important role in generating the high tensile growth stress required to maintain stem shape and branch

angles. The study found that in trees that develop G-layers, more G-fibres per unit of tissue area and thicker G-layers are correlated with higher longitudinal growth stress. Qiu et al. (2008) argued that G-layers may not serve a purpose of primary importance during tension wood formation due to the absence of G-layer formation not only in *Eucalyptus*, but also in other species (Clair et al., 2006b, Fisher and Stevenson, 1981, Prodhan et al., 1995).

In a normal wood fibre cell the secondary cell wall is divided into three layers - namely the S_1, S_2 and S_3 layers (Figure 1.2A). In each of the three layers, the cellulose microfibrils have a parallel arrangement. The S_1 layer, the thinnest of the secondary cell wall layers, has a transverse microfibril angle of 60° to 80° with regards to the cell axis. The S_2 layer is the thickest layer of the secondary cell wall and also the most important layer for the mechanical support of the tree. The microfibril angle of the S_2 layer ranges between 5° and 30° and is longitudinal to the cell axis. The S_3 layer again has a transverse microfibril arrangement that can vary between 60° and 90° (Plomion et al., 2001). The microfibril angle of the G-layer is greatly decreased so that the microfibrils lie parallel to the vertical cell axis, giving the G-layer a hard crystalline property (Goswami et al., 2008). As an example the G-layer microfibril angle of *Eucalyptus* trees used in a study by Baba et al. (2000) was reduced from 22.5° to 3.5°.

As mentioned above, the G-layer in tension wood fibres replaces the S_3-layer and most of the S_2 layer and is thickened to a high extent (Figure 1.2B, Andersson Gunnerås, 2005). G-layers may be present in developing tension wood fibres as early as three weeks after bending and can be visualised by staining with Astra blue which stains the cellulose content of the fibres blue (Figure 1.3, Jourez et al., 2001b). In the middle of the tension wood forming zone thickening of the G-layer may be so extreme that the lumen of the G-fibre is completely filled up (Figure 1.3A, Coutand et al., 2004, Goswami et al., 2008).

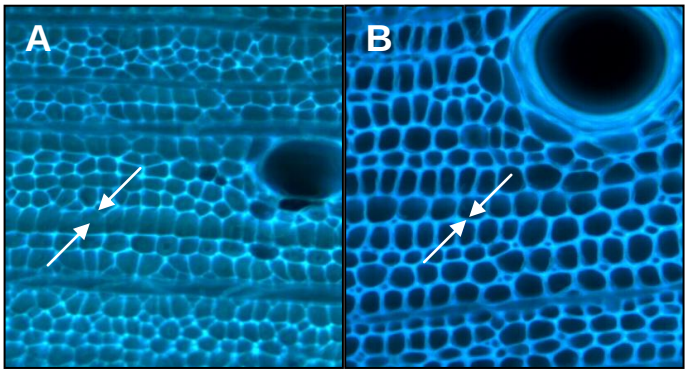


Figure 1.3: Anatomical difference between tension wood and opposite wood. Cross sections were taken from the *Eucalyptus* tree in Figure 1.1 and stained with Astra blue. The stained cellulose reveals the highly thickened G-layer (A) in contrast to the thinner cell walls of the opposite wood cross section (B). Arrows indicate the inner and outer borders of a wood cell fibre. Slides from the Forest and Molecular Genetics Programme, University of Pretoria

1.2.2 Chemical composition of the G-fibre

The three main components of a normal secondary fibre cell wall are altered in the G-layer. The normal 2:1:1 cellulose: hemicellulose: lignin ratio is altered so that the cellulose content is greatly increased. This increase in hard, crystalline cellulose is accompanied by an increase in the hemicelluloses xyloglucan (Mellerowicz et al., 2008) and rhamnogalacturonan (Bowling and Vaughn, 2008). Overall a decrease is observed in xylan and lignin content (Joseleau et al., 2004, Nishikubo et al., 2007). In a biochemical study of *Eucalyptus camaldulensis* conducted by Baba et al. (2000), the tension wood cellulose content was approximately twice the amount found in normal wood, and the tension wood lignin content was less than half the amount found in normal wood. Due to the decreased lignin content, the G-layer is very loosely associated with the S-layers resulting in a woolly appearance when tension wood is sawn (Gorišek and Straže, 2006). The S:G (syringyl: guaiacyl) ratio of the monolignol subunits in lignin is increased. Joseleau et al. (2004) demonstrated through an immunochemical study the presence of discrete syringyl units embedded in the crystalline cellulose of the G-layer. This property is of particular interest as it has been one of the goals of the biotechnology industry to develop trees that accumulate more extractable lignin. This is because the cellulose

extraction process uses immense amounts of chemicals and energy and is also very taxing to the environment (Pilate et al., 2004).

Molecular analyses have greatly improved our understanding of the biosynthetic pathways underlying cellulose, hemicellulose and lignin biosynthesis. The change from a normal fibre cell consisting of S-layers rich in cellulose, hemicellulose and lignin to G-fibres rich in crystalline cellulose is due to a change in carbon allocation pathways (Andersson-Gunnerås et al., 2006). The switch in carbon metabolism which results in an increase in cellulose and glucosamine biosynthesis as well as pectin degradation gives the opportunity to link gene expression profiles to observed wood properties (Andersson-Gunnerås et al., 2006). Pathways that show decreased activity when the carbon flux is altered by tension wood development are the pentose phosphate pathway, lignin biosynthesis and biosynthesis of cell wall matrix carbohydrates such as xylan (Andersson-Gunnerås et al., 2006). A few key tension wood studies have investigated gene expression profiles linked to a tension wood reaction to identify genes that affect tension wood properties (Andersson Gunnerås, 2005; Andersson-Gunnerås et al., 2006; Paux et al., 2005; Pilate et al., 2004; Qiu et al., 2008). Proteomic analysis of the poplar G-layer supports the tension wood expression profiles generated by Andersson-Gunneras et al. (2006; Kaku et al., 2009).

1.2.2.1 Cellulose biosynthesis

Cellulose microfibrils are synthesised by cellulose synthase complexes made up of rosettes of cellulose synthase (CESA) proteins (Taylor, 2008). The *CesA* gene family of higher plants consist of members that are functionally distinct. The functionally distinct CESA proteins affect either primary cell wall or secondary cell wall cellulose deposition (Scheible and Pauly, 2004). From studies conducted thus far there is an inconsistency in tension wood profiling data with regards to the cellulose biosynthesis machinery. Compared to normal wood the expression of the secondary wall *CesA* genes is either increased or decreased during tension wood formation. In an EST library constructed by Andersson-Gunnerås et al. (2006) none of the *CESA* transcripts were among the most

abundant transcripts. Together with the microarray data of the same study, a significant decrease in expression was observed for *PtrCesA3*, which codes for one of the secondary cell wall CESA proteins. A slight increase in expression was observed for *PtrCesA1* and is supported by Reverse-Transcription qPCR (RT-qPCR) in hybrid aspen (Andersson-Gunnerås et al., 2006, Djerbi et al., 2004). In contrast, Bhandari et al. (2006) observed strong up-regulation of all secondary cell wall *CesA* genes in aspen trees. Consistent with the latter observation are two *Eucalyptus* tension wood studies which showed up-regulation in *CesA1* and *CesA2* expression (Paux et al., 2004; Paux et al., 2005; Qiu et al., 2008). In the study of Paux et al. (2005) the expression level of *CesA1* was increased 6.5 over times that of the control after just 1 week. The third eucalypt *CesA* gene, *CesA3* did not show an altered response during tension wood formation (Qiu et al., 2008). A transcriptional increase in the *CesA* genes was supported by a gus-promoter fusion study which showed that the Poplar *CesA1* promoter was responsive to tension stress (Wu et al., 2000). To date, no clear understanding or explanation exists for the observed *CesA* expression patterns during tension wood formation. Andersson-Gunnerås et al. (2006) suggested that different CESA isoforms may be utilised in the cellulose biosynthetic rosettes during the formation of tension wood. Qiu et al. (2008) suggested that altered expression of CESA isoforms may affect the orientation of cellulose deposition. Alternatively, the cellulose synthase complex response during tension wood formation could be species specific.

Consistent with the notion that during tension wood formation carbon is channelled towards cellulose biosynthesis, an increase in gene expression of sucrose synthase genes *PttSus1* and *PttSus2* was seen. Other carbon allocation transcripts such as fructokinase transcripts and UDP-glucose pyrophosphorylase (UGP2) were also shown to be increased during tension wood formation (Andersson-Gunnerås et al., 2006). Sucrose synthase (SuSy enzymes) break down sucrose to produce fructose and UDP-glucose which is used by the cellulose synthase complex for the synthesis of cellulose (Haigler et al., 2001). Thus during tension wood formation an accumulation of fructose is an expected outcome of increased SuSy activity which would result in an increase of fructose metabolism (Delmer and Haigler, 2002). An increase of fructose metabolism was indeed observed by Andersson-Gunnerås et al. (2006). In agreement with the increased cellulose content in tension

wood, fructose was utilised for cellulose biosynthesis through sucrose regeneration and N-acetylglucosamine biosynthesis. The increased activity of these pathways that lead to cellulose biosynthesis may in part account for the conflicting expression patterns observed for *CesA* genes in these studies. Another short term carbon source implicated by (Andersson-Gunnerås et al., 2006) due to an increase in α-amylase and glucose-6-phosphate translocation transcripts may be starch metabolism.

1.2.2.2 Structural cell wall proteins

Transcripts of several cell wall proteins also show an increase during tension wood formation. Numerous fasciclin-like arabinogalactan proteins (FLAs) show increased expression during tension wood formation. It is notable that most of the FLAs that show an increase in expression during tension wood formation belong to subgroup A. The greater majority of FLAs that show an expression increase in poplar tension wood belong to this group (Andersson-Gunnerås et al., 2006, Lafarguette et al., 2004). The two subgroup A eucalypt FLAs, *EgrFLA1* and *EgrFLA2*, are also among the most up-regulated genes in *Eucalyptus* tension wood formation (Qiu et al., 2008). The highly expressed FLAs may play an important role in the altered microfibril angle found in the G-layer of tension wood (Mellerowicz and Sundberg, 2008, Qiu et al., 2008). The importance of these proteins during tension wood formation is further supported by the detection of these proteins in poplar G-layers during total peptide analysis (Kaku et al., 2009). The specific mechanism of action of FLAs in wood formation still remains unknown, but FLAs are also found to be commonly expressed in the xylem tissues of normal upright trees, which alternatively could indicate a general role in wood development (Andersson-Gunnerås et al., 2006, Yang et al., 2005).

Microtubule-associated proteins (MAPs) that interact with microtubules and are hypothesized to guide the deposition of cellulose the microfibrils are up-regulated during tension wood formation (Andersson-Gunnerås et al., 2006). The cortical microtubule organisation may also be influenced by the action of surface arabinogalactan proteins (Nguema-Ona et al., 2007). The

glycosylphosphatidylinositol (GPI)-anchored wall protein COBRA thought to be involved in microfibril orientation was also shown to be up-regulated in tension wood (Wasteneys and Fujita, 2006). Lastly, kinesins are another group of proteins important in cellulose microfibril orientation and up-regulated in tension wood (Paux et al., 2005, Zhong et al., 2002).

1.2.2.3 Hemicellulose biosynthesis

During normal wood formation the fructokinase product fructose-6-phosphate is generally used for lignin biosynthesis and for hemicellulose biosynthesis (Delmer and Haigler, 2002). During tension wood formation, fructose metabolism is channelled away from hemicellulose biosynthesis resulting in the observed decreased hemicellulose content (Andersson-Gunnerås et al., 2006). Carbohydrate active enzymes (CAZymes) have an essential role in all plant tissues and include many hemicellulose biosynthetic enzymes. CAZymes are utilised in the synthesis of sugars, starch and the biopolymers (pectins, cellulose, hemicellulose) found in the plant cell walls. They are also involved in the glycosylation of cell wall proteins. Geisler-Lee et al. (2006) studied the expression profiles of CAZymes in different poplar tissues and have found that woody tissues show the highest abundance of CAZyme expression. Andersson-Gunnerås et al. (2006) found 37 CAZyme families to be expressed during tension wood formation. This enforces the importance of this group of enzymes in carbon processing during xylogenesis.

In general, enzymes involved in hemicellulose biosynthesis were found to be decreased in tension wood (Andersson-Gunnerås et al., 2006, Paux et al., 2005). These include a decrease in the expression of nucleotide-sugar conversion enzymes, Golgi localised glycosyl transferases and β-xylosidase transcripts. In contrast to the decreased hemicellulose content in tension wood, some transcripts involved in hemicellulose biosynthesis were found to be increased. (1,4)-beta-galactan is exclusively localised to a thin area between the secondary cell wall and the G-layer and could be utilised as a highly specific marker of G layer formation in mechanically stressed trees (Arend, 2008). Other genes whose increased expression levels are associated with hemicellulose biosynthesis include

those coding for enzymes involved in pectin degradation, polysaccharide lyases and xyloglucan endotransglucosylase (XTH) (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007, Qiu et al., 2008). Conflicting with the observed increased expression levels of *XTH* transcripts are the findings of Kaku et al. (2009) who did not detect XTH proteins in isolated poplar G-layers. XTH cuts the xyloglucan backbone so that a new glycosidic bond can be formed with another xyloglucan chain. This allows cellulose microfibrils to 'glide' over each other so that the wall can expand (Cosgrove, 2005). Xyloglucan and (1,4)-beta-galactan are hypothesised to play an important structural role in tension wood by linking the G-layer to the S_2-layer (Arend, 2008, Baba et al., 2009). This enables tension stress to be transferred from the G-layer to the other cell wall layers and the rest of the G-fibre to create a bending moment. Increased activity of XTH in the G-layer is thought to serve the function of repairing G-S_2 bonds that may have been broken during shrinking of the G-layer (Baba et al., 2009, Nishikubo et al., 2007).

1.2.2.4 Lignin biosynthesis

During normal wood formation the excess fructose would be converted and channelled towards lignin biosynthesis through the pentose phosphate pathway and the shikimate pathway (Delmer and Haigler, 2002). During tension wood formation, genes in the shikimate and phenylpropanoid pathways leading to lignin formation are decreased in expression (Andersson Gunnerås, 2005). The activity of all the genes involved in the monolignol biosynthesis pathways during tension wood formation do not always correlate well to the decreased lignin content. *CAD* (Cinnamyl alcohol dehydrogenase) was shown to be weakly down-regulated and both *4CL* (4-coumarate:coenzyme A ligase) and *CCR* (cinnamoyl-CoA reductase) show an initial sharp and transient decrease in expression but soon return back to the same level of expression as the upright control (Andersson-Gunnerås et al., 2006, Paux et al., 2005). Paux et al. (2005) suggest that the decreased lignin content is due to methylation enzymes because of their more down-regulated expression profiles. Both *COMT* (caffeic acid O-methyltransferase) and *CCoAOMT* (caffeoyl-CoA O-methyltransferase) expression are strongly decreased during tension wood formation (Andersson-Gunnerås et al., 2006, Paux et al., 2005). The

observed decrease in expression of *COMT* is supported by proteomic analysis of the G-layer (Kaku et al., 2009). *CCoAOMT* decreases sharply in *Eucalyptus* to reach an expression level less than 2-fold of upright control trees (Paux et al., 2005).

In *Eucalyptus*, two MYB transcription factors, *EgMyb1* and *EgMyb2* bind to MYB binding sites in the promoters of *CCR* and *CAD*. MYB1 and MYB2 play antagonistic roles on lignification genes; MYB1 acts as an activator and MYB2 acts to suppress lignification genes (Boudet et al., 2003). Consistent with its suppressive role in lignin formation, *Myb2* shows increased transcription during tension wood formation (Andersson-Gunnerås et al., 2006, Karpinska et al., 2004, Paux et al., 2004). Another family of transcription factors that govern monolignol biosynthesis, the LIM transcription factors, show an increase in transcription during tension wood formation (Andersson-Gunnerås et al., 2006). This once again is confusing in light of tension wood formation as LIM transcription factors positively stimulate monolignol biosynthesis in transgenic plants (Kawaoka et al., 2000). This may indicate that LIM transcription factors have an alternative function during tension wood formation.

1.2.3 Hormonal regulation of tension wood formation

Auxin, ethylene and gibberellic acid (GA) are essential to cambial differentiation and growth. Several studies investigating the involvement of these hormones in tension wood formation have been performed. Knowledge on which hormones promote or inhibit tension wood formation is being gathered but a clear concurrence on whether all of these hormones play a direct role in tension wood formation is yet to be made (Du and Yamamoto, 2007).

1.2.3.1 Auxin

Auxin genes exhibit specific tissue expression patterns that change during tension wood formation (Moyle et al., 2002). Auxin serves as a positional signal for wood development by forming a

concentration gradient in the developing vascular tissues (Benjamins and Scheres, 2008). High levels of auxin regulate wood development by keeping the cambium in an undifferentiated state (this prevents the depletion of the meristematic stem cells), whereas low levels of auxin induce cambial division. *Aux/IAA* genes are auxin induced and high levels of *Aux/IAA* expression in tissues are correlated with an increased sensitivity to auxin (Abel and Theologis, 1996). The role of auxin in tension wood formation was demonstrated in a study by Moyle et al. (2002). Two hybrid aspen *Aux/IAA* genes, *PttIAA1* and *PttIAA2*, were rapidly decreased in expression upon the induction of tension wood formation indicating a decrease in auxin sensitivity. This indicates increased cambial division, consistent with new xylem development during tension wood formation. In support of the previous observation, *Eucalyptus Aux/IAA* genes also show decreased expression during tension wood formation (Paux et al., 2005). These studies may suggest that Angiosperms share a conserved mechanism with regards to the control of tension wood formation. Other genes found to be down-regulated during tension wood formation are auxin influx genes and auxin responsive proteins (Andersson-Gunnerås et al., 2006).

1.2.3.2 Ethylene

Ethylene interacts with the auxin signalling pathway and stimulates cell division in the cambial meristem for the induction of fibre cell differentiation (Love et al., 2009). An increase in bark and xylem formation is observed when ethrel (an ethylene releasing compound) is applied to trees (Junghans et al., 2004). Ethylene is synthesised through the action of two main enzymes; 1-aminocyclopropane-1carcoxylic acid (ACC) synthase and ACC oxidase (Kende, 1989). A high level of ACC oxidase activity has been found to be correlated with high levels of ethylene (Andersson-Gunnerås et al., 2003). Elevated levels of ethylene are observed in tension wood (Du and Yamamoto, 2003). This observation is paired with an increase in the transcription of ACC oxidase, ethylene receptors and signalling proteins during tension wood formation (Andersson-Gunnerås et al., 2006). In a study by Jiang et al. (2009), ethylene synthesis and action was inhibited to investigate the role of ethylene in tension wood formation. A significant decrease in the number of wood fibres was

observed in both the upper and lower sections of the bent *Fraxinus mandshurica* Rupr. var. *japonica* Maxim. stems. Tension wood formation however was unaffected. This indicates the important role of ethylene in xylem production, but also that ethylene may not be directly involved in the formation of G-layers.

1.2.3.3 Gibberellic acid

Gibberellins have been proposed to be required for normal wood formation and tension wood formation. Gibberellins that are applied exogenously demonstrate the ability of this group of hormones to affect the differentiation and lignification of wood cell fibres. Gibberellins also have the ability to induce tension wood formation and to accelerate the formation of well developed tension wood (Aloni, 1979, 1990; Baba et al., 2000; Funada et al., 2008; Jiang et al., 2008). The study of Funada et al. (2008) demonstrated how gibberellins applied to the vertical stems of angiosperm trees are able to induce the formation of tension wood. In drooping cherry tree mutants, vertical shoot growth was restored after the application of the gibberellin GA_3. During this treatment, tension wood formation was also observed (Baba et al., 2000). The study by Jiang et al. (2008) indicated that GA_3 and GA_4 were sufficient for a gravitropic response but not necessary for the formation of a G-layer. These two gibberellins affected the quantity of wood production more than the differentiation of the G-layer which is most likely controlled by other hormones. The formation of tension wood can also be decreased through the use of a GA biosynthesis blocker, uniconazole-P (Du and Yamamoto, 2007). This confirms the role that gibberellins plays during tension wood formation.

1.3. TAG BASED METHODS USED TO STUDY TRANSCRIPTOME CHANGES DURING WOOD FORMATION

The development, identity and role of different tissue types in multicellular organisms are in part determined by intricate differences in transcriptional activity of large numbers of genes. This includes

the variation of expressed transcripts and degree of expression of those transcripts. In addition, diversity of the 5' and 3' ends of genes serve as a regulatory mechanism which in turn controls mRNA turnover, translation and sub-cellular localisation of the expressed protein (Gustincich et al., 2006). The functional complexity of differing transcriptomes has increased over the years with the discovery of a range of expressed small RNAs. It is now understood that most of the genome is constantly transcribed at a very low level and a large number of genes are transcribed in a bidirectional manner (Gingeras, 2007). Different technologies are available to study transcriptional complexity of plant tissues. Among these are microarrays and DNA sequencing based expression profiling methods. The recently available ultra-high-throughput second generation DNA sequencing platforms have changed the face of genomic research not only by the magnitude of data output but also by an unparalleled exposure of the richness of eukaryotic genomes and transcriptomes (Morozova and Marra, 2008).

1.3.1 Expressed Sequence Tag (EST) analysis

Sanger sequencing, developed by Sanger et al. (1977) has been the dominant sequencing method used to decipher the DNA sequence of genes for over 30 years. Transcriptional analysis through cDNA sequencing provided the nucleotide sequence of genes investigated in different tissue and developmental stages and enabled novel gene discovery (Adams et al., 1991). EST sequencing has made a significant contribution to elucidating the process of secondary xylem formation in forest tree species such as pine, poplar and eucalypt by adding to the knowledge of genes recognised to be involved in this developmental process.

Sterky et al. (1998) obtained 3700 unique ESTs from poplar xylem and stated that 10% of the unique ESTs represented novel xylem differentiation genes. Through EST sequencing, Allona et al. (1998) identified a subset of novel pine ESTs in addition to detecting cellulose, lignin and other cell wall biosynthesis genes. Li et al. (2009) investigated genes involved in early and late wood development in pine and generated 3304 unigenes from 6389 ESTs. *Eucalyptus* genomics has also

drawn benefit from EST sequencing. Paux et al. (2004) detected ESTs of transcripts preferentially expressed in xylem as opposed to leaf tissue. Kirst et al. (2004) used EST information together with gene expression data and genetic map information in a genetical genomics approach to elucidate genetic and metabolic networks that regulate growth variation in *Eucalyptus*.

EST and full length cDNA sequencing have played a significant role in the annotation of previously sequenced genomes and will continue to have a valuable role in the annotation of genomes that are yet to be sequenced. ESTs are also useful in gene structure prediction (Jiang and Jacob, 1998, Kan et al., 2001), to investigate alternative splicing (Brett et al., 2000) and for SNP discovery and characterisation (Picoult-Newberg et al., 1999, Useche et al., 2001). EST sequencing however is generally not suitable to detect low abundance transcripts due to insufficient depth of coverage and has to be used in conjunction with other techniques to investigate an organism's transcriptional complexity (Sun et al., 2004).

1.3.2 Serial Analysis of Gene Expression (SAGE)

In 1995 Velculescu et al. (1995) described a sensitive method capable of detecting lower abundance transcripts. Serial Analysis of Gene Expression (SAGE) is a method which sequences and quantifies short transcript-specific tags cloned into plasmid vectors (Figure 1.4). In short, the tags are generated through two subsequent restriction enzyme digestions of double stranded cDNA. The first enzyme is generally a four base pair cutter (NlaIII or DpnII) and is known as the anchoring enzyme. After digestion, an adapter containing a type IIS restriction enzyme recognition site is ligated to the fragment. The type IIS tagging enzyme, BsmFI, digests distal from its recognition site, generating a short 14bp tag specific to the original transcript. The tags are then concatemerised and cloned into a plasmid vector. These concatamers consist of multiple short tags that are ligated together. Each of the 22-50 concatenated tags represents a distinct transcript and gives information on which genes are expressed and the level at which they are expressed (Velculescu et al., 1995). More recently, different tagging restriction enzymes have been used to create longer tags. LongSAGE uses a Type IIS

restriction enzyme, MmeI, to produce tags up to 21 bp and SuperSAGE uses a Type III restriction enzyme, EcoP15I, to produce a 26 bp tag (Matsumura et al., 2003, Saha et al., 2002).

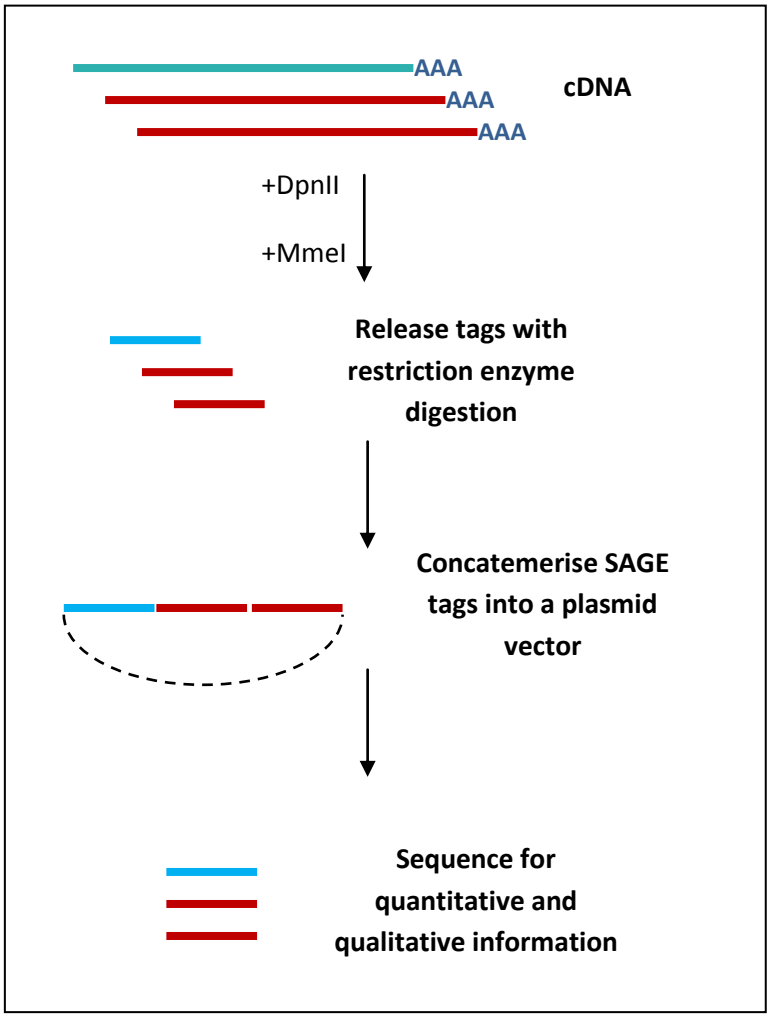


Figure 1.4: A schematic representation of the generation of SAGE data. (Redrawn from Wang, 2007). Tags are generated by digesting cDNA with an anchoring enzyme such as DpnII followed by a tagging enzyme such as MmeI. Tag sequences are obtained by sequencing concatamers of tags cloned into a plasmid vector and quantified by counting the number of unique tags representative of a gene.

Despite laborious, costly cloning and sequencing steps, SAGE has improved the number of genes being profiled from a few hundred to tens of thousands of genes. For example, in the tension wood microarray study by Paux et al. (2005) a total of 231 genes preferentially expressed in xylem

were profiled while approximately 11 000 genes were profiled by Bao et al. (2005) in a SAGE comparative study on super-hybrid rice (*Oryza sativa).* A SAGE study on juvenile *Eucalyptus grandis* produced expression profiles for 3066 genes of which 445 represented genes important to actively developing xylem tissue (de Carvalho et al., 2008).

Another tag-based method that is available is massively parallel signature sequencing (MPSS). This technology involves the cloning of unique cDNA fragments onto microbeads followed by construction and sequencing of signature tags on the beads. This generates 17-20bp signature sequences and is capable of analyzing millions of tags at a time (Gowda et al., 2006, Meyers et al., 2004). A study by Hene et al. (2007) specifically investigated the complexity of transcripts sampled during MPSS compared to that of LongSAGE. The study demonstrated that even though the LongSAGE library was half the size of the MPSS library, the LongSAGE library was able to detect twice as many unique transcripts. After a combination of three replicate MPSS libraries, the number of unique transcripts sampled totalled only 73% of the single LongSAGE library. MPSS therefore produces a larger sequence output in one experiment but suffers from a lack of complexity in the transcripts sampled (Hanriot et al., 2008, Siddiqui et al., 2006).

1.3.3 Digital Gene Expression (DGE)

Second generation DNA sequencing technologies have revolutionised genomic studies. Data generation of projects that were previously too expensive can now be completed in weeks instead of years. These technologies are constantly being improved and the throughput of these technologies is ever increasing (Mardis, 2008, Zhou et al., 2010). The sequencing by synthesis (SBS) technology of Illumina (San Diego, California, USA) started with a sequence output of 1.5-3 Gbp per instrument (Genome Analyzer, GA) run and a short sequence read length of 36 bp (Barski et al., 2007, Ossowski et al., 2008). In 2009, the GAII generated 50 Gbp per run with read lengths of up to 100 bp (Simpson et al., 2009). Using paired-end sequencing the GAIIx is currently (end 2010) able to generate ≈190 Gbp per run with read lengths of up to 150 bp (www.illumina.com). The high-throughput of second generation DNA sequencing technologies enables gene expression profiling by SAGE to be scaled up

to ultra-deep, whole-transcriptome profiling and eliminates the need to concatenate and clone multiple tags before sequencing. The depth of sequencing is increased by anchoring the fragment that is selected for sequencing at a specific restriction enzyme site on the cDNA (Eveland et al., 2008). Eveland et al. (2008) sequenced 3'-UTR signature tags from drought stressed *Zea mays* ovaries using a Roche GS 20 454-FLX Life Sciences sequencer (Branford, Connecticut, USA). Approximately 330,000 sequence tags identified approximately 15,000 unique transcripts represented by at least two tags. Profiling using the information rich 3'-UTR enabled successful resolution of the expression profiles of gene family members and allele-specific transcripts. Hanriot et al. (2008) sequenced longSAGE libraries prepared from mouse hypothalamus tissue with conventional Sanger sequencing and Illumina sequencing. The study demonstrated that Illumina sequencing combined with LongSAGE gives a reliable, deep sampling of the transcriptome overcoming the problem of low complexity associated with MPSS.

Illumina developed their own SAGE based expression profiling application and refer to it as Digital Gene Expression (DGE). DGE is therefore an improved version of Illumina's older MPSS technology ('t Hoen et al., 2008). DGE libraries are constructed by using the same method as the LongSAGE method, without concatenating or cloning the tags. Bead captured mRNA transcripts are converted to cDNA and digested with the appropriate LongSAGE restriction enzymes. Upon digestion with the tagging MmeI restriction enzyme the tag is released from the magnetic bead. After each restriction enzyme digestion, adapter sequences are ligated to both ends of the tag (Figure 1.5). This enables the molecules to be bound to a glass surface containing oligonucleotides complementary to the tag adapters. This is followed by solid-phase amplification and sequencing-by-synthesis with Illumina's Genome Analyser instrument.

't Hoen et al. (2008) was the first to make use of DGE by contrasting the hippocampal expression profiles of wild-type mice and mice transgenic for δ-C-doublecortin-like kinase. Using very stringent statistical criteria differential expression was detected for 3179 genes. This study also detected antisense transcription and alternative polyadenylation in 47% of the detected genes. In addition, 't Hoen et al. (2008) compared the results from their study to that of five previous microarray

studies. DGE expression profiling was found to be sensitive to smaller changes in gene expression levels and profiled the expression of genes over a wider dynamic range.

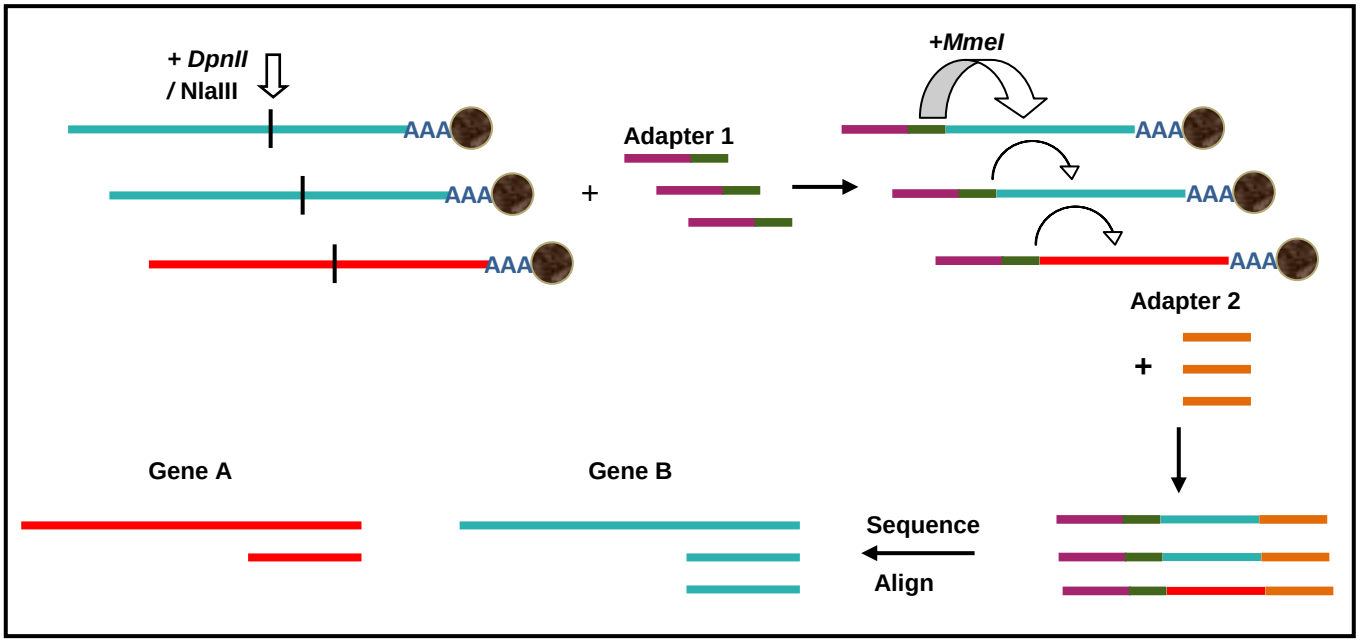


Figure 1.5: A schematic representation of Digital Gene Expression (DGE). Bead captured mRNA is converted to cDNA and digested with an anchoring restriction enzyme. The adapter that is ligated to the 5' end of the cDNA fragment contains the restriction enzyme site for the tagging enzyme MmeI. MmeI digestion releases the fragment from the magnetic bead to enable ligation of the second adapter to the 3' end of the fragment. Sequenced tags are then mapped to the reference sequence to generate expression profiles for detected genes.

A study by Hegedus et al. (2009) used DGE to investigate the transcriptional response to mycobacterium infection in zebrafish, a model for human tuberculosis. The experiment generated more than 5 million tags per sample which detected 70% of all genes represented in zebrafish transcript databases. Of the detected transcripts, 5000 were found to be differentially expressed. This represented approximately 2% of the transcripts in databases. Thirty transcripts produced alternative splice forms specific to the mycobacterium infection response. These transcript variants showed differential expression between the infected and control samples. In addition, 29 putative novel splice

variants were identified. The DGE approach once again showed an improved detection of transcripts over a wider range when compared microarray studies previously performed on this topic (Meijer et al., 2005).

The study by Ko et al. (2009) was the first study to employ DGE in plants. DGE was used to confirm expression profiles generated by ATH1 GeneChip data generated from *Arabidopsis* rosette leaves. The ATH1 GeneChip and DGE data identified a group of transcription factors that shows expression preceding and coinciding with secondary cell wall biosynthetic genes. The study used the ectopic expression of the master regulator MYB46 in *Arabidopsis* to identify three other transcription factors, AtC3H14, MYB52 and MYB63 that activate secondary cell wall biosynthesis.

During the year 2010, DGE profiling was applied to a diverse number of fields such as virology and parasitology (Xiao et al., 2010, David et al., 2010), equine physiology and disease (McGivney et al., 2010, Serteyn et al., 2010), mycology (Liu et al., 2010) and primate evolution (Babbitt et al., 2010). An important publication for the plant research community is a maize meristem differentiation study by Eveland et al. (2010). Twenty seven million tags detected expression of 37,117 genes and revealed 660 genes significantly differentially expressed between wild type maize and *RAMOSA 3* (*ra3*) mutant maize plants. This study also detected and quantified the expression of 14 primary microRNA transcripts originating from 10 different microRNA families. These findings demonstrated the ability of DGE to complement the expression profiles linked to protein coding genes with expression profiles of other expressed transcript species.

1.3.3.1 Data analysis

The management, analysis and visualisation of the enormous amounts of raw data produced by second generation sequencing techniques are emphasised as a hurdle to successfully complete projects (Horner et al., 2009). This is also reflected by the delay in DGE publications mentioned above. The large amounts of data stored in databases for analysis is enabling a surge in bioinformatics with the possible emergence of several projects from single wet-laboratory experiments (Barrett et al., 2005,

Minic et al., 2009). Each application of second generation sequencing generates a specific type of data that needs to be dealt with accordingly.

The nature of data produced by DGE may be established by examining the type of data produced by SAGE studies, from which DGE originated, and by examining pioneering DGE studies. The largest difference between SAGE and DGE data lies in the raw sequence output. A SAGE experiment generates a series of ditags set in long DNA concatamers (Lash et al., 2000). Tags generated from a DGE experiment are not concatenated but are produced as single sequences. Thus DGE tags only have to be sorted, counted and filtered for artefacts ('t Hoen et al., 2008). DGE studies make use of custom Perl, C^{++}, or linux shell scripts to trim adapter sequences and filter for tags with the correct count ('t Hoen et al., 2008, Hegedus et al., 2009, McGivney et al., 2010). As mentioned before, the output of second generation sequencing greatly supersedes that of conventional Sanger sequencing and this aspect greatly improves the statistical power and reproducibility of DGE (Dinel et al., 2005). Published DGE experiments have produced 2.4 – 27 million sequence tags per sample (Hanriot et al., 2008, Hegedus et al., 2009, 't Hoen et al., 2008, Eveland et al., 2010). The dynamic range of DGE data spans four to five orders of magnitude (Hegedus et al., 2009, 't Hoen et al., 2008, Eveland et al., 2010). Thus the detection of transcripts span from a few copies to a few hundred thousand copies.

Even though some studies (Chen et al., 2002, Wang, 2003) have suggested that sequencing errors are low, an assessment made by Keime et al. (2007) suggested that SAGE generated data may be “noisier” in terms of sequencing errors and artefacts than previously thought. The SAGE dataset generated by Faunes et al. (2009) produced 69% single copy tags and in the DGE dataset produced by Hegedus et al. (2009) 70-80% of unique sequences consisted of single copy tags which are regarded as tags containing sequencing errors. DGE data sets have been filtered differently to reduce the number of erroneous tags incorporated into down-stream analyses. Filtering too stringent with a high raw count will result in discarding expression profiles of lowly expressed genes whereas filtering with a raw count too low will result in the false positive expression of genes. The minimum raw count per signature tag has been set at 1 (Xiao et al., 2010), 2 (Ko et al., 2009), 10 (Pewzner-Jung et al., 2010, Eveland et al., 2010) or 20 (David et al., 2010) to be used in analyses in previous DGE studies.

Mapping short tag sequences back to a reference sequence remains one of the most difficult steps during the analysis of SAGE profiling data. Only 50-70% of 14 bp SAGE tags generally map to known genes (Wang, 2007). Faunes et al. (2009) mapped half of the generated SAGE tags with confidence and only 3% of all tags mapped at a single location within the genome. Improved mapping efficiency is obtained with DGE tags due to the increased sequencing depth of second generation sequencing and to the 20-21 bp tag length as opposed to that of conventional 14 bp SAGE tags (Dinel et al., 2005, Wahl et al., 2005). Eveland et al. (2010) mapped up to 60% of the DGE tags generated in their study to the maize B73 genome reference sequence. In general, allowing for one or two mismatches has resulted in the mapping of 70-85% of DGE tags to a reference sequence in various studies (Hegedus et al., 2009, 't Hoen et al., 2008, Eveland et al., 2010, Babbitt et al., 2010).

Processed DGE data looks similar to that of SAGE data: each unique tag represents an expressed gene/transcript and the number of times each unique tag is observed represents its expression value (Velculescu et al., 1995). It is expected that the tag would map to the 3'-most cut site of the anchoring enzyme, assuming that complete restriction enzyme digestion takes place during DGE library preparation. Variations on this observation enable additional inferences to be made from the data giving a richer picture of the sampled transcriptome. For instance as found by Hegedus et al. (2009) and 't Hoen et al. (2008), different unique tags mapping to separate locations on the same gene indicates alternative splicing. In addition, differences in allelic expression may be observed when two unique tags differing in sequence map to a location that contains a SNP (Eveland et al., 2008). An observation made in a SAGE study by Keime et al. (2007) was tags mapping to the negative strand indicating antisense transcription. 't Hoen et al. (2008) observed antisense transcription for 51% of the genes detected. Tags mapping to intron locations or to a junction between an intron and an exon have also been observed ('t Hoen et al., 2008, Keime et al., 2007). Babbitt et al. (2010) describe adding the counts of tags sequenced from multiple sites within the same transcript.

Published DGE studies make use of different approaches to identify statistically significantly differentially expressed genes. Pewzner-Jung et al. (2010) simply employed a 1.5 fold change cut off to the 5% genes most up- or down-regulated. To assess the significance of differential expression David et al. (2010) used a minimum change of 2 fold in expression in addition to a P-value calculated

by Fisher's noncentral hypergeometric distribution test. Hegedus et al. (2009) utilised an online differential expression analysis tool described by Lash et al. (2006). SAGEmap is an online open access SAGE data repository with several analysis tools. It has an online differential expression tool which makes use of a Bayesian statistical approach developed by Chen et al. (1998). To detect significant differential expression is a challenge as most DGE studies have not involved biologically replicated samples (Xiao et al., 2010, Liu et al., 2010, David et al., 2010, Hegedus et al., 2009, Pewzner-Jung et al., 2010, Wang et al., 2010). Four studies (Wang et al., 2010; Pewzner-Jung et al., 2010; Liu et al., 2010; Xiao et al., 2010) employed the statistical testing method of Audic and Claverie (1997). This test, initially used for significance testing of SAGE data, assesses the probability of the differential expression of a gene in two different libraries. This test along with five other statistical tests has been incorporated into an online tool, IDEG6, described by Romualdi et al. (2003). Significance testing may be performed with the option of applying a Bonferroni correction. In addition any of the selected tests are performed simultaneously and the results are displayed in a colour coded table which may be extracted in text format. Providing gene accession numbers links the results to GenBank, UniGene or LocusLink.

1.4. Conclusion

During tension wood formation in hardwoods, reprogramming of gene expression takes place throughout the transcriptome. This results in the alteration of important physical and chemical characteristics. Of particular interest is the increase in crystalline cellulose deposition in the development of the secondary cell wall. Tension wood is therefore a good gene discovery model for wood formation that can be compared to normal wood formation. This will enable the identification and verification of regulatory and structural genes involved in the most important aspects of wood formation. Studying the transcriptional differences between tension wood and normal wood formation will not only add to the current knowledge of cellulose biosynthesis and in a broader context wood formation as a whole, but also provide support for future biotechnology targets aimed at improving wood quality traits.

From literature there are still many uncertain points pertaining to the development of tension wood formation. Some species seem to exhibit different characteristics to what is known as the norm during tension wood formation. Poplar (Andersson-Gunnerås et al., 2006) and eucalypt trees (Paux et al., 2005) were found to produce G-layers during tension wood formation, while at least one study (Qiu et al., 2008) found that eucalypts did not produce G-layers. Specific gene expression patterns underlie the phenotype of a specific trait. It is therefore valid to expect that the gene expression patterns observed during poplar tension wood formation (Andersson-Gunnerås et al., 2006) may not coincide with those in *Eucalyptus* observed by Paux et al. (2005). An example of this is the uncertainty of whether the expression of the *CesA* genes are truly up- or down-regulated during tension wood formation. Furthermore, there are gaps in tension wood expression profile information especially for *Eucalyptus*. The bending experiment of Paux et al. (2005) only spanned one week in time. This has given adequate information on tension wood induction, but may not portray important later tension wood formation expression profiles. The study by Andersson-Gunnerås et al. (2006) does not involve a time series of expression profiles in tension wood and only investigates tension wood formation at three weeks after bending.

Illumina's DGE profiling application is able to detect transcripts expressed over a range of five to six orders of magnitude, but is also sensitive enough to identify low abundant transcripts and very small changes in transcript abundance. DGE analysis currently produces 5 to 12 million sequence tags per sample, which represents a very deep sampling of transcripts in the corresponding tissue or cell type. The data obtained from a DGE experiment is more versatile than other expression profiling approaches, which only give information on changes in gene expression profiles. From DGE data inferences can be made about the transcript abundance of expressed alleles, alternative splice variants and antisense transcription. Expression profiling with DGE is more advantageous than microarrays as the number of genes profiled is not limited to a chosen subset, but to transcripts that contain the anchoring restriction enzyme cut site. This allows for near whole-transcriptome profiling and for the profiling of novel transcripts.

The current challenge of genomic studies lies not in the experimental production of data but in the processing, analysis and presentation of the data produced. This is in part due to the enormous

data output obtained from next-generation DNA sequencing technologies. The similarities shared by DGE and SAGE can be used as a guideline for the processing of DGE data. Specific software for the comprehensive analysis of DGE sequence data is not yet publically available. From observations made in SAGE and DGE studies, the data generated with DGE may be more complex than expected. DGE data does not only translate into an expression profile from a number of tags that map to certain genes, but portray other processes such as antisense transcription ('t Hoen et al., 2008). It is therefore important to make a thorough investigation of the data generated by DGE to comprehend the nature of the data fully. This will allow correct processing of the data so that unexpected and interesting cases may be noted for further analysis. An interactive analysis pipeline that will enable the integration of the processing, analysis and presentation of raw DGE data is therefore a crucial requirement. The analysis output from the pipeline should be modular so that it can be updated with new experimental data. The new experimental data should also be subject to the re-examination with previously analysed data sets. Furthermore, the differentially expressed datasets should be integrated with available databases so that subsequent analysis is eased.

Expression profiling of the *Eucalyptus* xylem transcriptome during tension wood formation is valuable for the identification of genes involved in wood development and in particular cellulose biosynthesis. DGE profiling will produce an ultra deep sampling of the *Eucalyptus* transcriptome during tension wood formation. Expression profiling of a hybrid genotype will enable the data to be reused for the characterisation of allelic expression patterns in future studies. This will be feasible once pure *E. grandis* and *E. urophylla* transcriptomes become available. The resulting preferential allelic expression patterns may illuminate transcriptome differences between pure species and hybrids of *Eucalyptus* species.

1.5 References

ABEL, S. & THEOLOGIS, A. (1996) Early genes and auxin action. *Plant Physiology*, 111**,** 9-17.

ADAMS, M. D., KELLEY, J. M., GOCAYNE, J. D., DUBNICK, M., POLYMEROPOULOS, M. H., XIAO, H., MERRIL, C. R., WU, A., OLDE, B., MORENO, R. F. & ET, A. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252**,** 1651-1656.

ALONI, R. (1979) Role of Auxin and gibberellin in differentiation of primary phloem fibers. *Plant Physiology*, 63**,** 609-614.

ALONI, R., TOLLIER, M. T. & MONTIES, B. (1990) The role of auxin and gibberellin in controlling lignin formation in primary phloem fibers and in xylem of *Coleus blumei* stems. *Plant Physiology*, 94**,** 1743-1747.

ANDERSSON-GUNNERÅS, S., HELLGREN, J. M., BJÖRKLUND, S., REGAN, S., MORITZ, T. & SUNDBERG, B. (2003) Asymmetric expression of a poplar ACC oxidase controls ethylene production during gravitational induction of tension wood. *The Plant Journal*, 34**,** 339-349.

ANDERSSON-GUNNERÅS, S., MELLEROWICZ, E. J., LOVE, J., SEGERMAN, B., OHMIYA, Y., COUTINHO, P. M., NILSSON, P., HENRISSAT, B., MORITZ, T. & SUNDBERG, B. (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *The Plant Journal*, 45**,** 144-165.

ANDERSSON GUNNERÅS, S. (2005) Wood formation and transcript analysis with focus on tension wood and ethylene biology. *Department of Forest Genetics and Plant Physiology*. Umeå, Swedish University of Agricultural Sciences.

AREND, M. (2008) Immunolocalization of (1,4)-beta-galactan in tension wood fibers of poplar. *Tree Physiology*, 28**,** 1263-1267.

ASMANN, Y., KLEE, E., THOMPSON, E. A., PEREZ, E., MIDDHA, S., OBERG, A., THERNEAU, T., SMITH, D., POLAND, G., WIEBEN, E. & KOCHER, J.-P. (2009) 3' tag

digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genomics,* 10**,** 531.

AUDIC, S. & CLAVERIE, J.-M. (1997) The significance of digital gene expression profiles. *Genome research,* 7**,** 986-995.

BABA, K. I., ASADA, T. & HAYASHI, T. (2000) Relation between developmental changes on anatomical structure and on protein pattern in differentiating xylem of tension wood. *Journal of Wood Science,* 46**,** 1-7.

BABA, K. I., PARK, Y. W., KAKU, T., KAIDA, R., TAKEUCHI, M., YOSHIDA, M., HOSOO, Y., OJIO, Y., OKUYAMA, T., TANIGUCHI, T., OHMIYA, Y., KONDO, T., SHANI, Z., SHOSEYOV, O., AWANO, T., SERADA, S., NORIOKA, N., NORIOKA, S. & HAYASHI, T. (2009) Xyloglucan for generating tensile stress to bend tree stem. *Molecular Plant,* 2**,** 893-903.

BABBITT, C. C., FEDRIGO, O., PFEFFERLE, A. D., BOYLE, A. P., HORVATH, J. E., FUREY, T. S. & WRAY, G. A. (2010) Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biology and Evolution,* 2**,** 67-79.

BALA, G., CALDEIRA, K., WICKETT, M., PHILLIPS, T. J., LOBELL, D. B., DELIRE, C. & MIRIN, A. (2007) Combined climate and carbon-cycle effects of large-scale deforestation. *Proceedings of the National Academy of Sciences,* 104**,** 6550-6555.

BAMBER, R. K. (2001) A general theory for the origin of growth stresses in reaction wood: How trees stay upright. *IAWA Journal,* 22**,** 205-212.

BARRETT, T., TROUP, D. B., WILHITE, S. E., LEDOUX, P., RUDNEV, D., EVANGELISTA, C., KIM, I. F., SOBOLEVA, A., TOMASHEVSKY, M. & EDGAR, R. (2005) NCBI GEO: mining tens of millions of expression profiles - database and tools update. *Nucleic Acids Research,* 35**,** D760-D765.

BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T.-Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-837.

BENJAMINS, R. & SCHERES, B. (2008) Auxin: the looping star in plant development. *Annual Review of Plant Biology,* 59**,** 443-465.

BOUDET, A. M., 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.

BOWLING, A. J. & VAUGHN, K. C. (2008) Immunocytochemical characterization of tension wood: Gelatinous fibers contain more than just cellulose. *Am. J. Bot.,* 95**,** 655-663.

BRETT, D., HANKE, J., LEHMANN, G., HAASE, S., DELBRÜCK, S., KRUEGER, S., REICH, J. & BORK, P. (2000) EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Letters,* 474**,** 83-86.

BUTLER, R. A. & LAURANCE, W. F. (2008) New strategies for conserving tropical forests. *Trends in Ecology & Evolution,* 23**,** 469-472.

CHEN, H., CENTOLA, M., ALTSCHUL, S. F. & METZGER, H. (1998) Characterization of gene expression in resting and activated mast cells. *The Journal of Experimental Medicine,* 188**,** 1657-1668.

CHEN, J., SUN, M., LEE, S., ZHOU, G., ROWLEY, J. D. & WANG, S. M. (2002) Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 12257-12262.

CLAIR, B., ALMÉRAS, T., YAMAMOTO, H., OKUYAMA, T. & SUGIYAMA, J. (2006a) Mechanical behavior of cellulose microfibrils in tension wood, in relation with maturation stress generation. *Biophysical Journal,* 91**,** 1128-1135.

CLAIR, B., RUELLE, J., BEAUCHENE, J., PREVOST, M. F. & FOURNIER, M. (2006b) Tension wood and opposite wood in 21 tropical rain forest species - occurence and efficiency of the G-layer. *IAWA Journal,* 27**,** 329-338.

CLAIR, B., THIBAUT, B. & SUGIYAMA, J. (2005) On the detachment of the gelatinous layer in tension wood fiber. *Journal of Wood Science,* 51**,** 218-221.

COSGROVE, D. J. (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol,* 6**,** 850-861.

COUTAND, C., JERONIMIDIS, G., CHANSON, B. & LOUP, C. (2004) Comparison of mechanical properties of tension and opposite wood in *Populus*. *Wood Science and Technology,* 38**,** 11-24.

CRONN, R., LISTON, A., PARKS, M., GERNANDT, D., SHEN, R. & MOCKLER, T. (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research,* 36**,** e122.

DARLEY, C. P., FORRESTER, A. M. & MCQUEEN-MASON, S. J. (2001) The molecular basis of plant cell wall extension. *Plant Molecular Biology,* 47**,** 179-195.

DAVID, J.-P., COISSAC, E., MELODELIMA, C., POUPARDIN, R., RIAZ, M., CHANDOR-PROUST, A. & REYNAUD, S. (2010) Transcriptome response to pollutants and insecticides in the dengue vector *Aedes aegypti* using next-generation sequencing technology. *BMC Genomics,* 11**,** 216.

DE CARVALHO, M. C. D. C. G., CALDAS, D. G. G., CARNEIRO, R. T., MOON, D. H., SALVATIERRA, G. R., FRANCESCHINI, L. M., DE ANDRADE, A., CELEDON, P. A. F., ODA, S. & LABATE, C. A. (2008) SAGE transcript profiling of the juvenile cambial region of *Eucalyptus grandis*. *Tree Physiology,* 28**,** 905-919.

DELMER, D. P. & HAIGLER, C. H. (2002) The regulation of metabolic flux to cellulose, a major sink for carbon in plants. *Metabolic Engineering,* 4**,** 22-28.

DINEL, S., BOLDUC, C., BELLEAU, P., BOIVIN, A., YOSHIOKA, M., CALVO, E., PIEDBOEUF, B., SNYDER, E. E., LABRIE, F. & ST-AMAND, J. (2005) Reproducibility, bioinformatic analysis and power of the SAGE method to evaluate changes in transcriptome. *Nucleic Acids Research,* 33**,** e26.

DJERBI, S., ASPEBORG, H., NILSSON, P., SUNDBERG, B., MELLEROWICZ, E., BLOMQVIST, K. & TEERI, T. T. (2004) Identification and expression analysis of genes encoding putative cellulose synthases (CesA) in the hybrid aspen, *Populus tremula* (L.) × *P. tremuloides* (Michx.). *Cellulose,* 11**,** 301-312.

DU, S. & YAMAMOTO, F. (2003) Ethylene evolution changes in the stems of *Metasequoia glyptostroboides* and *Aesculus turbinata* seedlings in relation to gravity-induced reaction wood formation. *Trees - Structure and Function,* 17**,** 522-528.

DU, S. & YAMAMOTO, F. (2007) An overview of the biology of reaction wood formation. *Journal of Integrative Plant Biology,* 49**,** 131-143.

EMRICH, S. J., BARBAZUK, W. B., LI, L. & SCHNABLE, P. S. (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research,* 17**,** 69-73.

EVELAND, A. L., MCCARTY, D. R. & KOCH, K. E. (2008) Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiology,* 146**,** 32-44.

EVELAND, A. L., SATOH-NAGASAWA, N., GOLDSHMIDT, A., MEYER, S., BEATTY, M., SAKAI, H., WARE, D. & JACKSON, D. (2010) Digital gene expression signatures for maize development. *Plant Physiology,* 154**,** 1024-1039.

FANG, C., CLAIR, B., GRILL, J. & LIU, S. (2008) Growth stresses are highly controlled by the amount of G-layer in poplar tension wood. *IAWA Journal,* 29**,** 237-246.

FAUNES, F., SANCHEZ, N., CASTELLANOS, J., VERGARA, I., MELO, F. & LARRAIN, J. (2009) Identification of novel transcripts with differential dorso-ventral expression in *Xenopus gastrula* using serial analysis of gene expression. *Genome Biology,* 10**,** R15.

FEARNSIDE, P. M. (2008) Deforestation in brazilian amazonia and global warming. *Annals of Arid Zone,* 47**,** 1-20.

FENNING, T. M. & GERSHENZON, J. (2002) Where will the wood come from? Plantation forests and the role of biotechnology. *Trends in Biotechnology,* 20**,** 291-296.

FISHER, J. B. & STEVENSON, J. W. (1981) Occurrence of reaction wood in branches of dicotyledons and its role in tree architecture. *Botanical Gazette,* 142**,** 82-95.

FUNADA, R., MIURA, T., SHIMIZU, Y., KINASE, T., NAKABA, S., KUBO, T. & SANO, Y. (2008) Gibberellin-induced formation of tension wood in angiosperm trees. *Planta,* 227**,** 1409-1414.

GEIST, H. J. & LAMBIN, E. F. (2002) Proximate causes and underlying driving forces of tropical deforestation. *BioScience,* 52**,** 143-150.

GINGERAS, T. R. (2007) Origin of phenotypes: genes and transcripts. *Genome Research,* 17**,** 682-690.

GORIŠEK, Ž. & STRAŽE, A. (2006) Sorption and swelling characteristics of normal and tension beech wood (*Fagus sylvatica* L.). *Wood Structure and Properties,* 1**,** 227-231.

GOSWAMI, L., DUNLOP, J. W. C., JUNGNIKL, K., EDER, M., GIERLINGER, N., COUTAND, C., JERONIMIDIS, G., FRATZL, P. & BURGERT, I. (2008) Stress generation in the tension wood of poplar is based on the lateral swelling power of the G-layer. *The Plant Journal,* 56**,** 531-538.

GOWDA, M., VENU, R. C., RAGHUPATHY, M., NOBUTA, K., LI, H., WING, R., STAHLBERG, E., COUGLAN, S., HAUDENSCHILD, C., DEAN, R., NAHM, B.-H., MEYERS, B. & WANG, G.-L. (2006) Deep and comparative analysis of the mycelium and appressorium transcriptomes of *Magnaporthe grisea* using MPSS, RL-SAGE, and oligoarray methods. *BMC Genomics,* 7**,** 310.

GUPTA, P. K. (2008) Single-molecule DNA sequencing technologies for future genomics research. *Trends in Biotechnology,* 26**,** 602-611.

GUSTINCICH, S., SANDELIN, A., PLESSY, C., KATAYAMA, S., SIMONE, R., LAZAREVIC, D., HAYASHIZAKI, Y. & CARNINCI, P. (2006) The complexity of the mammalian transcriptome. *The Journal of Physiology,* 575**,** 321-332.

HAIGLER, C. H., IVANOVA-DATCHEVA, M., HOGAN, P. S., SALNIKOV, V. V., HWANG, S., MARTIN, K. & DELMER, D. P. (2001) Carbon partitioning to cellulose synthesis. *Plant Molecular Biology,* 47**,** 29-51.

HAMILTON, M., RAYMOND, C., HARWOOD, C. & POTTS, B. (2009) Genetic variation in *Eucalyptus nitens* pulpwood and wood shrinkage traits. *Tree Genetics & Genomes,* 5**,** 307-316.

HANRIOT, L., KEIME, C., GAY, N., FAURE, C., DOSSAT, C., WINCKER, P., SCOTE-BLACHON, C., PEYRON, C. & GANDRILLON, O. (2008) A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. *BMC Genomics,* 9**,** 418.

HEGEDUS, Z., ZAKRZEWSKA, A., ÁGOSTON, V. C., ORDAS, A., RÁCZ, P., MINK, M., SPAINK, H. P. & MEIJER, A. H. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Molecular Immunology,* 46**,** 2918-2930.

HENE, L., SREENU, V., VUONG, M., ABIDI, S. H., SUTTON, J., ROWLAND-JONES, S., DAVIS, S. & EVANS, E. (2007) Deep analysis of cellular transcriptomes - LongSAGE versus classic MPSS. *BMC Genomics,* 8**,** 333.

HORNER, D. S., PAVESI, G., CASTRIGNANÃ², T., DE MEO, P. D. O., LIUNI, S., SAMMETH, M., PICARDI, E. & PESOLE, G. (2009) Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in Bioinformatics,* 11**,** 181-197.

IMAI, N., SAMEJIMA, H., LANGNER, A., ONG, R. C., KITA, S., TITIN, J., CHUNG, A. Y. C., LAGAN, P., LEE, Y. F. & KITAYAMA, K. (2009) Co-benefits of sustainable forest management in biodiversity conservation and carbon sequestration. *PLoS ONE,* 4**,** e8267.

JANA, B. K., BISWAS, S., MAJUMDER, M., ROY, P. K. & MAZUMDAR, A. (2009) Carbon sequestration rate and aboveground biomass carbon potential of four young species. *Journal of Ecology and Natural Environment,* 1**,** 015-024.

JIANG, J. & JACOB, H. J. (1998) EbEST: An automated tool using expressed sequence tags to delineate Genome structure. *Genome Research,* 8**,** 268-275.

JIANG, S., XU, K., WANG, Y.-Z., REN, Y.-P. & GU, S. (2008) Role of GA3, GA4 and uniconazole-p in controlling gravitropism and tension wood formation in *Fraxinus mandshurica* Rupr. var. *japonica* Maxim. seedlings. *Journal of Integrative Plant Biology,* 50**,** 19-28.

JOSELEAU, J.-P., IMAI, T., KURODA, K. & RUEL, K. (2004) Detection in situ and characterization of lignin in the *G*-layer of tension wood fibres of *Populus deltoides*. *Planta,* 219**,** 338-345.

JOUREZ, B., RIBOUX, A. & LECLERCQ, A. (2001a) Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* cv 'Ghoy'). *International Association of Wood Anatomists,* 22**,** 133-157.

JOUREZ, B., RIBOUX, A. & LECLERCQ, A. (2001b) Comparison of basic density and longitudinal shrinkage in tension wood and opposite wood in young stems of *Populus euramericana* cv. Ghoy when subjected to gravitational stimulus. *Canadian Journal of Forest Research,* 31**,** 1676-1683.

JUNGHANS, U., LANGENFELD_HEYSER, R., POLLE, A. & TEICHMANN, T. (2004) Effect of auxin transport inhibitors and ethylene on the wood anatomy of poplar. *Plant Biology,* 6**,** 22-29.

KAKU, T., SERADA, S., BABA, K. I., TANAKA, F. & HAYASHI, T. (2009) Proteomic analysis of the G-layer in poplar tension wood. *Journal of Wood Science,* 55**,** 250-257.

KAN, Z., ROUCHKA, E. C., GISH, W. R. & STATES, D. J. (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Research,* 11**,** 889-900.

KARPINSKA, B., KARLSSON, M., SRIVASTAVA, M., STENBERG, A., SCHRADER, J., STERKY, F., BHALERAO, R. & WINGSLE, G. (2004) MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. *Plant Molecular Biology,* 56**,** 255-270.

KAWAOKA, A., KAOTHIEN, P., YOSHIDA, K., ENDO, S., YAMADA, K. & EBINUMA, H. (2000) Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis. *The Plant Journal,* 22**,** 289-301.

KEIME, C., SEMON, M., MOUCHIROUD, D., DURET, L. & GANDRILLON, O. (2007) Unexpected observations after mapping LongSAGE tags to the human genome. *BMC Bioinformatics,* 8**,** 154.

KENDE, H. (1989) Enzymes of ethylene biosynthesis. *Plant Physiology,* 91**,** 1-4.

KO, J.-H., KIM, W.-C. & HAN, K.-H. (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. *The Plant Journal,* 60**,** 649-665.

KULHEIM, C., HUI YEOH, S., MAINTZ, J., FOLEY, W. & MORAN, G. (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics,* 10**,** 452.

LAFARGUETTE, F., LEPLÉ, J.-C., DÉJARDIN, A., LAURANS, F., COSTA, G., LESAGE-DESCAUSES, M.-C. & PILATE, G. (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytologist,* 164**,** 107-121.

LASH, A. E., TOLSTOSHEV, C. M., WAGNER, L., SCHULER, G. D., STRAUSBERG, R. L., RIGGINS, G. J. & ALTSCHUL, S. F. (2000) SAGEmap: A public gene expression resource. *Genome Research,* 10**,** 1051-1060.

LASH, A. E., TOLSTOSHEV, C. M., WAGNER, L., SCHULER, G. D., STRAUSBERG, R. L., RIGGINS, G. J. & ALTSCHUL, S. F. (2006) SAGEmap: A public gene expression resource.

LIU, X., JIANG, J., SHAO, J., YIN, Y. & MA, Z. (2010) Gene transcription profiling of *Fusarium graminearum* treated with an azole fungicide tebuconazole. *Applied Microbiology and Biotechnology,* 85**,** 1105-1114.

LOVE, J., BJÖRKLUND, S., VAHALA, J., HERTZBERG, M., KANGASJÄRVI, J. & SUNDBERG, B. (2009) Ethylene is an endogenous stimulator of cell division in the cambial meristem of Populus. *Proceedings of the National Academy of Sciences,* 106**,** 5984-5989.

MANSFIELD, S. D. (2009) Solutions for dissolution-engineering cell walls for deconstruction. *Current Opinion in Biotechnology,* 20**,** 286-294.

MARDIS, E. (2008) The impact of next-generation sequencing technology on genetics. *Trends in genetics,* 24**,** 133-141.

MATSUMURA, H., REICH, S., ITO, A., SAITOH, H., KAMOUN, S., WINTER, P., KAHL, G., REUTER, M., KRUGER, D. H. & TERAUCHI, R. (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 15718-15723.

MCGIVNEY, B., MCGETTIGAN, P., BROWNE, J., EVANS, A., FONSECA, R., LOFTUS, B., LOHAN, A., MACHUGH, D., MURPHY, B., KATZ, L. & HILL, E. (2010) Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC Genomics,* 11**,** 398.

MEIJER, A. H., VERBEEK, F. J., SALAS-VIDAL, E., CORREDOR-ADÁMEZ, M., BUSSMAN, J., VAN DER SAR, A. M., OTTO, G. W., GEISLER, R. & SPAINK, H. P. (2005) Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to *Mycobacterium marinum* infection. *Molecular Immunology,* 42**,** 1185-1203.

MELLEROWICZ, E. J., IMMERZEEL, P. & HAYASHI, T. (2008) Xyloglucan: the molecular muscle of trees. *Annals of Botany*, 102**,** 659-665.

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

MEYERS, B. C., TEJ, S. S., VU, T. H., HAUDENSCHILD, C. D., AGRAWAL, V., EDBERG, S. B., GHAZAL, H. & DECOLA, S. (2004) The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. *Genome Research,* 14**,** 1641-1653.

MINIC, Z., JAMET, E., SAN-CLEMENTE, H., PELLETIER, S., RENOU, J.-P., RIHOUEY, C., OKINYO, D., PROUX, C., LEROUGE, P. & JOUANIN, L. (2009) Transcriptomic analysis of *Arabidopsis* developing stems: a close-up on cell wall genes. *BMC Plant Biology,* 9**,** 6.

MOROZOVA, O. & MARRA, M. A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92**,** 255-264.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5**,** 621-628.

MOYLE, R., SCHRADER, J., STENBERG, A., OLSSON, O., SAXENA, S., SANDBERG, G. & BHALERAO, R. P. (2002) Environmental and auxin regulation of wood formation involves members of the Aux/IAA gene family in hybrid aspen. *The Plant Journal,* 31**,** 675-685.

NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science,* 320**,** 1344-1349.

NELSON, C. D. & JOHNSEN, K. H. (2008) Genomic and physiological approaches to advancing forest tree improvement. *Tree Physiology,* 28**,** 1135-1143.

NGUEMA-ONA, E., BANNIGAN, A., CHEVALIER, L., BASKIN, T. I. & DRIOUICH, A. (2007) Disruption of arabinogalactan proteins disorganizes cortical microtubules in the root of *Arabidopsis thaliana*. *The Plant Journal,* 52**,** 240-251.

NISHIKUBO, N., AWANO, T., BANASIAK, A., BOURQUIN, V., IBATULLIN, F., FUNADA, R., BRUMER, H., TEERI, T. T., HAYASHI, T., SUNDBERG, B. R. & MELLEROWICZ, E. J. (2007) Xyloglucan endo-transglycosylase (XET) functions in gelatinous layers of tension wood fibers in poplar - a glimpse into the mechanism of the balancing act of trees. *Plant and Cell Physiology,* 48**,** 843-855.

NOVAES, E., DROST, D., FARMERIE, W., PAPPAS, G., GRATTAPAGLIA, D., SEDEROFF, R. & KIRST, M. (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics,* 9**,** 312.

OSSOWSKI, S., SCHNEEBERGER, K., CLARK, R. M., LANZ, C., WARTHMANN, N. & WEIGEL, D. (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Research,* 18**,** 2024-2033.

PAUX, E., CAROCHA, V., MARQUES, C., MENDES DE SOUSA, A., BORRALHO, N., SIVADON, P. & GRIMA-PETTENATI, J. (2005) Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytologist,* 167**,** 89-100.

PAUX, E., TAMASLOUKHT, M. B., LADOUCE, N., SIVADON, P. & GRIMA-PETTENATI, J. (2004) Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant Molecular Biology,* 55**,** 263-280.

PEWZNER-JUNG, Y., BRENNER, O., BRAUN, S., LAVIAD, E. L., BEN-DOR, S., FELDMESSER, E., HORN-SABAN, S., AMANN-ZALCENSTEIN, D., RAANAN, C., BERKUTZKI, T., EREZ-ROMAN, R., BEN-DAVID, O., LEVY, M., HOLZMAN, D., PARK, H., NYSKA, A., MERRILL, A. H. & FUTERMAN, A. H. (2010) A critical role for ceramide synthase 2 in liver homeostasis. *Journal of Biological Chemistry,* 285**,** 10911-10923.

PICOULT-NEWBERG, L., IDEKER, T. E., POHL, M. G., TAYLOR, S. L., DONALDSON, M. A., NICKERSON, D. A. & BOYCE-JACINO, M. (1999) Mining SNPs from EST databases. *Genome Research,* 9**,** 167-174.

PILATE, G., CHABBERT, B., CATHALA, B., YOSHINAGA, A., LEPLÉ, J.-C., LAURANS, F., LAPIERRE, C. & RUEL, K. (2004) Lignification and tension wood. *Comptes Rendus Biologies,* 327**,** 889-901.

PLOMION, C., LEPROVOST, G. & STOKES, A. (2001) Wood Formation in Trees. *Plant Physiology,* 127**,** 1513-1523.

POP, M. & SALZBERG, S. L. (2009) Bioinformatics challenges of new sequencing technology. *Trends in Genetics,* 24**,** 142-149.

PRODHAN, A. K. M. A., OHTANI, J., FUNADA, R., ABE, H. & FUKAZAWA, K. (1995) Ultrastructural investigation of tension wood fibre in *Fraxinus mandshurica* Rupr. var. *japonica* Maxim. *Annals of Botany,* 75**,** 311-317.

QIU, D., WILSON, I. W., GAN, S., WASHUSEN, R., MORAN, G. F. & SOUTHERTON, S. G. (2008) Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. *New Phytologist,* 179**,** 94-103.

REINHARDT, J. A., BALTRUS, D. A., NISHIMURA, M. T., JECK, W. R., JONES, C. D. & DANGL, J. L. (2009) De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. oryzae. *Genome Research,* 19**,** 294-305.

ROMUALDI, C., BORTOLUZZI, S., D'ALESSI, F. & DANIELI, G. (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiological Genomics,* 12**,** 159-162.

RUDEL, T. K. (2006) Shrinking tropical forests, human agents of change, and conservation policy. *Conservation Biology,* 20**,** 1604-1609.

RUDEL, T. K. (2007) Changing agents of deforestation: From state-initiated to enterprise driven processes, 1970-2000. *Land Use Policy,* 24**,** 35-41.

SAHA, S., SPARKS, A. B., RAGO, C., AKMAEV, V., WANG, C. J., VOGELSTEIN, B., KINZLER, K. W. & VELCULESCU, V. E. (2002) Using the transcriptome to annotate the genome. *Nature Biotechnology,* 20**,** 508-512.

SCHEIBLE, W.-R. & PAULY, M. (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Current Opinion in Plant Biology,* 7**,** 285-295.

SEO, Y.-K., CHONG, H. K., INFANTE, A. M., IM, S.-S., XIE, X. & OSBORNE, T. F. (2009) Genome-wide analysis of SREBP-1 binding in mouse liver chromatin reveals a preference for promoter proximal binding to a new motif. *Proceedings of the National Academy of Sciences,* 106**,** 13765-13769.

SERTEYN, D., PIQUEMAL, D., VANDERHEYDEN, L., LEJEUNE, J.-P., VERWILGHEN, D. & SANDERSEN, C. (2010) Gene expression profiling from leukocytes of horses affected by osteochondrosis. *Journal of Orthopaedic Research,* 28**,** 965-970.

SHEPHERD, M., CROSS, M., DIETERS, M. J., HARDING, K., KAIN, D. & HENRY, R. (2003) Genetics of physical wood properties and early growth in a tropical pine hybrid. *Canadian Journal of Forest Research,* 33**,** 1923-1932.

SIDDIQUI, A. S., DELANEY, A. D., SCHNERCH, A., GRIFFITH, O. L., JONES, S. J. M. & MARRA, M. A. (2006) Sequence biases in large scale gene expression profiling data. *Nucleic Acids Research,* 34**,** e83.

SIMPSON, J. T., WONG, K., JACKMAN, S. D., SCHEIN, J. E., JONES, S. J. M. & BIROL, Ä. N. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research,* 19**,** 1117-1123.

SUN, M., ZHOU, G., LEE, S., CHEN, J., SHI, R. & WANG, S. (2004) SAGE is far more sensitive than EST for detecting low-abundance transcripts. *BMC Genomics,* 5**,** 1.

'T HOEN, P. A. C., ARIYUREK, Y., THYGESEN, H. H., VREUGDENHIL, E., VOSSEN, R. H. A. M., DE MENEZES, R. E. X., BOER, J. M., VAN OMMEN, G.-J. B. & DEN DUNNEN, J. T. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Research,* 36**,** e141.

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

USECHE, F. J., GAO, G., HANAFEY, M. & RAFALSKI, A. (2001) High-throughput identification, database storage and analysis of SNPs in EST sequences. *Genome Informatics*, 12**,** 194-203.

VASTA, V., NG, S. B., TURNER, E. H., SHENDURE, J. & HAHN, S. H. (2009) Next generation sequence analysis for mitochondrial disorders. *Genome medicine*, 1**,** 100.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. (1995) Serial analysis of gene expression. *Science*, 270**,** 484-487.

WAHL, M. B., HEINZMANN, U. & IMAI, K. (2005) LongSAGE analysis significantly improves genome annotation: identifications of novel genes and alternative transcripts in the mouse. *Bioinformatics*, 21**,** 1393-1400.

WANG, S. M. (2003) Response: the new role of SAGE in gene discovery. *Trends in biotechnology*, 21**,** 57-58.

WANG, S. M. (2007) Understanding SAGE data. *Trends in Genetics*, 23**,** 42-50.

WANG, X.-W., LUAN, J.-B., LI, J.-M., BAO, Y.-Y., ZHANG, C.-X. & LIU, S.-S. (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, 11**,** 400.

WANG, Z., ZANG, C., CUI, K., SCHONES, D. E., BARSKI, A., PENG, W. & ZHAO, K. (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*, 138**,** 1019-1031.

WASHUSEN, R. (2002) Tension wood occurence in *Eucalyptus globulus* Labill. II. The spatial distribution of tension wood and its association with stem form. *Australian Forestry*, 65**,** 127-134.

WASHUSEN, R. & EVANS, R. (2001) Prediction of wood tangential shrinkage from celllulose crystallite width and density in one 11-year-old tree of *Eucalyptus globulus* Labill. *Australian Forestry*, 64**,** 123-126.

WASHUSEN, R. & ILIC, J. (2001) Relationship between transverse shrinkage and tension wood from three provenances of Eucalyptus globulus Labill. *European Journal of Wood and Wood Products,* 59**,** 85-93.

WASTENEYS, G. & FUJITA, M. (2006) Establishing and maintaining axial growth: wall mechanical properties and the cytoskeleton. *Journal of Plant Research,* 119**,** 5-10.

WILSON, B. F. (2000) Apical control of branch growth and angle in woody plants. *American Journal of Botany,* 87**,** 601-607.

WU, L., JOSHI, C. P. & CHIANG, V. L. (2000) A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. *The Plant Journal,* 22**,** 495-502.

XIAO, S., JIA, J., MO, D., WANG, Q., QIN, L., HE, Z., ZHAO, X., HUANG, Y., LI, A., YU, J., NIU, Y., LIU, X. & CHEN, Y. (2010) Understanding PRRSV infection in porcine lung based on genome-wide transcriptome response identified by deep sequencing. *PLoS ONE,* 5**,** e11377.

YAMAMOTO, H., YOSHIDA, M. & OKUYAMA, T. (2002) Growth stress controls negative gravitropism in woody plant stems. *Planta,* 216**,** 280-292.

YANG, S.-H., WANG, H., SATHYAN, P., STASOLLA, C. & LOOPSTRA, C. A. (2005) Real-time RT-PCR analysis of loblolly pine (*Pinus taeda*) arabinogalactan-protein and arabinogalactan-protein-like genes. *Physiologia Plantarum,* 124**,** 91-106.

ZHONG, R., BURK, D. H., MORRISON, W. H., III & YE, Z.-H. (2002) A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. *Plant Cell,* 14**,** 3101-3117.

ZHOU, X., REN, L., LI, Y., ZHANG, M., YU, Y. & YU, J. (2010) The next-generation sequencing technology: A technology review and future perspective. *Science China Life Sciences,* 53**,** 44-57.

Chapter 2

ANALYSIS OF DIGITAL GENE EXPRESSION (DGE) DATA DERIVED FROM EUCALYPTUS TENSION WOOD FORMING TISSUES

J.F. Silberbauer[1], E. Mizrachi[1], C. Hefer[2], F. Joubert[2], D.K. Berger[3] and A.A. Myburg[1]

[1]Department of Genetics, [3]Department of Plant Science, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 2000, South Africa

[2]Bioinformatics and Computational Biology Unit, Department of Biochemistry, University of Pretoria, Pretoria, 2000, South Africa

Author contributions

I performed the laboratory work which included the RNA extractions, DGE library preparation and RT-qPCR verification. I performed the data analysis assisted by E. Mizrachi who also guided me with the planning of the work. C. Hefer performed the bioinformatic programming for the data analysis. Prof. A.A. Myburg, Prof. D.K Berger and Prof F. Joubert provided advice, direction and supervision in the planning of the project, and performed critical review of the written chapter. Prof. Myburg conceived the main concept for the study and obtained funding for the research. All other technical assistance is acknowledged at the beginning of the thesis.

2.1 Introduction

Differences in transcriptional activity are one of the main regulatory factors that influence the development, identity and functions of tissues and cell types in multicellular organisms. This includes the structural variation of expressed transcripts and degree of expression of those transcripts (Gustincich et al., 2006, Yamada et al., 2003). Expression profiling studies not only lead to insight with regards to the functioning of cell types and tissues, but provide an approach to identify genes that play central roles in a specific cellular process. The outcomes of expression profiling studies validate existing observations and drive the development of new hypotheses. Genome-wide transcript profiling with microarrays and massively parallel DNA sequencing technologies, known as second generation DNA sequencing, enable modelling of the complex transcriptional networks (Nelson and Johnsen, 2008). The validity of observations made with expression profiling applications is therefore of great importance as incorrect results will hinder the modelling of complex biological systems and support false hypotheses. Biologically meaningful expression profiling studies depend on correct experimental design, proper execution of experimental procedures and application of appropriate data analysis techniques (Bustin et al., 2009, Leakey et al., 2009).

Since the first publication in 1995, microarrays have become one of the most efficient methods for the quantification of transcriptome-wide gene expression (Schena et al., 1995). Microarrays continue to be utilised in a wide variety of biological fields (Coppée, 2008, Gregory et al., 2008). Online databases open to public access such as ArrayExpress (Parkinson et al., 2009) or GeneExpressionOmnibus (Barret and Edgar, 2006) serve as repositories where researchers share data with the global scientific community. In the past decade, concerns have been raised with regards to the reproducibility of microarray experiments (Klebanov and Yakovlev, 2007, Michiels et al., 2005, Tan et al., 2003). This has put the reliability of published results under scrutiny and led to the MicroArray Quality Control MAQC project (Chen et al., 2007, Ioannidis et al., 2009, MAQC.Consortium, 2006). The MAQC project was initiated to address problems surrounding the reproducibility of microarray studies. The project aimed to establish confidence in microarray results by implementing a robust methodology for the use of microarrays so that this technology may be used

as a clinical diagnostic application. Employing different means of data normalisation and significance testing to datasets have been shown to produce considerable differences in the microarray results obtained (Mecham et al., 2010, Shi et al., 2005, Tan et al., 2003). Therefore one of the focal points of the MAQC project was the evaluation of data analysis methods employed by different researchers in different laboratories (Chen et al., 2007).

Reverse transcriptase real-time quantitative PCR (RT-qPCR) has found wide application in research and as a clinical diagnostic tool over the past fourteen years (Bernard and Wittwer, 2002, Caldana et al., 2007, Heid et al., 1996, Whiley et al., 2009). Similar to microarrays, the reliability and reproducibility of RT-qPCR studies have been questioned (Bustin and Nolan, 2004, Garson et al., 2009, Hendriks-Balk et al., 2007). The MIQE (Minimum Information for Publication of Quantitative real-time PCR Experiments) guidelines are a set of recommendations put forth by Bustin et al. (2009). These guidelines address important steps of experimental procedure and analysis of RT-qPCR data that, when performed inconsistently, may produce unreliable results. Inconsistent methodology used for RT-qPCR data analysis is one of the sources of non-reproducible results (Karlen et al., 2007, Vandesompele et al., 2002, Yuan et al., 2006).

Digital Gene Expression (DGE, Figure 2.1), a sequencing-based expression profiling application developed by Illumina (Hene et al., 2007), shares the concept of Serial Analysis of Gene Expression (SAGE, Velculescu et al., 1995). The basic concept of DGE is to generate sequence tags from a pool of diverse poly-A captured mRNA transcripts that are counted and mapped back onto an annotated reference sequence. This is done by digesting captured cDNA converted transcripts with restriction enzymes such as DpnII and MmeI, and ligating adapters to the end of the fragment used for the linear amplification and sequencing processes (Figure 2.1, 't Hoen et al., 2008). The unique tag sequence is dependent on the position of the DpnII restriction enzyme recognition site (which may occur once per 256 bp) within the transcript sequence and the expression of a transcript within a tissue is therefore identified by a corresponding tag.

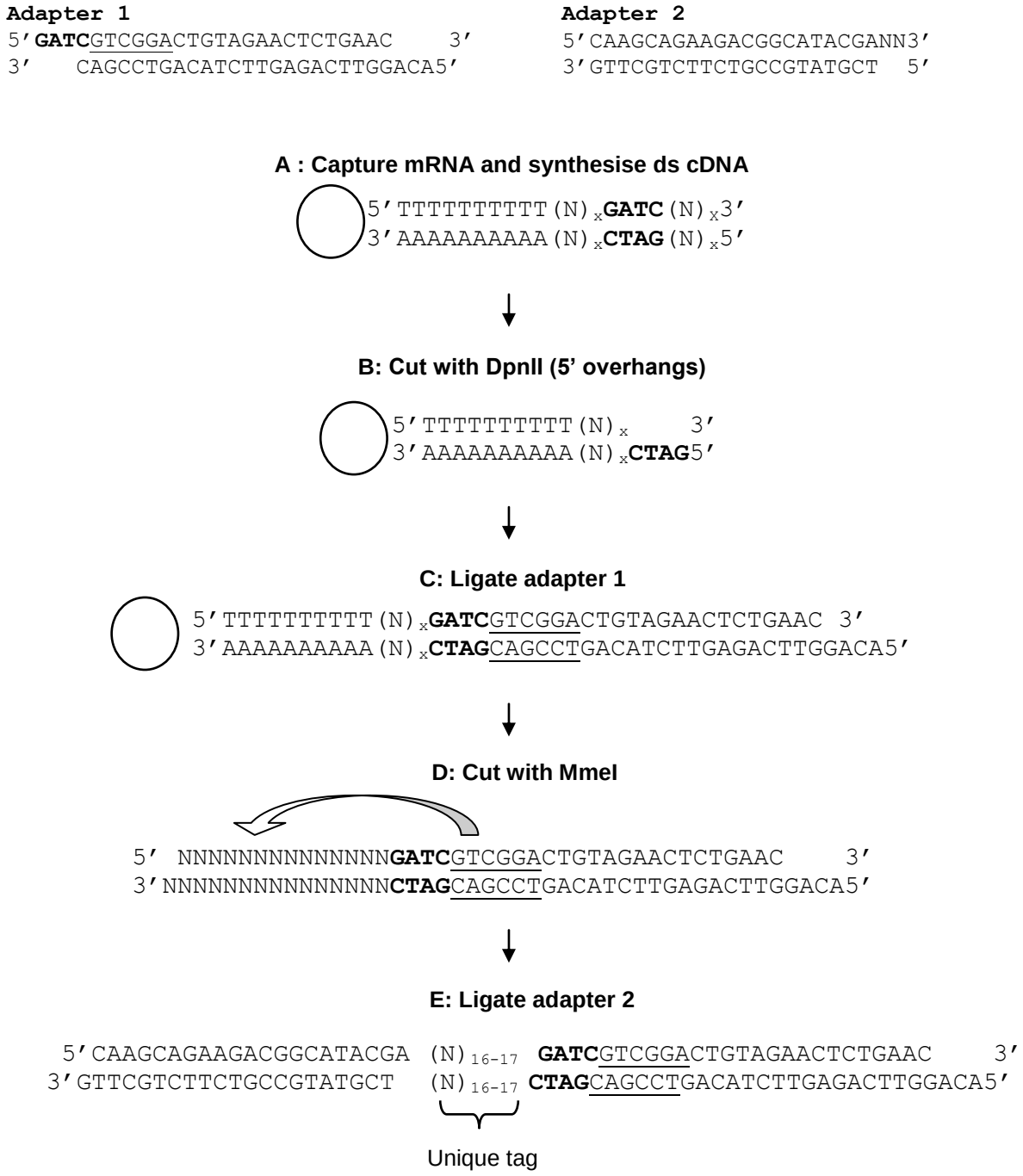


Figure 2.1: Schematic showing the generation of a DGE tag at the sequence level as outlined in the Materials and Methods, Section 2.3.4. The sequences of adapter 1 and adapter 2 are indicated at the top of the figure. The DpnII restriction site located on the mRNA and adapter 1 sequences are depicted in bold. The MmeI recognition site located on adapter 1 is underlined.

DGE has been applied in studies of primate brain evolution (Babbitt et al., 2010), gene expression in human brain tissues (Asmann et al., 2009, 't Hoen et al., 2008), responses to viral and bacterial infections (Hegedus et al., 2009, Xiao et al., 2010), equine skeletal muscle physiology (McGivney et al., 2010), liver homeostasis and cell physiology (Pewzner-Jung et al., 2010), and the response of *Fusarium* to azole fungicides (Liu et al., 2010). Currently no set standards have been published for DGE-based expression profiling. To date, DGE data is processed by a series of custom software scripts (Feng et al., 2010, Hanriot et al., 2008, Hegedus et al., 2009, Wang et al., 2010). Requirements for the standardisation of microarrays and RT-qPCR due to non-reproducible results necessitate precise analysis and interpretation of all expression profiling techniques. A hurdle to the successful use of next generation DNA sequencing applications is that many researchers lack the expertise in bioinformatics required for the data processing and the responsibility rests on bioinformaticians to develop user-friendly analysis tools (Cantacessi et al., 2010). Investigation of the characteristics of DGE data is needed to identify any unexpected technical artefacts which may be hidden in the sequence information. This will aid in the development of standardised analysis procedures aimed at reducing technical variation so that the true biological variation between systems may be detected.

Well designed expression profiling experiments should estimate technical and biological variation (Nettleton, 2006). The biological variation contained in an experiment is ideally estimated by obtaining separate measurements for each individual genotype or individual assessed, and by including multiple genotypes or individuals as biological replicates in the first place. However, cost remains a limiting factor for the profiling of multiple biological replicates (Cantacessi et al., 2010, Zhang et al., 2007). Thus at the risk of compromising the power of expression profiling studies, pooling of biological samples is often performed ('t Hoen et al., 2008, Zhang et al., 2007). It is assumed that expression profiles measured for a pooled sample represent the average expression profiles of the individual biological replicates. This assumption known as the “biological averaging assumption” has been examined with microarray studies with conflicting results (Kendziorski et al., 2005, Mary-Huard et al., 2007, Shih et al., 2004, Zhang et al., 2007). For some DGE studies,

biological averaging has been assumed and biological replicates were pooled before DGE library preparation (David et al., 2010, Hegedus et al., 2009, Pewzner-Jung et al., 2010, Xiao et al., 2010). Other DGE studies have profiled libraries prepared from individual biological replicates (Asmann et al., 2009, Babbitt et al., 2010, Feng et al., 2010). The study by 't Hoen et al. (2008) attempted to measure the effect of assumed biological averaging using a DGE platform. However, contamination of the pooled sample was discovered post-analysis which invalidated the attempt leaving the question unanswered ('t Hoen et al., 2008).

Tension wood is a specialised type of reaction wood formed by Angiosperm trees and serves the function of maintaining a tree's morphology. Tension wood is characterised by asymmetrical growth on the upper side of a branch or a bent stem (Pilate et al., 2004). Mechanical growth stress interactions between different cell wall layers in tension wood stem cells result in the stabilisation of a branch and are able to re-orientate a bent stem to an upright position (Clair et al., 2006). The generation of growth stresses within tension wood is due to altered chemical and physical characteristics in the secondary cell walls of tension wood fibres. Tension wood exhibits an increased amount of highly crystalline cellulose and an increased amount of specific hemicellulose polymers such as xyloglucan (Baba et al., 2009). The overall content of lignin and remaining hemicelluloses are decreased along with a great decrease in the cellulose microfibril angle (Andersson-Gunnerås et al., 2006, Joseleau et al., 2004). The formation of tension wood compared with normal wood is of particular interest as key genes involved in cellulose, hemicellulose and lignin biosynthesis are differentially regulated and can therefore be identified and characterised.

This study was aimed at identifying genes that are differentially regulated during tension wood formation by inducing tension wood in the hybrid Angiosperm *Eucalyptus grandis x Eucalyptus urophylla*, a widely grown eucalypt hybrid combination. The first aim of this study was to establish a set of data processing guidelines for the DGE data produced in this study. This was done through detailed investigation of the technical features of DGE data obtained from the *Eucalyptus* tension wood induction trial. These guidelines were then validated by generating expression profiles for the extensively studied primary and secondary cell wall related cellulose synthase (*CesA*) genes in

Eucalyptus. The second aim was to assess the level of biological variation among biological replicates in *Eucalyptus*. In addition, the biological averaging assumption was investigated by testing the correlation between a pooled sample and the average of individually sequenced biological replicates. Real-time RT-qPCR was performed to technically validate DGE results, as well as validation with regards to the biological averaging assumption. This was done by bending 18-month-old stems at a 45° angle for the duration of 6 hours, 24 hours, 1 week, 2 weeks, 3 weeks and 6 months for comparison against unbent control trees. Seven DGE libraries were constructed from pooled samples (three independant ramets) for each of the bending trial time points including the unbent control. Separate libraries were constructed for each of the three biological replicates of the unbent control. In addition a second DGE library was prepared from one of the unbent control biological replicates to serve as a technical replicate.

2.2 Materials and Methods

2.2.1 Tension wood induction trial

A tree bending trial was conducted in a clonal plantation near Kwambonambi in Northern Kwazulu-Natal (Sappi Forestry) to induce tension wood formation in *Eucalyptus* trees. The tension wood induction time trial consisted of unbent trees to serve as controls (timepoint 0) and trees bent for six different time periods: 6 hours, 24 hours, 1 week, 2 weeks, 3 weeks and 6 months. At each time point, the main stems of three 18-month-old ramets of a single F1 hybrid *Eucalyptus grandis* x *urophylla* clone (GUSAP1, Sappi Forestry) were bent at an angle of 45°. To avoid temporal gene expression variation sampling of all time points was completed within three hours on the same day under the same environmental conditions.

2.2.2 Sampling of plant material

Stems were debarked immediately before taking xylem scrapings to avoid wounding-related gene expression and oxidation of the samples. Differentiating xylem tissue was collected by cutting out the section of the stem bent at 45° (approximately 50 cm), removing the bark and immediately scraping the exposed total xylem 4-5 mm deep. For each bent stem, the upper (tension wood) side was scraped as well as the corresponding locations (height) on the three unbent controls. Samples were taken from both sides of each of the three unbent control tree stems and were bulked separately. All samples were immediately frozen in liquid nitrogen and stored at -80°C.

2.2.3 RNA Isolation

Total RNA was isolated from the xylem samples using a cetyl trimethylammonium bromide (CTAB) based method (Chang et al., 1993). The frozen wood samples were ground to a fine powder in liquid nitrogen using a high speed grinder (IKA-Werke, Staufen, Germany). Fifteen ml of extraction buffer was mixed with three grams of ground tissue. RNA quantity and purity were assessed by using a Nanodrop spectrophotometer (Nanodrop Technologies ND 1000, Wilmington DE), Agilent Bioanalyser 2100 RNA 6000 pico total RNA kits (Agilent Technologies, Santa Clara CA) and 1.5% agarose gels. In addition, the samples were tested for DNA contamination as described below. First strand cDNA was synthesised using ImProm-II™ Reverse Transcriptase (Promega, Madison, WI) according to the manufacturer's instructions. Polymerase Chain Reaction (PCR) amplification was performed using primers designed for a *Eucalyptus* ADP Ribosylation Factor gene (GenBank Acc: AY770746) (Forward: TTCTGGTGCCATGCTGAGAA, Reverse: GATGCTGTGTTGCTCGTCTT). Genomic DNA extracted from the same Sappi *Eucalyptus grandis x urophylla* tree (GUSAP1) was used as a positive control. The PCR products were resolved on a 1.5% agarose gel.

2.2.4 Digital Gene Expression (DGE) library preparation

Ten DGE libraries were prepared from the tension wood induction time trial. These include a control library from pooled RNA of three unbent biological replicate trees, a separate library from each of the three unbent control biological replicates and six pooled tension wood tag libraries (one library containing a pool of three biological replicate RNA samples for each time point in the bending trial).

For each of the pooled DGE libraries, 20 µg total RNA from each of the three biological replicates per time point were pooled. Library preparation was done using Illumina's DGE Tag Profiling for DpnII Sample Prep kit (Illumina, San Diego, CA) according to the manufacturer's instructions with the exception of alterations to the number cycles used in the enrichment PCR (see below). Two micrograms of total RNA were incubated with oligo-dT beads to capture the polyadenylated RNA transcripts. First- and second-strand cDNA synthesis was performed using the captured transcripts as template (Figure 2.1A). The double-stranded cDNA products were treated with DpnII to produce fragments digested at the DpnII recognition cut site GATC (Figure 2.1B). The 5' ends of the digested fragments not bead bound were washed away. The GEX DpnII adapter was ligated to the 5' ends of the fragments still attached to the oligo-dT beads (Figure 2.1C). The GEX DpnII adapter contained the recognition site for the tagging restriction enzyme MmeI. MmeI, which digests 16-17 bp downstream of the GATC site (towards the poly-A tail), was used to release the fragments from the oligo-dT beads (Figure 2.1D). The oligonucleotide tags generated after MmeI digestion were 20-21 bp long (16-17 bp plus the GATC DpnII restriction site). After dephosphorylation and a phenol extraction of the released fragments, the GEX adapter 2 was ligated to the free 3' end of each fragment (Figure 2.1E). A linear PCR using 18-20 amplification cycles, instead of 15 amplification cycles, was performed to selectively enrich the cDNA fragments that had adapter sequences on both ends. The entire adapter-ligated construct was amplified with the two adapter-specific primers supplied in the kit. The resulting 92-96 bp fragments were purified by gel excision and elution after separation on a 6% polyacrylamide TBE gel. The eluant was purified from the gel debris using Spin-X Cellulose Acetate Filters (2 ml, 0.45 µm, Sigma-Aldrich). The ligated fragments were precipitated by adding 1 µl of glycogen, 10 µl of 3 M sodium acetate (pH 5.2) and

330 µl of 100% ethanol (-20°C), followed by centrifugation at 11 000 x g for 20 min at room temperature. The pellet was subsequently washed with 70% ethanol (room temperature), and resuspended in 10 µl of the resuspension buffer provided in the kit. Purity and yield of the eluted libraries were assessed using a 2100 Bioanalyser DNA 1000 kit with the fragment detection range 25 to 1000 bp (Agilent Technologies).

Approximately 2.5 pm of cDNA was used per DGE library for cluster generation using the DGE-Tag Profiling DpnII Cluster Generation Kit. Sequencing by synthesis was performed for each of the libraries using the Illumina Genome Analyser II system at the University of Western Cape (UWC, Cape Town, South Africa), and 18-Cycle Sequencing Kits (Illumina) according to the manufacturer's instructions. One library, a biological replicate of the unbent control, was selected to be sequenced in duplicate to yield a technical replicate. Image analysis and base calling were performed using the Illumina pipeline.

2.2.4.1 Mapping, processing and analysis of DGE data

The raw FASTQ sequence files containing the sequence text and quality scores from the Illumina base calling was converted to FASTA format which contains sequence text only. Customised Python scripts (P. Law, University of Pretoria) were used to trim off the adapter sequences and to filter for the correct length of 16-17 bp per DGE tag. The DpnII cut site (GATC) was added to the 5' end of each tag (Figure 2.1) to enhance mapping specificity.

The resulting 20-21 bp DGE tags were mapped to the reference transcriptome sequence using Bowtie (Langmead et al., 2009). A total of 18,894 mRNA-Seq assembled cDNA contigs (transcript models) produced for the same hybrid *Eucalyptus grandis x E.urophylla* (GUSAP1) clonal genotype (Mizrachi et al., 2010) was used as the reference sequence to identify the transcripts from which the tags were generated. To allow for sequencing error and allelic differences between the *Eucalyptus grandis* and *Eucalyptus urophylla* alleles in the GUSAP1 transcriptome, each DGE dataset was mapped to the transcriptome with a maximum of two mismatches allowed per tag.

DGE tag mapping data for seven *Eucalyptus* cellulose synthase (*CesA*) genes (Ranik and Myburg, 2006) were exported to Microsoft Excel 2007 for further investigation. Four primary cell wall *CesA* genes, *EgCesA4* (ortholog of *AtCesA3*: AT5G05170.1), *EgCesA5* (ortholog of *AtCesA1*: AT4G32410.1), *EgCesA6* (ortholog of *AtCesA2*: AT4G39350.1), *EgCesA7* (ortholog of *AtCesA2*: AT4G39350.1) and three secondary cell wall *CesA* genes, *EgCesA1* (ortholog of *AtCesA8*: AT4G18780.1), *EgCesA2* (ortholog of *AtCesA4*: AT5G44030.1) and *EgCesA3* (*AtCesA7*: AT5G17420.1) were selected for detailed analysis of tag mapping per transcript.

Subsequent to the analysis of DGE mapping data to the seven *CesA* genes, a customised Python script was written for the automated processing of the entire tension wood DGE data set (C. Hefer, University of Pretoria). The Python script was developed to extract the positions and counts for each tag mapped to each cDNA contig from SAM formatted files (Li et al., 2009). Tags that overlapped by 1 bp or none were joined into a single position count, and only positions where more than 10 tags mapped to the same position were considered in the downstream analysis. The sum of the counts of the tags that respectively mapped in the sense and antisense direction on the transcript were calculated and converted to parts per million (ppm).

2.2.5 Reverse transcription quantitative real-time PCR (RT-qPCR) analysis

Nine targets were selected for RT-qPCR analysis and validation of the DGE data. All of the genes, except the *EgCesA* genes were named based on *Arabidopsis thaliana* ortholog descriptions. These targets included a primary cell wall cellulose synthase (*EgCesA5*; AT4G32410.1), a secondary cell wall cellulose synthase (*EgCesA3*; AT5G17420.1), korrigan (*KOR1*; AT5G49720.1), sucrose synthase (*SUS1*; AT3G43190.1), xyloglucan endotranscglucosylase (*XTR6*; AT4G25810.1), galacturonosyltransferase (*GAUT12*; AT5G54690.1), an auxin response transcription factor (*ARF6*; AT1G30330.1), a knotted-like homeobox transcription factor from *Arabidopsis thaliana* (*KNAT7*; AT1G62990.1) and caffeoyl-CoA 3-O-methyltransferase (*CCoAOMT*; AT4G34050.1), a lignin biosynthesis gene. Two genes with constitutive expression based on in the DGE data set were selected as control genes. These are actin (*ACT7*; AT5G09810.1) and adenine phosphoribosyl

transferase (*APT1*; AT1G27450.1). Five genes were represented by primer pairs optimized in previous studies in our research group. These were *EgCesA5*, *EgCesA3*, *KOR1*, *SUS1* and *CCoAOMT*. Gene-specific primer pairs were designed for the remaining genes using Primer Designer 5 (Scientific and Educational Software, Cary, NC). All primer sequences are recorded in Table 2.1.

RT-qPCR validation was performed using the LightCycler® 480 instrument and LightCycler® 480 SYBR green I Master kit (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturer's instructions. cDNA synthesis was carried out in duplicate for each

Table 2.1: Genes profiled with RT-qPCR, the sequences of the oligonucleotide primers used and the observed size of the amplified fragment.

Gene	Forward Primer	Reverse Primer	Amplicon length (bp)
Cellulose biosynthesis			
cellulose synthase 3	CTGTGATTGCTGCCCATGCT	CGCCACCCTGTTCCATCAAA	194
cellulose synthase 5	GGGAAGGGTGGCAATAAGAA	AACAGGAGACTGACCGAATC	177
Korrigan 1 (endo-1,4-β-glucanase)	GGACCGCTAAGATCGACAAG	CCAGCTCTGGATTGTGTCAT	120
Sugar metabolism			
UDP-glycosyltransferase/ sucrose synthase	GCCAACTTCTTCGAGAAGTG	TTGAGCAGCCTCTCGGAGTA	206
Hemicellulose biosynthesis			
Xyloglucan endotransglycosylase 6	GTGACGAGCTGGCCGTTATC	CATCGGCTCTTCCGTGGTTG	104
Galacturonosyltransferase 12	AAGGAGAACCACGCTTGCAT	TGCCAGATTGCAGCTTCCAT	206
Transcription factors			
Auxin Response Factor 6	TCGAGATGGAGACCGAGGAA	CATGCTGCACAGGCATAAGG	199
Knotted-like Homeobox of *Arabidopsis thaliana* 7	AGTCCAAGCGCAAGAGGTAG	CCACCACCACAAGTGTTCAG	155
Lignin biosynthesis			
Caffeoyl-CoA 3-O-Methyltransferase	TGCACGAGCTGATCAAGGAG	ACGCCAAGAACACCATGGAG	211
Reference genes			
Actin 7	CCTCTGTCTCCGCCATACTT	CACGCCTTCTTCTGCTTCTG	139
Adenine phosphoribosyl transferase 1	GCATCTCCATCGTGTCTGTT	TCTCTGTTGTGGCAGGTGTT	164

sample. A total of four pooled cDNA samples were synthesised as described in Section 2.3.3 from the RNA extracted from the three biological replicates of the unbent control samples and from the pool of the three control samples. Each 11 µl reaction contained 5 µl of 2 x SYBR Green, 0.4 mM forward and reverse primers and 1 µl cDNA. The PCR reaction conditions included an initial denaturation step at 95°C for 5 min followed by 45 cycles of denaturation for 10 sec at 94°C, primer annealing for 10 sec at 60°C and elongation for 10 sec at 72°C. A final elongation step of 5 min at 72°C was performed. The RT-qPCR reactions were carried out in triplicate (technical replicates) and melting curve analysis followed by agarose gel electrophoresis of the amplified product was performed. The qBase software (Biogazelle NV, Ghent, Belgium) was used for raw data quality control, PCR amplification efficiency correction and multiple reference gene normalisation (Hellemans et al., 2007). The results table was exported and graphs were produced using Microsoft Excel.

2.3 Results

2.3.1 RNA isolation and quality control

High quality and quantity RNA was isolated from the *Eucalyptus* control and tension wood tissues (Table 2.2). An average amount of 3750 µg total RNA was extracted with an $OD_{260/280}$ absorbance ratio of approximately 2.0 indicating pure RNA. The 28S and 18S rRNA bands were clearly resolved and intact with agarose gel electrophoresis assays (Figure 2.2A). All RNA samples had RNA integrity (RIN) numbers of 8.0 or higher indicating that the RNA samples were not degraded (Table 2.2). The amplification of a region of an ADP-ribosylation factor gene (GenBank Acc: AY770746) from *Eucalyptus* confirmed the absence of contaminating genomic DNA for all cDNA samples (Figure 2.2B). The presence of the 280 bp fragment amplified from mRNA was clearly distinguished from the 1190 bp intron-containing fragment amplified from genomic DNA.

2.3.2 DGE library construction

Nine DGE libraries were produced (Table 2.2). The concentration of the six libraries produced from pooled RNA samples (from three biological replicates) was assessed with the 2100 Agilent Bioanalyser instrument (Table 2.2). The six libraries include the unbent control, 6 hours, 24 hours, 1 week, 2 weeks and 6 months samples. The Bioanalyser electrophoretic assay (Figure 2.3) revealed a pool of 96 bp fragments indicating that the DGE libraries were pure and devoid of other contaminating fragment species.

Table 2.2: Total RNA quality and the concentrations of the DGE libraries subsequently produced.

Sample Name	Total RNA (µg)	260/280 [a]	RIN value [b]	DGE library concentration (ng/µl) [c]
Unbent control	3716.36	2.07	8.3	1.80
6 hours	3462.97	2.10	8.8	1.84
24 hours	3430.80	2.11	8.6	1.87
1 week	3748.51	2.05	8.5	0.80
2 weeks	4432.52	1.81	8.7	1.36
3 weeks	4549.37	1.62	8.3	3.00
6 months	2890.56	2.09	8.0	0.66
Unbent control biological replicate 1	3489.65	2.10	NA[d]	2.14
Unbent control biological replicate 2	4296.49	1.93	NA[d]	7.30
Unbent control biological replicate 3	3515.79	2.10	NA[d]	2.80

[a] Determined with a Nanodrop spectrophotometer (Nanodrop Technologies ND 1000, DE, USA).

[b] RIN: RNA Integrity Number determined using the Agilent Bioanalyser 2100 total pico RNA kit.

[c] Determined using the Agilent Bioanalyser 2100 DNA 1000 kit (size detection range: 25 -1000 bp).

[d] The RNA integrity was not assessed for the individual biological replicates of the unbent control. An estimate of the RNA integrity was obtained for the pooled unbent control sample (8.3).

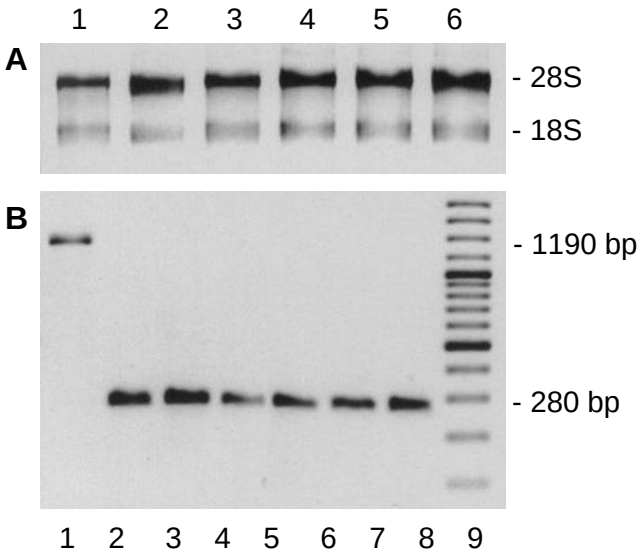


***Figure 2.2*:** RNA quality assessment. (A) Total RNA from six normal and tension wood tissues analysed by 1.5% agarose gel electrophoresis. Intact 28S and 18S bands indicate intact, high quality RNA samples. Lane 1: Unbent Control, Lane 2: 6 hours, Lane 3: 24 hours, Lane 4: 1 week, Lane 5: 2 weeks, Lane 6: 6 months bent samples.

(B) Genomic DNA contamination assay. Quality of the cDNA samples were assessed by performing an intron-spanning PCR on an ADP Ribosylation Factor gene (Forward primer: TTCTGGTGCCATGCTGAGAA, Reverse primer: GATGCTGTGTTGCTCGTCTT). *Eucalyptus grandis x E.urophylla* genomic DNA was used as a template control (Lane 1). cDNA from unbent control (Lane 2), and 6 hours, 24 hours, 1 week, 2 weeks and 3 weeks (Lanes 3-8) tension wood samples yielded a 280 bp product. The absence of the 1190 bp intron-spanning product in the cDNA samples confirms the absence of genomic DNA contamination in the samples. Lane 9: Molecular Weight standard (100 bp size standard, Fermentas, Ontario, Canada).

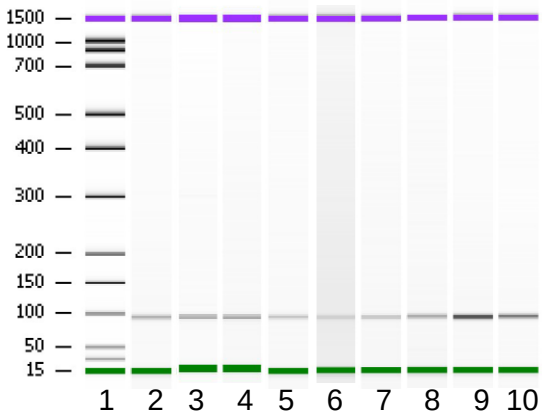


Figure 2.3: Agilent 2100 Bioanalyser DNA electrophoretic assay showing the pool of 96 bp DGE fragments purified for each of the constructed DGE libraries. Lane 1: Molecular Weight Standard, Lane 2: Unbent control, Lane 3: 6 hours, Lane 4: 24 hours, Lane 5: 1 week, Lane 6: 2 weeks, Lane 7: 6 months, Lane 8: Unbent control biological replicate 1, Lane 9: Unbent control biological replicate 2, Lane 10: Unbent control biological replicate 3.

2.3.3 Illumina GAII sequencing statistics

Nine *Eucalyptus* DGE libraries were sequenced with the Illumina Genome Analyser II at the University of the Western Cape to produce ten data sets (Table 2.3). The 3 weeks tension wood library failed and was excluded from further analyses. Biological replicate 2 of the unbent control, the DGE library with the highest concentration (Table 2.2), was selected as the library to be re-sequenced as a technical replicate to produce DGE data set number ten (Table 2.3). The total number of raw sequences produced per library ranged from 1.8 million to 9.1 million. After *in silico* removal of adapter sequences and filtering for the correct length of 16-17 bp, the total number of sequences was reduced to a range between 1.6 million and 8.6 million. For mapping to the *Eucalyptus grandis x Eucalyptus urophylla* (F1 hybrid, GUSAP1) transcriptome sequence, a maximum of two mismatches were allowed to account for sequencing error and for single nucleotide polymorphisms (SNPs) in the F1 hybrid transcriptome. Two mismatches allowed 80-85 % of the filtered sequences to map to the transcriptome (Table 2.3).

***Table 2.3*:** Illumina GAII sequencing statistics of each DGE library.

DGE library	Total [a]	Filtered [b]	Mapped [c]	% mapped [d]
Unbent Control	9,120,380	8,648,222	7,066,919	82
6 hours	8,363,198	7,869,491	6,422,940	82
24 hours	6,503,004	6,183,498	5,089,899	82
1 week	9,242,468	6,673,018	5,499,007	82
2 weeks	3,211,315	3,013,923	2,425,993	80
6 months	4,358,865	4,063,167	3,444,010	85
Unbent control biological rep 1	2,650,444	2,461,005	2,060,951	84
Unbent control biological rep 2	2,695,247	2,466,707	2,038,467	83
Unbent control biological rep 3	1,845,784	1,695,377	1,393,218	82
Technical repeat of biological rep 2	1,802,216	1,622,374	1,344,052	83

[a] Total initial sequences.

[b] Filtered for adapter and correct length.

[c] Number of reads that mapped to the *Eucalyptus grandis x Eucalyptus urophylla* transcriptome allowing for two mismatches.

[d] The percentage of filtered tags that mapped to the *Eucalyptus grandis x Eucalyptus urophylla* transcriptome assembly (18,894 transcript-based contigs >200 bp in length).

2.3.4 Trends in DGE data revealed by the manual analysis of seven well characterized genes

To gain a more detailed understanding of the nature and distribution of the DGE tags produced, seven well characterized genes were selected for manual analysis of the data produced by DGE profiling. These genes included four *Eucalyptus grandis* cellulose synthase (*EgCesA*) genes involved in primary cell wall formation and three *Eucalyptus grandis EgCesA* genes involved in secondary cell wall formation. Full length transcript assemblies were available for these genes (Mizrachi et al., 2010).

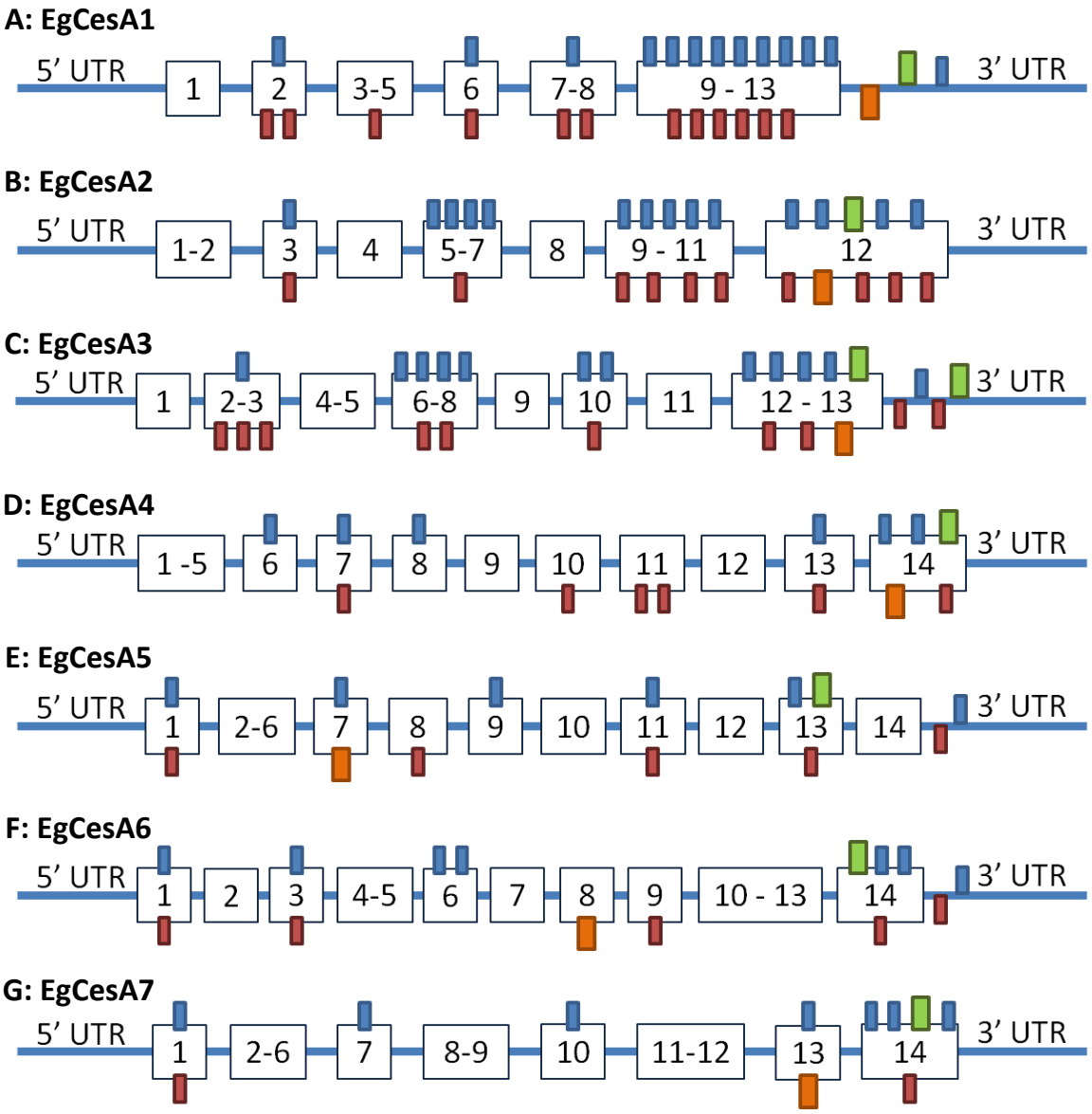


Figure 2.4: Distribution of positions that produced DGE tags (n≥1) for each of the *EgCesA* genes. *EgCesA1-3,* expressed at higher levels in secondary xylem tissues than *EgCesA4-7* produced tags from a greater number of DpnII recognition sites. Exons are depicted as boxes with the exon number inside the box and positions that produced tags are depicted as bars above or below a line that represents the genomic sequence. Blue bars above the sequence represent positions that produced sense tags and red bars below represent positions that produced antisense tags. Thick green bars represent the sense strand position that produced the tag with the highest count and thick orange bars represent the antisense strand position that produced the highest count. (Note: the bars are not scaled and do not indicate the level of expression).

Multiple tags were generated from each *EgCesA* transcript. These tags mapped to both the sense and antisense strands (Figure 2.4). The maximum number of DpnII recognition sites per gene is equal to the number of "GATC" tetranucleotide occurrences in the transcript sequence (which should be on average the transcript length divided by 256 bp in random DNA sequence). For the seven *EgCesA* genes the maximum number of DpnII restriction sites available was 28 with a DpnII site frequency as high as one every 140 bp (Table 2.4). The number of sense-strand positions that produced tags (n=1) ranged from 9 to 16 per *EgCesA* gene. Between 6 and 18 antisense positions produced tags on the transcripts. The antisense tags occurred with very low counts and, with the exception of *EgCesA6*, the sum of the counts of these antisense tags made up less than 10% of the total tag count of each *EgCesA* gene (Figure 2.5). The antisense tag count was not proportional to the number of positions that produced antisense tags. For example 73% of the *EgCesA2* antisense DpnII sites produced tags (Table 2.4), but these tags only constituted a maximum of 2% of the total tag count for *EgCesA2* (Figure 2.5) in any DGE library.

In each case, the position that produced the sense tag with the highest count (the major position) was located towards the far 3' end of the transcript and mapped to the sense strand (Figure 2.4). However, in contrast to what is expected from the DGE tag production procedure, this position was not always the most terminal 3' restriction site. The sum of the counts of all the tags at the major position accounted for up to 95% of the sum of all the tags across the entire transcript (Table 2.5). Only 60-81% of the total tag count mapped to the major position on *EgCesA2*, *EgCesA3* and *EgCesA6*. This was due to the presence of a second position that contained a substantial proportion of to the total count. The proportionally-higher levels of antisense counts detected for *EgCesA6* was in addition to the two positions that produced the majority of the total tag count (Figure 2.5). The second major position on *EgCesA2* that produced tags with high counts overlapped with the first major position and were four base pairs apart (Figure 2.6). Across all the DGE libraries the more 3' position contributed an average of 85% of the counts of the overlapping tags, while the adjacent 5' recognition site contributed 15% of the counts of the overlapping tags (Figure 2.7).

Table 2.4: Distribution of DpnII restriction sites for each *Eucalyptus CesA* gene and the frequency at which DGE tags were produced from the DpnII restriction sites. The maximum number of positions that produced sense and antisense tags is indicated for each *EgCesA* gene. The number of positions that produced sense and antisense tags are indicated as a percentage of the total number of DpnII restriction sites in each *EgCesA* gene.

	EgCesA1	***EgCesA2***	***EgCesA3***	***EgCesA4***	***EgCesA5***	***EgCesA6***	***EgCesA7***
Transcriptome assembly[a]							
Transcript assembly length (nt)[a]	3363	3499	3432	4417	3825	4214	4437
Number of DpnII restriction sites	23	25	21	22	21	20	28
Frequency of DpnII restriction sites (bp)	146	140	163	187	182	211	158
DGE data							
Max sense positions per gene	14	16	12	9	12	13	11
Max antisense positions per gene	17	18	14	12	6	6	6
% Sense positions observed	61	64	57	41	57	65	39
% Antisense positions observed	74	72	67	55	29	30	21
Average number of unique tags per gene	71	82	79	30	20	24	18
Average number of unique tags per position per gene	3.1	3.3	3.7	1.4	0.9	1.2	0.6

[a] Mizrachi et al., 2010.

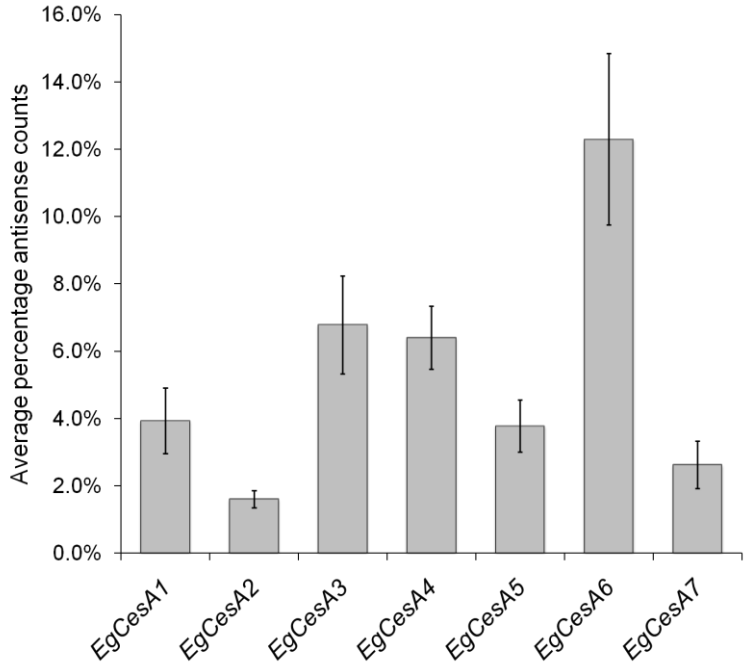


***Figure 2.5*:** The average percentage of antisense counts for each *EgCesA* gene across all ten DGE data sets (error bars indicate the standard error, n=10).

***Table 2.5*:** The percentage of unique tags that map to the position with the highest count and the percentage of the total count generated by this position.

	Ces A1	Ces A2	Ces A3	Ces A4	Ces A5	Ces A6	Ces A7
Average number of unique tags per gene	71	82	79	30	20	24	18
% of unique tags per position with the highest count	65	53	45	55	51	32	47
% of total tag count mapping to the position with the highest count	95	81	78	90	90	60	90

Nucleotide position on gene: 3148 - GATCGATCCGGTCCTGCCGAA
Nucleotide position on gene: 3152 - GATCCGGTCCTGCCGAAGCAA

***Figure 2.6*:** Nucleotide sequences of the overlapping tags from two DpnII recognition sites located close together on the *EgCesA2* gene.

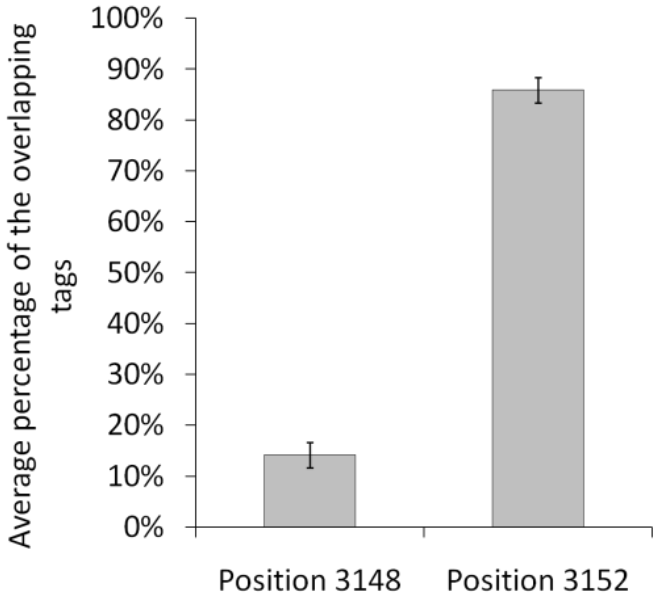


Figure 2.7: The percentage of overlapping counts produced from each of two adjacent overlapping DpnII restriction sites on *EgCesA2*. Position 3152 is the more 3' DpnII recognition site and represented tags from fully digested cDNAs. The low frequency of tags produced from the more 5' recognition site, position 3148 may represent cDNAs that were partially digested.

Unlike the two overlapping major tags of *EgCesA2*, the two major (non-overlapping) tags of *EgCesA3* exhibited a pattern of expression which may be indicative of alternative splicing (Figure 2.8). The two DpnII cut sites that contained 90% of the tags mapping to *EgCesA3* are located 300 bp apart. The more 5' position is located within the last exon of *EgCesA3* and the more 3' position is located 205 bp downstream from the last exon (Figure 2.4). At the beginning of the bending time series, the more 3' position contained approximately 80% of the tag count (Figure 2.8 A-B). The tag count produced at this position decreased progressively in the 6 hours, 24 hours, 1 week and 2 weeks DGE libraries (Figure 2.8 C-E). In the last library of the bending time series the situation is reversed with the position located more 5' producing more than 90% of the tag counts (Figure 2.8 F). For *EgCesA5* and *EgCesA6* antisense tags with the highest count were produced from the middle of the transcript, exon 7 and 8 respectively (Figure 2.4 E and F). In the rest of the genes, the position that produced the antisense tag with the highest count was located towards the far 3'-end of the transcript close to, or at the same position that produced the sense strand tag with the highest count (Figure 2.4 A – D and G).

An obvious correlation linked the number of positions that produced tags and the total number of produced tags to the size of the DGE library and the level of expression of the gene. A higher number of positions produced tags (Figure 2.9) and a greater number of tags mapped to each *EgCesA* gene (Figure 2.10) in DGE libraries with a greater total number of tags. A higher total number of tags mapped to *EgCesA1*, *EgCesA2* and *EgCesA3* (Figure 2.10) as would be expected from their levels of expression in secondary xylem tissues. For these three *EgCesA* genes, the only exons that did not produce tags were those that did not contain a DpnII recognition site. The tags that mapped to the major sense position had at least one sequencing error at every position along the tag sequence (Table 2.6). Positions with a count less than 10 make up less than 16% of the total tag count per gene (Figure 2.11).

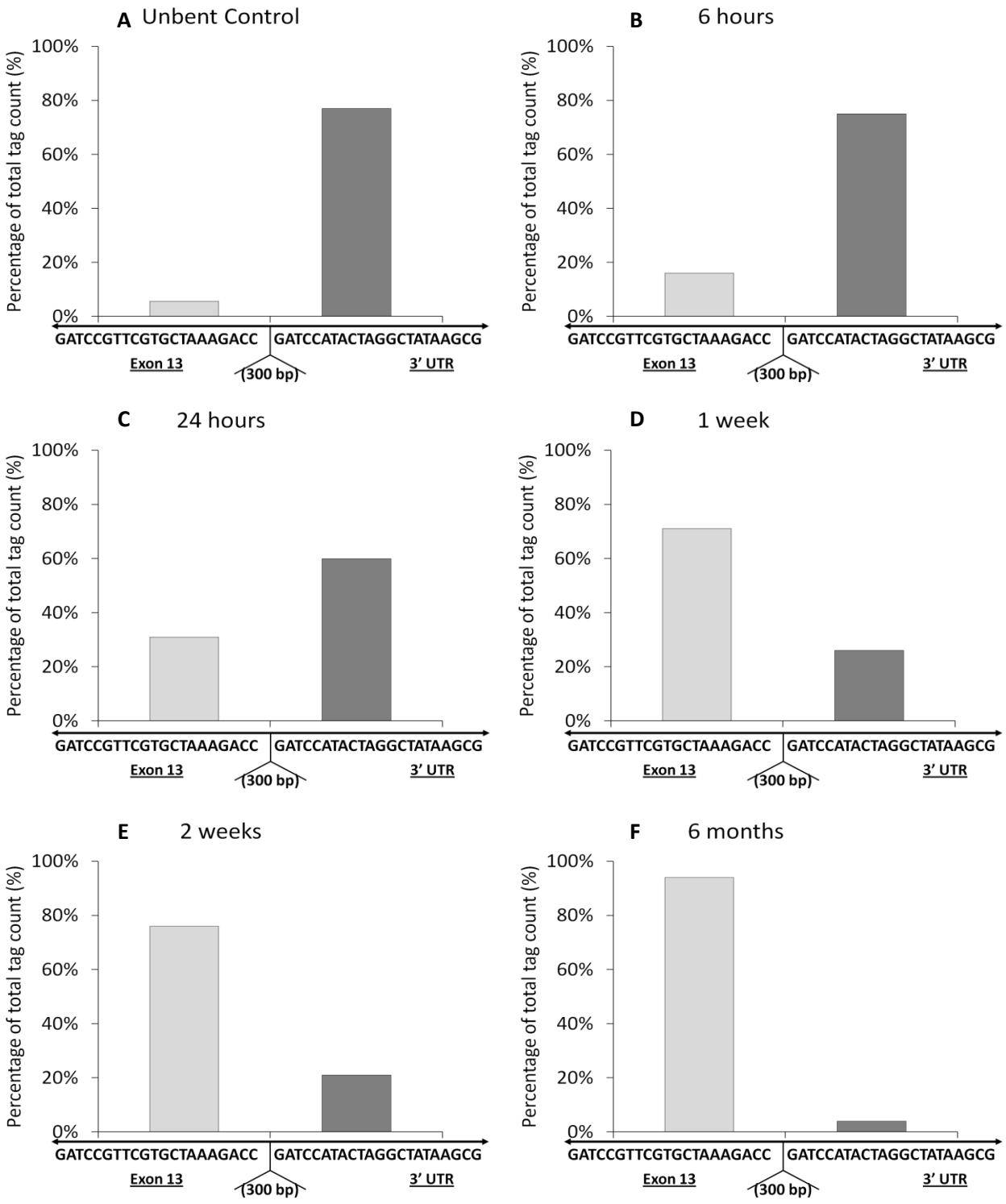


Figure 2.8: Two positions 300 bp apart, one located in the last exon and one located in the 3' UTR, alternately produce the majority of the tag count for *EgCesA3*. The tag sequence derived from each position is indicated below the histogram bar indicating the percentage of the total tag count. (A) Unbent control, (B) 6 hours, (C) 24 hours, (D) 1 week, (E) 2 weeks, (F) 6 months.

Table 2.6: Example of single nucleotide differences among the unique tags that mapped to position 3013 on *EgCesA1*, the position with the highest count. The unique tag with the highest count (91% of the total tag count at this position) is printed in bold and the count of each unique tag is given. The tag count is also represented as a percentage of the count at this position. All other unique tags are aligned to the most abundant tag and base pair differences are highlighted showing that sequencing errors occur throughout the entire tag sequence.

Sequence tag	Count	% of count	Sequence	Count	% of count
GATCTCAAGCTGTTTTTGCAG	10030	91	GATCTCCAGCTGTTTTTTCA	2	<1
GATCCCAAGCTGTTTTTGCAG	20	<1	GATCTCACGCTGTTTTTGCAG	3	<1
GATCTTAAGCTGTTTTTGCA	10	<1	GATCTCAAACTGTTTTTGCA	1	<1
GATCTCCAGCTGTTTTTGCAG	76	1	GATCTTAAGTTGTTTTTGCAG	8	<1
GATCTCANGCTGTTTTTGCAG	5	<1	GATCTCAAGCTNTTTTTGCA	2	<1
GATCTCAACCTGTTTTTGCAG	9	<1	GATCTCAAGCTGTTATTGCA	7	<1
GATCTCAAGGTGTTTTTGCA	1	<1	GATCTCAAGCTGTTTNTGCAG	4	<1
GATCTCAAGCCGCTTTTGCA	3	<1	GATCTCAAGCTGTTTTTCCAG	5	<1
GATCTCAAGCTCTTTTTGCAG	3	<1	GATCTCAAGCTGTTTTTCCA	5	<1
GATCTCAAGCTGNTTTTGCA	1	<1	GATCTCAAGCTGTTTTTTCC	6	<1
GATCTCAAGCTGTNTTTGCA	8	<1	GATCTCAAGCTGTTTTTTTCAG	171	2
GATCTCAAGCTGTTCTTGCA	1	<1	GATCTCAAGCTGTTTTTGCCG	348	3
GATCTCAAGCTGTTTNTGCA	1	<1	GATCTCAAGCTGTTTTTGCC	2	<1
GATCTCAAGCTGTTTTCGCAG	3	<1	GATCTCAAGCTGTTTTTGCN	4	<1
GATCTCAAGCTGTTTTTACA	2	<1	GATCTCACGCTGTTTTTGCC	2	<1
GATCTCAAGCTGTTTTTGNA	1	<1	GATCTCAAGCTGTTTTTTCA	1	<1
GATCTCAAGCTGTTTTTGCNG	2	<1	GATCTCAAGCTGTTTATGCA	8	<1
GATCTCAAGCTGTTTTTGCG	10	<1	GATCTCAAGCTGCTCTTGCA	3	<1
GATCTCAAGCTGTTTTTGTA	40	<1	GATCTCAAGCTTTTTTTGCA	25	<1
GATCTCAAGCTGTTTTTNCA	6	<1	GATCTCAACCTGTTTTTGCA	1	<1
GATCTCAAGCTGTTTCCGCA	7	<1	GATCTCACGCTGTCTTTGCA	51	<1
GATCTCAAGCTGTTTGTGCA	5	<1	GATCTCNAGCTGTTTTTGCA	5	<1
GATCTCAAGCTGTTNTTGCA	4	<1	GATCTNAAGCTGTTTTTGCA	1	<1
GATCTCAAGCTGTATTTGCA	7	<1	GATCCCAAGCTGTTTTTTCA	2	<1
GATCTCAAGCTGGTTTTGCA	4	<1	GATCGCAAGCTGTTTTTGCAG	2	<1
GATCTCAAGCTCTTTTTGCA	7	<1	GATCGCAAGCTGTTTTTGCA	3	<1
GATCTCAAGATGTTTTTTCAG	5	<1	GATCTCGAGCTGTTTTTGCA	37	<1
GATCTCAANCTGTTTTTGCAG	6	<1	GATCTCCAGCTGTTTTTGCC	1	<1
GATCTCAGGCTGTTTTTTCA	3	<1	GATCTCACCCTGTTTTTGCA	1	<1
GATCTCCAGCTGTTTTTGCCG	8	<1	GATCTCAGGCTGTTTGTGCA	2	<1
GATCTGAAGCTGTTTTTGAA	6	<1	GATCTCACGCTGTTTTTNCA	15	<1
GATCCCAAGCTGTTTTTGCA	1	<1	GATCTCAATCTGTTTTTTCA	15	<1
GATCTAAAGCTGTTTTTGCA	7	<1	GATCTCAANCTGTTTTTGCA	2	<1

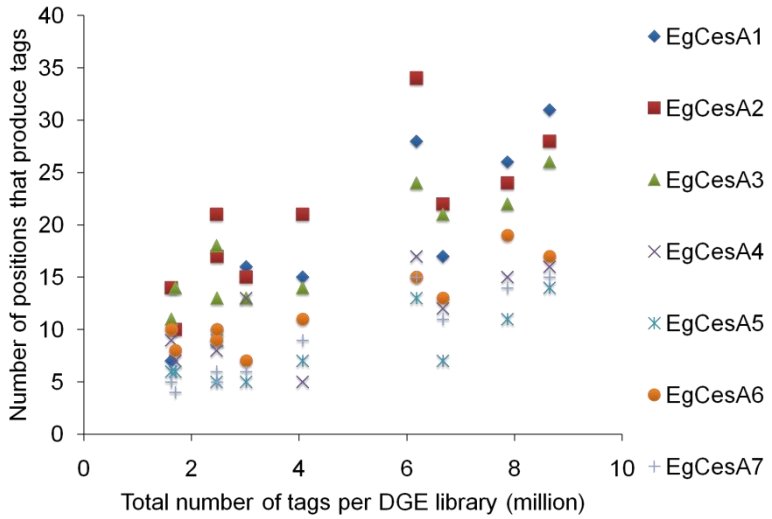


Figure 2.9: A linear increase was observed in the number of postions that produce tags in each *CesA* gene as the total number of sequenced tags increase per DGE library. x-axis: the total number of tags generated in each DGE library in millions. y-axis: the number of positions that produced tags in each *CesA* transcript.

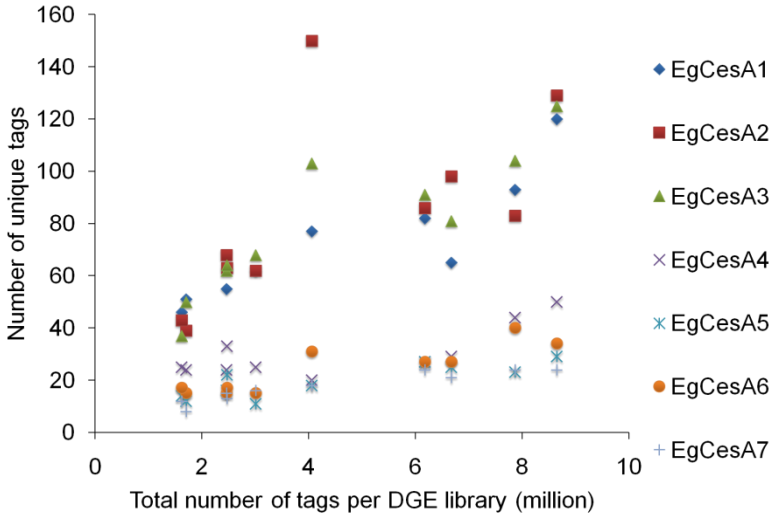


Figure 2.10: The total number of unique tags mapped to each *CesA* gene increased linearly with an increase in the total number of sequenced tags per DGE library. x-axis: the total number of sequenced tags per DGE library. y-axis: the number of unique tags produced from each *CesA* transcript.

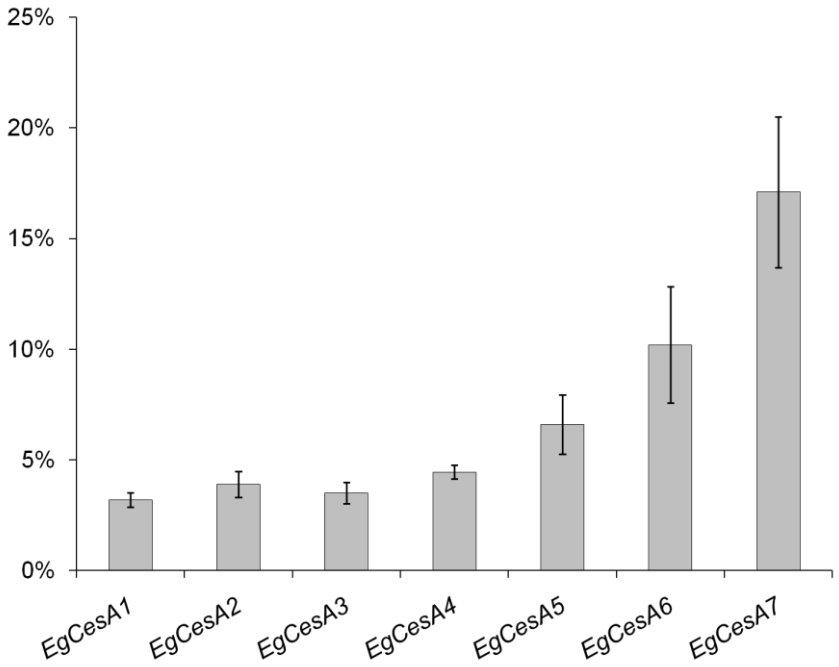


***Figure 2.11*:** The average percentage of the total count ascribed to tags with a count of 10 or less. Error bars indicate standard error, n=10.

2.3.5 Extraction of normalized expression values from DGE data and RT-qPCR verification

Manual analysis of the DGE data mapping to the secondary cell wall *EgCesA* genes allowed the development of processing guidelines for DGE data. The following guidelines for the automated processing of raw DGE data were developed: For each DGE library, 1. collapse tags overlapping by a minimum of 1 bp as a single position, 2. Discard positions that produce a tag count less than 10, 3. For each gene transcript sum the sense-strand position counts exceeding 10 into a single value, 4. For each gene transcript sum the antisense-strand position counts exceeding 10 into a single value and 5. Convert the total sense and antisense counts to parts per million (ppm) using the total number of tags (in millions) filtered for correct length. This yields two separate normalised expression values for each gene: a sense strand expression value and an antisense strand expression value.

The distribution of the counts per transcript after application of the DGE data processing rules is shown in Figure 2.12. The distribution of the tag counts showed similar trends for all of the DGE libraries and serves as an indication of reproducibility of DGE library sequencing. Transcription was detected for a maximum of 11,711 genes (Table 2.7). This number included transcripts to which sense and antisense tags mapped with a minimum count of 10. Antisense transcription was recorded for up to 54% of the genes detected by DGE and 11% showed antisense transcription only (Table 2.7). The detected dynamic range spanned three to four orders of magnitude from 1 ppm to 19,234 ppm. The highest expressed known transcript was that of a fasciclin-like arabinogalactan (FLA) protein in the 6 months DGE library, an ortholog of *AtFLA12* which exhibited 19,234 ppm sense strand counts. The average expression per gene in each DGE library was the same at 89 ppm indicating that normalisation was successful.

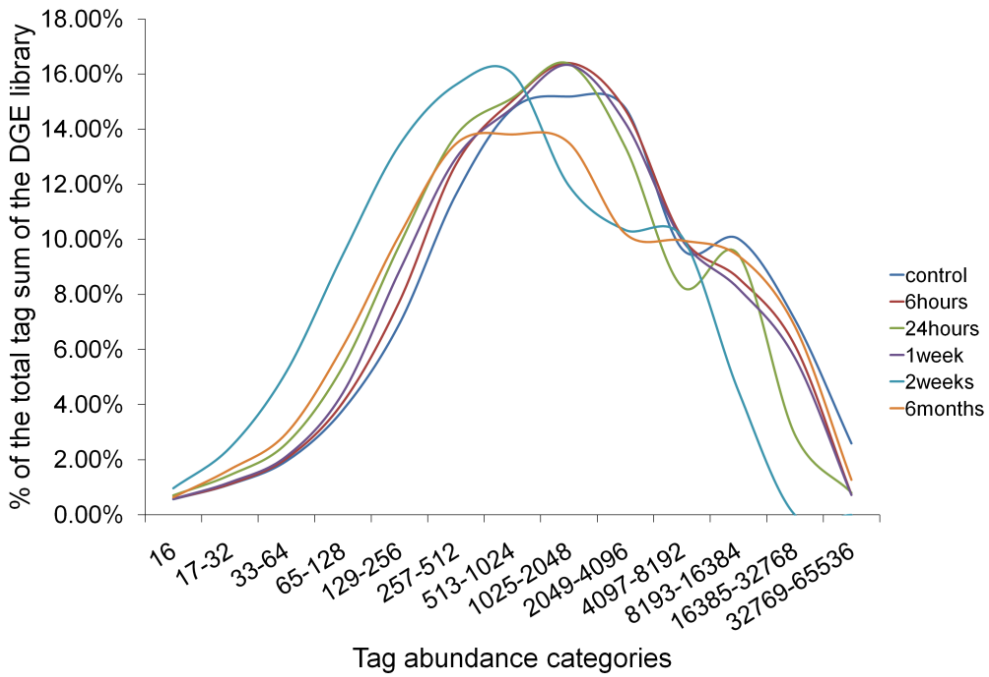


Figure 2.12: Distribution of tag counts in the DGE libraries. x-axis: Transcripts were grouped according to count in tag abundance categories. Each category increases in size exponentially. y-axis: the percentage of each tag count abundance category within the DGE library.

Table 2.7: The number and proportion of genes for which transcription was detected by DGE analysis. The detected genes are shown as a percentage of the total assembled *Eucalyptus grandis x E.urophylla* cDNA contigs (18,894). Also indicated is the percentage of detected genes that show sense, antisense and antisense transcription only.

DGE library	Genes detected[a]	% of transcriptome[b]	% Sense transcription	% Antisense transcription	% Antisense transcription only[c]
Unbent Control	11,711	62	89	54	11
6 hours	11,650	62	90	50	10
24 hours	11,320	60	90	49	10
1 week	11,126	59	89	43	11
2 weeks	9,759	52	90	29	10
6 Months	10,080	53	89	23	11

[a] At least one unique tag with a tag count exceeding 10.

[b] Percentage of 18,894 cDNA contigs (>200 bp) produced for the same F1 hybrid genotype by Mizrachi et al. (2010).

[c] Percentage of detected genes with at least one antisense tag exceeding 10 counts and no sense tags exceeding 10 counts.

The *EgCesA* DGE expression values in the unbent control revealed that the secondary cell wall *EgCesA* genes were expressed at higher levels than the primary cell wall *EgCesA* genes (Figure 2.13A) in normal xylem tissues as expected from previous studies (Ranik and Myburg, 2006). Results from an mRNA-seq study of some of the same tissues in the same *Eucalyptus grandis* x *Eucalyptus urophylla* genotype (GUSAP1, Mizrachi et al., 2010) are provided in Figure 2.13B. Expression values are presented as a ratio to the highest expressed gene in the set, *EgCesA1* to allow direct comparison between DGE and mRNA-seq data. The secondary cell wall *EgCesA* genes and the primary cell wall *EgCesA* genes showed similar ranges of expression in the DGE and mRNA-seq data sets. The minimum criteria for detection as outlined by the guidelines for processing DGE data (at least one tag exceeding 10 counts) resulted in the detection of low levels of antisense transcription for the *EgCesA* genes (Figure 2.14).

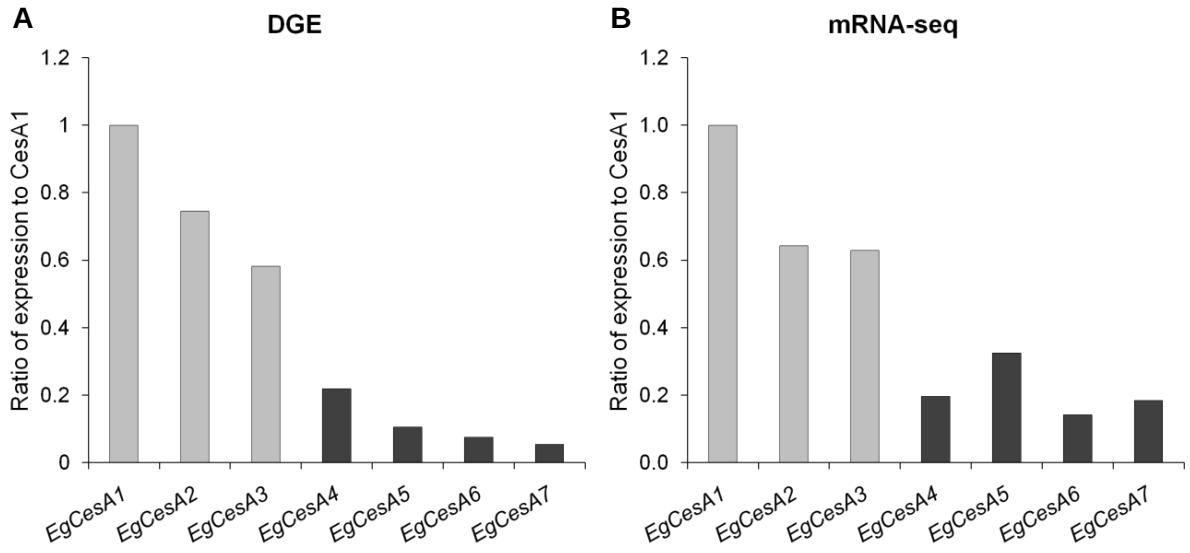


Figure 2.13: Ratio of the expression of the *EgCesA* genes to that of *EgCesA1*, the highest expressed gene in the set as quantified by DGE and mRNA-Seq. The DGE expression values are derived from the unbent control library. Comparison of expression profiles obtained by DGE (A) and mRNA-Seq (B) (Mizrachi et al., 2010), reveal a similar range of expression for the secondary cell wall *EgCesA* genes (grey) and the primary cell wall *EgCesA* genes (black).

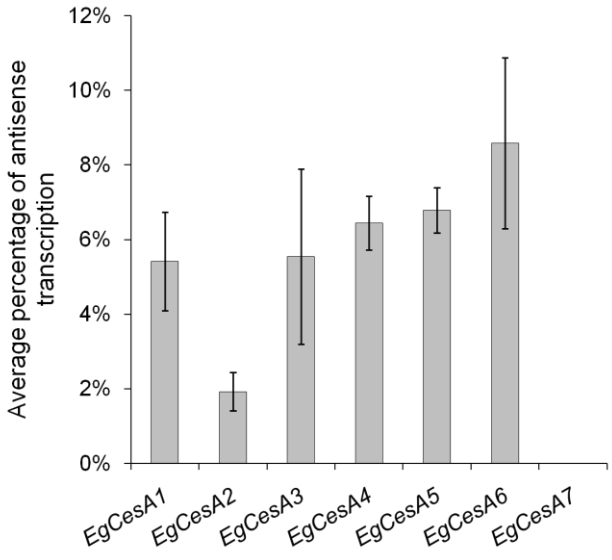


Figure 2.14: The average percentage of antisense transcription for each *EgCesA* gene across all DGE libraries after the application of the guidelines for the processing of DGE data. Error bars indicate standard error, n=10.

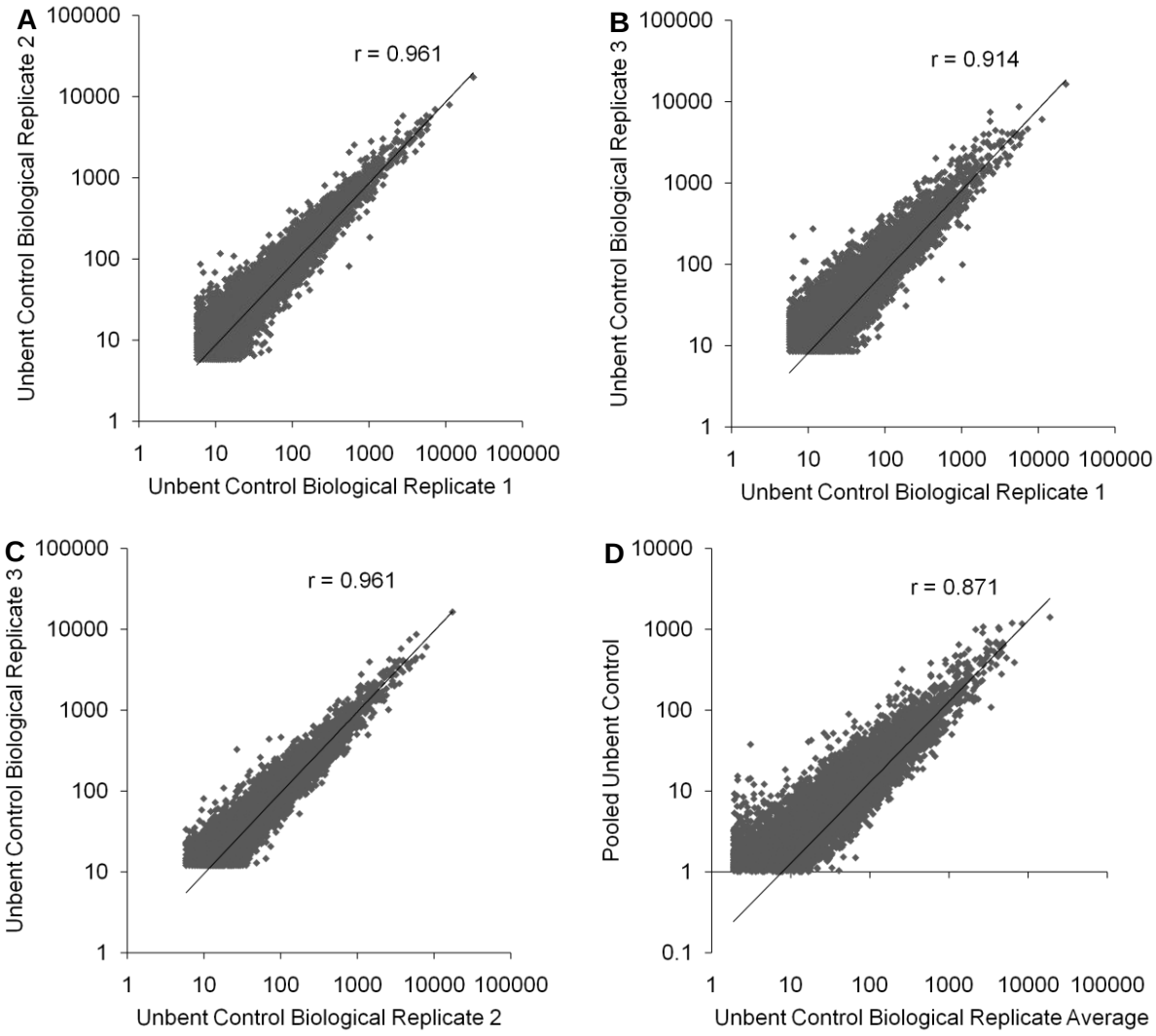


Figure 2.15: Biological and technical correlation of the Log transformed expression level in parts per million (ppm). Pearson correlation coefficients (r) are indicated on the graphs. Correlations are as follows: (A) Biological replicate 1 and Biological replicate 2 (B) Biological replicate 1 and Biological replicate 3 (C) Biological replicate 2 and Biological replicate 3 (D) Pooled library of unbent control and average values of the three biological replicates (E) Technical replicates of Biological replicate 2.

Across the libraries as a whole, application of the data analysis guidelines revealed pairwise correlations as high as 0.96 between the biological replicates of the unbent control sample (Figure 2.15A-C). The correlation between the pooled unbent DGE library and the average of the three biological replicates was 0.87 (Figure 2.15D). Technical reproducibility was marked by a correlation of 0.99 (Figure 2.15E).

The level of biological variation among the biological replicates of the unbent control and the pooled unbent control library were investigated for nine genes relevant to the process of tension wood formation. This includes genes involved in cellulose, hemicellulose and lignin biosynthesis, carbohydrate metabolism and transcriptional regulation (Table 2.1). The level of biological variation for these genes was also assessed using RT-qPCR. Melting curve analysis was performed to confirm the amplification of a single product (Figure S2.1). Each primer pair yielded a single melting curve peak. The amplification product for each primer pair was bulked and analysed with gel electrophoresis to confirm the amplification of the correctly sized fragment (Figure 2.16 and Table 2.1). The gene expression levels were normalised across tissues relative to the gene expression levels of the reference genes *ACT7* and *APT1*. The qBASE*plus* v1.0 average gene stability value (M) and average coefficient of variation (CV) were 0.123 and 4%, lower than 0.5 and 25% respectively, typically observed for stably expressed reference genes (Hellemans et al., 2007). Up to 70% biological variation was observed for the *KOR1* gene in the DGE dataset (Figure 2.17A), and up to 80% biological variation was observed for the *ARF6* gene in the RT-qPCR dataset (Figure 2.17B). Other genes however showed less variation among biological replicates. The biological replicates of *EgCesA3*, *EgCesA5* and *CCoAOMT* in the DGE set and *SUS1* in the RT-qPCR set differed by approximately 40%. Less than 20% biological variation was observed among the biological replicates of *GAUT12* and *SUS1* of the DGE dataset and *KNAT7* of the RT-qPCR dataset.

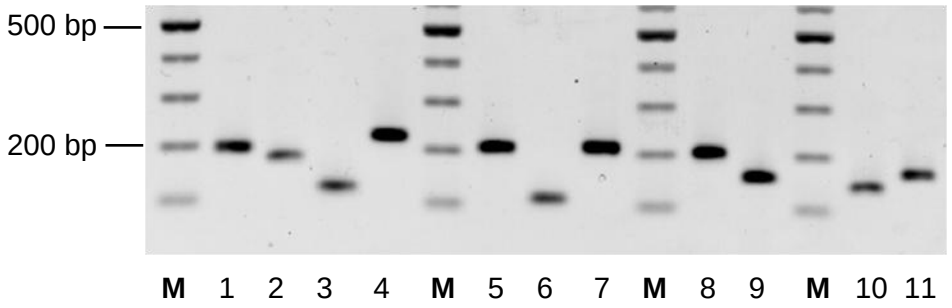


Figure 2.16: Agarose gel electrophoresis of bulked amplified fragments from genes analysed by RT-qPCR (Table 2.1). Lanes M: 100 bp DNA size standard, Lane 1: *CesA3*, Lane 2: *CesA5*, Lane 3: *KOR1*, Lane 4: *CCoAOMT*, Lane 5: *GAUT12*, Lane 6: *XTH23*, Lane 7: *SUS1*, Lane 8: *ARF6*, Lane 9: *KNAT7*, Lane 10: *ACT7*, Lane 11: *APT1*.

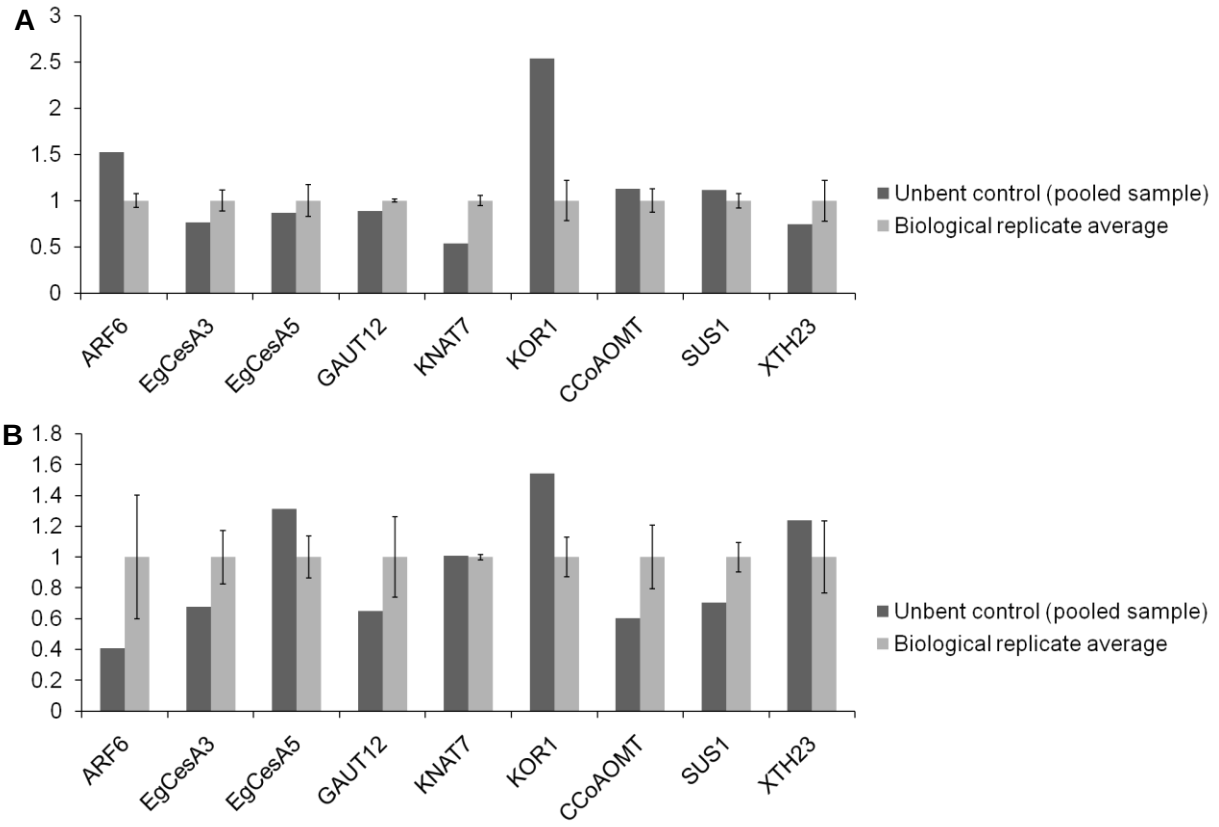


Figure 2.17: The observed biological variation for DGE (A) and RT-qPCR (B) among three biological replicates of the unbent control and the pooled unbent control sample. The pooled unbent control sample is represented as a ratio to the average expression of the three biological replicates of the unbent control. Error bars indicate standard error for the three biological replicates, n=3

2.4 Discussion

The technical nature of data generated by the Illumina DGE expression profiling technology was examined using the DGE data from a *Eucalyptus grandis x Eucalyptus urophylla* F1 hybrid tension wood formation time trial. Ten DGE data sets were generated from five pooled tension wood libraries, a pooled unbent control library, three biological replicates of the unbent control and a technical replicate of one of the unbent biological replicates. The five pooled DGE libraries were prepared from tension wood tissues collected from the *Eucalyptus* trees bent for 6 hours, 24 hours, 1 week, 2 weeks and 6 months. As DGE was a novel technology at the start of this study, a limited number of publications were available for comparison and guidance of the processing and analysis of DGE data (Hegedus et al., 2009, Ko et al., 2009, 't Hoen et al., 2008). Since then, several successful DGE-based studies have been published in a variety of fields including human development and genetics (Asmann et al., 2009, Pewzner-Jung et al., 2010), molecular virology (Xiao et al., 2010), evolutionary biology (Babbitt et al., 2010), and equine physiology (McGivney et al., 2010). The technical characteristics of the DGE data mapped to the *Eucalyptus EgCesA* genes were examined with the aim of developing data processing guidelines for the rest of the genes. *EgCesA* expression profiles obtained by processing the DGE data according to the developed guidelines compared well to the *EgCesA* expression profiles obtained with mRNA-Seq expression profiling in tissues of the same *Eucalyptus grandis x Eucalyptus urophylla* F1 hybrid (Figure 2.13). Biological variation among the unbent biological replicates was examined for individual genes and across the data set. The data of the biological replicates correlated well (Figure 2.15). The biological variation of a small subset of nine genes profiled with both DGE and RT-qPCR ranged from high (*ARF6* and *KOR1*) to very low (less than 20% for *GAUT12* and *KNAT7*, Figure 2.17).

2.4.1 Trends in DGE data revealed by the manual analysis of tags mapping to seven *CesA* genes

The proteins expressed from the *CesA* genes are essential to the formation of cellulose, one of the three main constituents of a plant cell wall. Based on their expression patterns, *CesA* genes are assigned to one of two groups (Somerville, 2006). During secondary xylem formation in *Eucalyptus* trees, the *CesA* genes are known to be highly transcribed with the primary cell wall *EgCesA* genes being transcribed at lower levels than the secondary cell wall *EgCesA* genes (Ranik and Myburg, 2006). Thus the manual analysis of the DGE data for to the *CesA* genes was relevant as these genes are known to be expressed during normal and altered xylem development. In addition, the *EgCesA* genes are well characterised which lends confidence to the interpretation of expression profiling data from these genes.

Tags produced from each *EgCesA* transcript mapped to the sense and antisense strands (Figure 2.4). DGE and SAGE studies have previously been shown capable of detecting antisense transcription (Babbitt et al., 2010, Birkeland et al., 2010, Ge et al., 2006, Hegedus et al., 2009, Quéré et al., 2004, Ruzanov and Riddle, 2010, 't Hoen et al., 2008, Xiao et al., 2010). The antisense tag production for each of the *EgCesA* genes (Figure 2.5) was in accord with other DGE studies where the sense-strand counts exceeded the antisense count by up to two fold (Babbitt et al., 2010, Hegedus et al., 2009, 't Hoen et al., 2008). Antisense transcription of up to 30% of the *Arabidopsis* transcriptome has previously been demonstrated (Yamada et al., 2003). This included transcription from antisense strands of annotated pseudogenes. We observed antisense transcription for 54% of previously assembled, expressed known *Eucalyptus grandis x Eucalyptus urophylla* genes. Crucial roles for non-coding RNAs in complex networks of transcriptional regulation are being revealed (Kurihara et al., 2009, Lanet et al., 2009, Song et al., 2009). Therefore the sense and antisense-strand DGE tags represent two functionally distinct outcomes of transcription and should not be considered together when measuring the transcript abundance produced by a gene.

Gene expression is a complex process affected by many transcriptional and post-transcriptional controls (Shyu et al., 2008). The DGE technology measures steady state transcript abundance, a single aspect of many that affects the final phenotypic expression of a gene. Therefore in a pool of sequenced tags, the frequency of the tag is an estimation of the abundance of the sampled transcript from which the tag was derived (Figure 1.5). However DGE tag generation is a multistep process and inefficiencies or errors at each step could introduce technical artefacts in the final data. The generation of multiple tags over the length of the *EgCesA* transcripts and not just at the 3' most GATC site (Figure 2.4) may be ascribed to two possible reasons. Sampling depth of the DGE technology may be so high that naturally 3'-end degraded RNAs with poly-A tails are sampled. The transient polyadenylation of truncated degradation intermediates as a means to assist exoribonucleases in rapid RNA degradation have been observed in bacterial (Wilusz and Wilusz, 2008), plant (Yehudai-Resheff et al., 2001) and human cells (Slomovic et al., 2010). The low abundance tags generated from multiple restriction sites on the *EgCesA* transcripts may serve as an indication that this process is utilised by *Eucalyptus* too. Asmann et al. (2009) estimated that for a transcript expressed at one copy per cell, a sequencing depth of 10 million tags per library will detect 30 tags for such a transcript. This would also be true for low abundance poly-adenylated degradation forms of a transcript. Alternatively, incomplete digestion by DpnII of the captured *CesA* transcripts would contribute to the generation of multiple tags at DpnII sites other than the 3' most site ('t Hoen et al., 2008, Xiao et al., 2010). The nucleotide composition of the *CesA* transcripts produce a frequency of DpnII recognition sites that is higher than the theoretical estimation of one restriction site every 256 bp (Karlin et al., 1992). This, along with partial digestion of the transcripts during the DGE library preparation, would increase the number of observed tags for each gene. Previous studies have either discarded tags from partially digested fragments, i.e. from alternative DpnII sites ('t Hoen et al., 2008), or considered these alternative tags in the calculation of the final abundance for a transcript (Xiao et al., 2010). From Figure 2.4 it is clear that for genes such as *EgCesA1*, *EgCesA2*, *EgCesA6* and *EgCesA7* considering only tags that map to the most 3' DpnII recognition site will disregard the main site producing the highest number of tags. Steric hinderance may be an explanation for the generation of the majority of the tags at a restriction site other than the most 3' site. It was

demonstrated that the enzymatic capability of DpnII, which functions as a dimer (Tran et al., 1998), is decreased greatly when insufficient space (less than DpnII dimer diameter) is allowed between surface bound DNA sequences (Castronovo et al., 2008). A restriction site too close to the bead may be protected from efficient binding of the DpnII dimer resulting in adjacent restriction sites 5' to these sites to be preferentially digested instead. The highest antisense counts for each *EgCesA* gene was not generated from the 5' most site, but close to or from the same DpnII site that generated the highest count sense tag. For *EgCesA5* and *EgCesA6* the highest count antisense tag was generated from the middle of the transcripts.

For *EgCesA2, EgCesA3* and *EgCesA6* only considering tags mapping to the major position would disregard a large number of tags (Table 2.5). The two positions that produce 90% of the total number of tags from *EgCesA2* are 4 bp apart - less than the 17 bp length of a single tag (Figure 2.4). The more 3' position produces 90% of the overlapping tags (Figure 2.7). This may indicate that the remaining 10% of overlapping tags produced from the adjacent site are not due to chance binding of DpnII to either of the sites but may be due to transcripts that escape the binding of DpnII at the other recognition site at the end of a digested fragment, leaving these fragments partially digested. In relation to an entire transcript, overlapping tags are essentially produced from the same position. Therefore mapped tags that overlap in sequence may be collapsed and summed as tags from a single position.

The two positions that alternately produced the major tag in *EgCesA3* revealed an interesting pattern. The tags, one located within the 3' UTR and the other located in the last exon, presented the highest count in the unbent control and 6 months libraries, respectively (Figure 2.8 A and F). These two libraries represent the extremes of the tension wood induction series. Also noticeable is a transition of the highest expressed tag in the intervening libraries of the time series (Figure 2.8 B - E). Alternative splicing was detected for the cellulose synthase gene *ZmCesA4c* in drought stressed maize (Eveland et al., 2008) and putative alternative splicing of *Zea mays CesA* genes *ZmCesA2* and *ZmCesA8* has also been recorded (Holland et al., 2000). An alternative splice variant of the ACC synthase gene (*PtaACS1*) is expressed in loblolly pine in response to bending stress (Barnes et al.,

2008). Furthermore, hormone treatment and numerous abiotic stresses have been found to induce alternative splicing in *Arabidopsis* (Filichkin et al., 2010, Palusa et al., 2007a, Palusa et al., 2007b). DGE has previously been demonstrated to detect tuberculosis infection-induced alternative splicing in zebrafish (Hegedus et al., 2009) and differences in hippocampus expression between wild type mice and transgenic mice over-expressing a splice variant of the doublecortin-like kinase-1 gene ('t Hoen et al., 2008). Therefore further investigation is needed to confirm the possibility of tension stress responsive alternative splicing in *EgCesA3*. Whether the incidence of multiple tags are ascribed to technical artefacts such as incomplete enzymatic digestion or to a biologically functional event such as alternative splicing, the tags still denote a quantifiable occurrence of that transcript in a pool of mRNA and should be considered as part of the expression value for that transcript.

Another source of multiple tag generation is due to sequencing error (Table 2.6). The number of positions and multiple tags produced from each position is correlated with, but not caused by, the sequencing depth and the level of expression of the gene (Figure 2.9 and Figure 2.10). It has been noted before that more erroneous tags differing by a single basepair are produced by highly abundant tags, simply because these are sequenced more (Philippe et al., 2009). High-throughput sequencing allows the DGE technology to detect all genes containing the DpnII recognition site and increases the detection sensitivity of low abundance transcripts. However from Table 2.6 and Figures 2.9-10 it can be inferred that increasing the sequencing depth beyond a certain point will not significantly increase the detection sensitivity of low abundance transcripts (Feng et al., 2010). Philippe et al. (2009) demonstrated that the DGE sequencing error rate of approximately 36% was halved when discarding tags with a single occurrence. Applying a widespread filtering criterion to exclude tags with a single occurrence allows for some erroneous tags to remain in the used data set. Thus sequencing error can be safely accounted for by discarding alternative tags with a count less than 10 without a large effect on the expression value for a gene. This can be done as, with the exception of *EgCesA7*, less than 10% of the total tag count for a gene consists of tags with a count less than 10 (Figure 2.11).

2.4.2 Guidelines for the processing of DGE data into expression values

2.4.2.1 Mapping and dynamic range

The number of genes detected by DGE analysis ranged between 52% and 62% of the total number (18,894) of assembled *Eucalyptus grandis x E.urophylla* cDNA contigs from a previous transcriptome sequencing effort involving the same clonal genotype (Mizrachi et al,. 2010). Ninety two percent of the total number of assembled transcripts contain a DpnII recognition site. Thus 92% of the transcripts have the potential of being detected, however it should be noted that two thirds of the *Eucalyptus grandis x Eucalyptus urophylla* transcripts are not full length (Mizrachi et al., 2010). Therefore genes for which no expression was detected may (i) be truly silent (i.e not expressed in xylem tissue), (ii) not contain a DpnII recognition site in its nucleotide sequence or (iii) be represented by a partially sequenced transcript. In a DGE study by Eveland et al., (2010) on maize meristem development, 66% of genes in the B73 genome catalog were detected by at least nine reads. Fifty four to sixty percent of the DGE data mapped to the maize reference sequence with no mismatches and one mismatch allowed for 85% to map to the reference sequence (Eveland et al., 2010). Human brain tissue represents a specialised tissue with complex gene expression patterns and the 54% of DGE data mapping to human brain protein-coding transcripts was lower than the 82% mapping achieved in this study (Asmann et al., 2009). The mapping of 80% of the total number of tags obtained by Hegedus et al. (2009) is consistent with the 82% mapping of total tags in this study (Table 2.3). The number of mapped tags should increase as the *Eucalyptus grandis x Eucalyptus urophylla* transcriptome sequence is updated and improved (Mizrachi et al., 2010). The pooled unbent control library was the library with the highest number of total tags generated and detected genes indicating a clear relationship between the sequencing depth and the number of genes detected. Thus the proportion of detected genes will improve with improvements to the sequencing chemistry and Illumina Pipeline Software (Asmann et al., 2009).

The distribution of tag abundance categories was similar for the six DGE libraries with few tags at either end of the expression range (Figure 2.12). The dynamic range of four orders of magnitude (1 to 19,234) is in accord with other DGE studies (Feng et al., 2010, Hegedus et al., 2009,

't Hoen et al., 2008). As a sequencing-based quantification method DGE quantification cannot be saturated like hybridization-based quantification technologies. This gives DGE a higher limit to the detection range which increases with the sequencing depth (Asmann et al., 2009, Babbitt et al., 2010).

2.4.2.2 Primary and secondary cell wall CesA expression in unbent Eucalyptus grandis x Eucalyptus urophylla

Processing of the DGE data using the developed guidelines for automated data processing generated expression profiles in accordance with profiles previously observed for the *Eucalyptus CesA* genes (Ranik and Myburg, 2006, Mizrachi et al., 2010). This validated the *EgCesA* expression profiles observed with DGE. In normal xylem, the secondary cell wall *CesA* genes are expressed higher than the primary cell wall *CesA* genes (Figure 2.13A). For *Eucalyptus* this was first demonstrated in *Eucalyptus grandis* trees using RT-qPCR (Ranik and Myburg, 2006). These expression patterns in normal xylem tissues are further supported by highly similar expression ratios of the primary and secondary *CesA* genes using DGE and mRNA-Seq quantification (Figure 2.13B, Mizrachi et al., 2010). This also provides good validation of the DGE data and lends confidence to the other tension wood DGE libraries, which were not biologically replicated.

2.4.2.3 Biological and technical variation

The biological variation for individual genes ranged from low (less than 20% variation) to high (up to 70% variation) using the DGE and RT-qPCR platforms (Figure 2.17). Inconsistent levels of variation were recorded by the DGE and RT-qPCR platforms. These inconsistencies may be due to the inherent differences between DGE and RT-qPCR. DGE is a tag-based sequencing approach and RT-qPCR relies on relative quantification based on fluorescence measurements. Normalisation of RT-qPCR data is more complex taking into account stable reference gene expression and gene-specific amplification efficiencies (Hellemans et al., 2007). For comparison between libraries, DGE normalisation involves a conversion of tag counts to parts per million.

The trial for this study, even though conducted with genetically identical ramets located next to each other, was conducted in a field where the environment is less controlled compared to trials conducted in controlled environments such as greenhouses. These results support previous arguments that expression profiles derived from a pooled sample does not always represent the average expression profile of biological replicates (Mary-Huard et al., 2007, Shih et al., 2004). Correlation between biological replicates as high as 0.96 was excellent (Figure 2.15A-C). Despite high biological variance among the pooled unbent sample and the unbent biological replicates for the individual genes profiled, the correlation between the pooled unbent sample and the average of the three unbent biological replicates was 0.87 (Figure 2.15D). It is therefore possible that the genes inspected individually showed high levels of biological variation, but that the majority of genes showed low levels of biological variation.

As previously demonstrated, regardless of a multistep procedure and the production of technical artefacts, the technical reproducibility of the DGE technology is excellent (Asmann et al., 2009, Feng et al., 2010, Hegedus et al., 2009, 't Hoen et al., 2008). The correlation of 0.99 between technical replicates for this study further supports the high technical reproducibility of DGE (Figure 2.15 E).

2.5 Conclusion

The guidelines presented here for the processing of DGE data performed well as the generation of expression profiles for the *CesA* genes during normal wood formation (unbent control) is comparable to two other studies (Ranik and Myburg, 2006, Mizrachi et al., 2010). The mapping statistics and detection of antisense transcription is in accordance with that of previous studies. However, guideline number 2: discard tags with a count of 10 or less, may be too stringent. The lower number of 62% detected genes as compared to the ≈75% of Hegedus et al., (2009), indicates that expression profiles of lowly expressed gene may have been discarded at this step. This is based on the observation that Hegedus et al. (2009) detected expression for a higher percentage of genes also making use of an

incomplete transcriptome. Other studies estimated that tags with a count as low as two represents a reliable measure of transcript abundance (Ko et al., 2009, Xiao et al., 2010), and Philippe et al. (2009) demonstrated that discarding singleton tags lowers the rate of false positive mapping events to below 15%. Thus lowering the level at which tags are filtered and discarded to two counts may be better suited to account for sequencing errors and include tags which represent true genes expressed at low levels in the investigated data set.

The technical variation was estimated separately with a correlation of 0.99. The correlations of 0.91 – 0.96 among the biological replicates show that these replicates are highly similar. The correlation of 0.87 between the pooled unbent control samples and the average of the three unbent biological replicates demonstrate that pooled samples may adequately represent the biological replicate average. Despite this high correlation, the genes investigated individually emphasise that the reliability and power of an expression profiling study is best when biological replicates are assessed separately.

2.6 References

ANDERSSON-GUNNERÅS, S., MELLEROWICZ, E. J., LOVE, J., SEGERMAN, B., OHMIYA, Y., COUTINHO, P. M., NILSSON, P., HENRISSAT, B., MORITZ, T. & SUNDBERG, B. (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *The Plant Journal,* 45, 144-165.

ASMANN, Y., KLEE, E., THOMPSON, E. A., PEREZ, E., MIDDHA, S., OBERG, A., THERNEAU, T., SMITH, D., POLAND, G., WIEBEN, E. & KOCHER, J.-P. (2009) 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genomics,* 10**,** 531.

BABA, K. I., PARK, Y. W., KAKU, T., KAIDA, R., TAKEUCHI, M., YOSHIDA, M., HOSOO, Y., OJIO, Y., OKUYAMA, T., TANIGUCHI, T., OHMIYA, Y., KONDO, T., SHANI, Z., SHOSEYOV, O., AWANO, T., SERADA, S., NORIOKA, N., NORIOKA, S. & HAYASHI, T. (2009) Xyloglucan for generating tensile stress to bend tree stem. *Molecular Plant,* 2**,** 893-903.

BABBITT, C. C., FEDRIGO, O., PFEFFERLE, A. D., BOYLE, A. P., HORVATH, J. E., FUREY, T. S. & WRAY, G. A. (2010) Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biology and Evolution,* 2**,** 67-79.

BARNES, J. R., LORENZ, W. W. & DEAN, J. F. D. (2008) Characterization of a 1-aminocyclopropane-1-carboxylate synthase gene from loblolly pine (*Pinus taeda* L.). *Gene,* 413**,** 18-31.

BARRETT, T. & EDGAR, R. (2006) Gene Expression Omnibus (GEO): Microarray data storage, submission, retrieval, and analysis. *Methods in Enzymology,* 411**,** 352-369.

BERNARD, P. S. & WITTWER, C. T. (2002) Real-time PCR technology for cancer diagnostics. *Clinical Chemistry,* 48**,** 1178-1185.

BIRKELAND, S. R., PREHEIM, S. P., DAVIDS, B. J., CIPRIANO, M. J., PALM, D., REINER, D. S., SVÄRD, S. G., GILLIN, F. D. & MCARTHUR, A. G. (2010) Transcriptome analyses of the *Giardia lamblia* life cycle. *Molecular and Biochemical Parasitology,* 174**,** 62-65.

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. & WITTWER, C. T. (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry,* 55**,** 611-622.

BUSTIN, S. A. & NOLAN, T. (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques,* 15**,** 155-166.

CALDANA, C., SCHEIBLE, W.-R., MUELLER-ROEBER, B. & RUZICIC, S. (2007) A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors. *Plant Methods,* 3**,** 7.

CANTACESSI, C., JEX, A. R., HALL, R. S., YOUNG, N. D., CAMPBELL, B. E., JOACHIM, A., NOLAN, M. J., ABUBUCKER, S., STERNBERG, P. W., RANGANATHAN, S., MITREVA, M. & GASSER, R. B. (2010) A practical, bioinformatic workflow system for large data sets generated by next-generation sequencing. *Nucleic Acids Research,* 38**,** e171.

CASTRONOVO, M., RADOVIC, S., GRUNWALD, C., CASALIS, L., MORGANTE, M. & SCOLES, G. (2008) Control of Steric Hindrance on Restriction Enzyme Reactions with Surface-Bound DNA Nanostructures. *Nano Letters,* 8**,** 4140-4145.

CHANG, S., PURYEAR, J. & CAIRNEY, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter,* 11**,** 113-116.

CHEN, J., HSUEH, H.-M., DELONGCHAMP, R., LIN, C.-J. & TSAI, C.-A. (2007) Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data. *BMC Bioinformatics,* 8**,** 412.

CLAIR, B., ALMÉRAS, T., YAMAMOTO, H., OKUYAMA, T. & SUGIYAMA, J. (2006) Mechanical behavior of cellulose microfibrils in tension wood, in relation with maturation stress generation. *Biophysical Journal,* 91**,** 1128-1135.

COPPÉE, J.-Y. (2008) Do DNA microarrays have their future behind them? *Microbes and Infection,* 10**,** 1067-1071.

DAVID, J.-P., COISSAC, E., MELODELIMA, C., POUPARDIN, R., RIAZ, M., CHANDOR-PROUST, A. & REYNAUD, S. (2010) Transcriptome response to pollutants and insecticides in the dengue vector Aedes aegypti using next-generation sequencing technology. *BMC Genomics,* 11**,** 216.

EVELAND, A. L., MCCARTY, D. R. & KOCH, K. E. (2008) Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiology.,* 146**,** 32-44.

EVELAND, A. L., SATOH-NAGASAWA, N., GOLDSHMIDT, A., MEYER, S., BEATTY, M., SAKAI, H., WARE, D. & JACKSON, D. (2010) Digital gene expression signatures for maize development. *Plant Physiology.,* 154**,** 1024-1039.

FENG, L., LIU, H., LIU, Y., LU, Z., GUO, G., GUO, S., ZHENG, H., GAO, Y., CHENG, S., WANG, J., ZHANG, K. & ZHANG, Y. (2010) Power of deep sequencing and agilent microarray for gene expression profiling study. *Molecular Biotechnology,* 45**,** 101-110.

FILICHKIN, S. A., PRIEST, H. D., GIVAN, S. A., SHEN, R., BRYANT, D. W., FOX, S. E., WONG, W.-K. & MOCKLER, T. C. (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Research,* 20**,** 45-58.

GARSON, J. A., HUGGETT, J. F., BUSTIN, S. A., PFAFFL, M. W., BENES, V., VANDESOMPELE, J. & SHIPLEY, G. L. (2009) Unreliable real-time PCR analysis of human endogenous retrovirus-w (HERV-W) RNA expression and DNA copy number in multiple sclerosis. *AIDS Research and Human Retroviruses,* 25**,** 377-378.

GE, X., WU, Q., JUNG, Y.-C., CHEN, J. & WANG, S. M. (2006) A large quantity of novel human antisense transcripts detected by LongSAGE. *Bioinformatics,* 22**,** 2475-2479.

GREGORY, B. D., YAZAKI, J. & ECKER, J. R. (2008) Utilizing tiling microarrays for whole-genome analysis in plants. *The Plant Journal,* 53**,** 636-644.

GUSTINCICH, S., SANDELIN, A., PLESSY, C., KATAYAMA, S., SIMONE, R., LAZAREVIC, D., HAYASHIZAKI, Y. & CARNINCI, P. (2006) The complexity of the mammalian transcriptome. *The Journal of Physiology,* 575**,** 321-332.

HANRIOT, L., KEIME, C., GAY, N., FAURE, C., DOSSAT, C., WINCKER, P., SCOTE-BLACHON, C., PEYRON, C. & GANDRILLON, O. (2008) A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. *BMC Genomics,* 9**,** 418.

HEGEDUS, Z., ZAKRZEWSKA, A., ÁGOSTON, V. C., ORDAS, A., RÁCZ, P., MINK, M., SPAINK, H. P. & MEIJER, A. H. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Molecular Immunology,* 46**,** 2918-2930.

HEID, C. A., STEVENS, J., LIVAK, K. J. & WILLIAMS, P. M. (1996) Real time quantitative PCR. *Genome Research,* 6**,** 986-994.

HELLEMANS, J., MORTIER, G., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology,* 8**,** R19.

HENDRIKS-BALK, M., MICHEL, M. & ALEWIJNSE, A. (2007) Pitfalls in the normalization of real-time polymerase chain reaction data. *Basic Research in Cardiology,* 102**,** 195-197.

HENE, L., SREENU, V., VUONG, M., ABIDI, S. H., SUTTON, J., ROWLAND-JONES, S., DAVIS, S. & EVANS, E. (2007) Deep analysis of cellular transcriptomes - LongSAGE versus classic MPSS. *BMC Genomics,* 8**,** 333.

HOLLAND, N., HOLLAND, D., HELENTJARIS, T., DHUGGA, K. S., XOCONOSTLE-CAZARES, B. & DELMER, D. P. (2000) A Comparative Analysis of the Plant Cellulose Synthase (CesA) Gene Family. *Plant Physiology,* 123**,** 1313-1324.

IOANNIDIS, J. P. A., ALLISON, D. B., BALL, C. A., COULIBALY, I., CUI, X., CULHANE, A. C., FALCHI, M., FURLANELLO, C., GAME, L., JURMAN, G., MANGION, J., MEHTA, T., NITZBERG, M., PAGE, G. P., PETRETTO, E. & VAN NOORT, V. (2009) Repeatability of published microarray gene expression analyses. *Nature Genetics,* 41**,** 149-155.

JIANG, S., XU, K., WANG, Y.-Z., REN, Y.-P. & GU, S. (2008) Role of GA3, GA4 and Uniconazole-P in Controlling gravitropism and tension wood formation in *Fraxinus mandshurica* Rupr. var. japonica Maxim. seedlings. *Journal of Integrative Plant Biology,* 50**,** 19-28.

JOSELEAU, J.-P., IMAI, T., KURODA, K. & RUEL, K. (2004) Detection in situ and characterization of lignin in the *G*-layer of tension wood fibres of *Populus deltoides*. *Planta,* 219**,** 338-345.

KARLEN, Y., MCNAIR, A., PERSEGUERS, S., MAZZA, C. & MERMOD, N. (2007) Statistical significance of quantitative PCR. *BMC Bioinformatics,* 8**,** 131.

KARLIN, S., BURGE, C. & CAMPBELL, A. M. (1992) Statistical analyses of counts and distributions of restriction sites in DNA sequences. *Nucleic Acids Research,* 20**,** 1363-1370.

KENDZIORSKI, C., IRIZARRY, R. A., CHEN, K. S., HAAG, J. D. & GOULD, M. N. (2005) On the utility of pooling biological samples in microarray experiments. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 4252-4257.

KLEBANOV, L. & YAKOVLEV, A. (2007) How high is the level of technical noise in microarray data? *Biology Direct,* 2**,** 9.

KO, J.-H., KIM, W.-C. & HAN, K.-H. (2009) Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in Arabidopsis. *The Plant Journal,* 60**,** 649-665.

KURIHARA, Y., MATSUI, A., HANADA, K., KAWASHIMA, M., ISHIDA, J., MOROSAWA, T., TANAKA, M., KAMINUMA, E., MOCHIZUKI, Y., MATSUSHIMA, A., TOYODA, T., SHINOZAKI, K. & SEKI, M. (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proceedings of the National Academy of Sciences,* 106**,** 2453-2458.

LANET, E., DELANNOY, E., SORMANI, R., FLORIS, M., BRODERSEN, P., CRETE, P., VOINNET, O. & ROBAGLIA, C. (2009) Biochemical evidence for translational repression by *Arabidopsis* microRNAs. *Plant Cell,* 21**,** 1762-1768.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* 10**,** R25.

LEAKEY, A. D. B., AINSWORTH, E. A., BERNARD, S. M., MARKELZ, R. J. C., ORT, D. R., PLACELLA, S. A., ROGERS, A., SMITH, M. D., SUDDERTH, E. A., WESTON, D. J.,

WULLSCHLEGER, S. D. & YUAN, S. (2009) Gene expression profiling: opening the black box of plant ecosystem responses to global change. *Global Change Biology,* 15**,** 1201-1213.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LIU, X., JIANG, J., SHAO, J., YIN, Y. & MA, Z. (2010) Gene transcription profiling of *Fusarium graminearum* treated with an azole fungicide tebuconazole. *Applied Microbiology and Biotechnology,* 85**,** 1105-1114.

MAQC.CONSORTIUM (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology,* 24**,** 1151-1161.

MARY-HUARD, T., DAUDIN, J.-J., BACCINI, M., BIGGERI, A. & BAR-HEN, A. (2007) Biases induced by pooling samples in microarray experiments. *Bioinformatics,* 23**,** i313-i318.

MCGIVNEY, B., MCGETTIGAN, P., BROWNE, J., EVANS, A., FONSECA, R., LOFTUS, B., LOHAN, A., MACHUGH, D., MURPHY, B., KATZ, L. & HILL, E. (2010) Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC Genomics,* 11**,** 398.

MECHAM, B. H., NELSON, P. S. & STOREY, J. D. (2010) Supervised normalization of microarrays. *Bioinformatics,* 26**,** 1308-1315.

MICHIELS, S., KOSCIELNY, S. & HILL, C. (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *The Lancet,* 365**,** 488-492.

MIZRACHI, E., HEFER, C., RANIK, M., JOUBERT, F. & MYBURG, A. (2010) De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics,* 11**,** 681.

MOYLE, R., SCHRADER, J., STENBERG, A., OLSSON, O., SAXENA, S., SANDBERG, G. & BHALERAO, R. P. (2002) Environmental and auxin regulation of wood formation involves members of the Aux/IAA gene family in hybrid aspen. *The Plant Journal,* 31**,** 675-685.

NELSON, C. D. & JOHNSEN, K. H. (2008) Genomic and physiological approaches to advancing forest tree improvement. *Tree Physiology,* 28**,** 1135-1143.

NETTLETON, D. (2006) A Discussion of statistical methods for design and analysis of microarray experiments for plant scientists. *Plant Cell,* 18**,** 2112-2121.

PALUSA, S. G., ALI, G. S. & REDDY, A. S. N. (2007a) Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *The Plant Journal,* 49**,** 1091-1107.

PALUSA, S. G., GOLOVKIN, M., SHIN, S.-B., RICHARDSON, D. N. & REDDY, A. S. N. (2007b) Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in Arabidopsis. *New Phytologist,* 174**,** 537-550.

PARKINSON, H., KAPUSHESKY, M., KOLESNIKOV, N., RUSTICI, G., SHOJATALAB, M., ABEYGUNAWARDENA, N., BERUBE, H., DYLAG, M., EMAM, I., FARNE, A., HOLLOWAY, E., LUKK, M., MALONE, J., MANI, R., PILICHEVA, E., RAYNER, T. F., REZWAN, F., SHARMA, A., WILLIAMS, E., BRADLEY, X. Z., ADAMUSIAK, T., BRANDIZI, M., BURDETT, T., COULSON, R., KRESTYANINOVA, M., KURNOSOV, P., MAGUIRE, E., NEOGI, S. G., ROCCA-SERRA, P., SANSONE, S.-A., SKLYAR, N., ZHAO, M., SARKANS, U. & BRAZMA, A. (2009) ArrayExpress update - from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Research,* 37**,** D868-D872.

PEWZNER-JUNG, Y., BRENNER, O., BRAUN, S., LAVIAD, E. L., BEN-DOR, S., FELDMESSER, E., HORN-SABAN, S., AMANN-ZALCENSTEIN, D., RAANAN, C., BERKUTZKI, T., EREZ-ROMAN, R., BEN-DAVID, O., LEVY, M., HOLZMAN, D., PARK, H., NYSKA, A., MERRILL, A. H. & FUTERMAN, A. H. (2010) A critical role for ceramide synthase 2 in liver homeostasis. *Journal of Biological Chemistry,* 285**,** 10911-10923.

PHILIPPE, N., BOUREUX, A., BRÉHÉLIN, L., TARHIO, J., COMMES, T. & RIVALS, E. (2009) Using reads to annotate the genome: influence of length, background distribution, and sequence errors on prediction capacity. *Nucleic Acids Research,* 37**,** e104.

QUÉRÉ, R., MANCHON, L., LEJEUNE, M., CLÉMENT, O., PIERRAT, F., BONAFOUX, B., COMMES, T., PIQUEMAL, D. & MARTI, J. (2004) Mining SAGE data allows large-scale, sensitive screening of antisense transcript expression. *Nucleic Acids Research,* 32**,** e163.

RANIK, M. & MYBURG, A. A. (2006) Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. *Tree Physiology,* 26**,** 545-556.

RUZANOV, P. & RIDDLE, D. L. (2010) Deep SAGE analysis of the *Caenorhabditis elegans* transcriptome. *Nucleic Acids Research,* 38**,** 3252-3262.

SCHENA, M., SHALON, D., DAVIS, R. W. & BROWN, P. O. (1995) Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science,* 270**,** 467-470.

SHI, L., TONG, W., FANG, H., SCHERF, U., HAN, J., PURI, R., FRUEH, F., GOODSAID, F., GUO, L., SU, Z., HAN, T., FUSCOE, J., XU, Z. A., PATTERSON, T., HONG, H., XIE, Q., PERKINS, R., CHEN, J. & CASCIANO, D. (2005) Cross-platform comparability of microarray technology: Intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics,* 6**,** S12.

SHIH, J. H., MICHALOWSKA, A. M., DOBBIN, K., YE, Y., QIU, T. H. & GREEN, J. E. (2004) Effects of pooling mRNA in microarray class comparisons. *Bioinformatics,* 20**,** 3318-3325.

SHYU, A.-B., WILKINSON, M. F. & VAN HOOF, A. (2008) Messenger RNA regulation: to translate or to degrade. *The european molecular biology organistaion journal,* 27**,** 471-481.

SLOMOVIC, S., FREMDER, E., STAALS, R. H. G., PRUIJN, G. J. M. & SCHUSTER, G. (2010) Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. *Proceedings of the National Academy of Sciences,* 107**,** 7407-7412.

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

SONG, D., YANG, Y., YU, B., ZHENG, B., DENG, Z., LU, B.-L., CHEN, X. & JIANG, T. (2009) Computational prediction of novel non-coding RNAs in *Arabidopsis thaliana*. *BMC Bioinformatics,* 10**,** S36.

TAN, P. K., DOWNEY, T. J., SPITZNAGEL JR, E. L., XU, P., FU, D., DIMITROV, D. S., LEMPICKI, R. A., RAAKA, B. M. & CAM, M. C. (2003) Evaluation of gene expression

measurements from commercial microarray platforms. *Nucleic Acids Research,* 31**,** 5676-5684.

'T HOEN, P. A. C., ARIYUREK, Y., THYGESEN, H. H., VREUGDENHIL, E., VOSSEN, R. H. A. M., DE MENEZES, R. E. X., BOER, J. M., VAN OMMEN, G.-J. B. & DEN DUNNEN, J. T. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Research,* 36**,** e141.

TRAN, P. H., KORSZUN, Z. R., CERRITELLI, S., SPRINGHORN, S. S. & LACKS, S. A. (1998) Crystal structure of the DpnM DNA adenine methyltransferase from the DpnII restriction system of Streptococcus pneumoniae bound to S-adenosylmethionine. *Structure,* 6**,** 1563-1575.

VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology,* 3**,** research0034.1 - research0034.11.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. (1995) Serial analysis of gene expression. *Science,* 270**,** 484-487.

WANG, X.-W., LUAN, J.-B., LI, J.-M., BAO, Y.-Y., ZHANG, C.-X. & LIU, S.-S. (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics,* 11**,** 400.

WHILEY, D. M., BIALASIEWICZ, S., BLETCHLY, C., FAUX, C. E., HARROWER, B., GOULD, A. R., LAMBERT, S. B., NIMMO, G. R., NISSEN, M. D. & SLOOTS, T. P. (2009) Detection of novel influenza A(H1N1) virus by real-time RT-PCR. *Journal of Clinical Virology,* 45**,** 203-204.

WILUSZ, C. J. & WILUSZ, J. (2008) New ways to meet your (3') end - oligouridylation as a step on the path to destruction. *Genes & Development,* 22**,** 1-7.

XIAO, S., JIA, J., MO, D., WANG, Q., QIN, L., HE, Z., ZHAO, X., HUANG, Y., LI, A., YU, J., NIU, Y., LIU, X. & CHEN, Y. (2010) Understanding PRRSV Infection in Porcine Lung

Based on Genome-Wide Transcriptome Response Identified by Deep Sequencing. *PLoS ONE*, 5**,** e11377.

YAMADA, K., LIM, J., DALE, J. M., CHEN, H., SHINN, P., PALM, C. J., SOUTHWICK, A. M., WU, H. C., KIM, C., NGUYEN, M., PHAM, P., CHEUK, R., KARLIN-NEWMANN, G., LIU, S. X., LAM, B., SAKANO, H., WU, T., YU, G., MIRANDA, M., QUACH, H. L., TRIPP, M., CHANG, C. H., LEE, J. M., TORIUMI, M., CHAN, M. M. H., TANG, C. C., ONODERA, C. S., DENG, J. M., AKIYAMA, K., ANSARI, Y., ARAKAWA, T., BANH, J., BANNO, F., BOWSER, L., BROOKS, S., CARNINCI, P., CHAO, Q., CHOY, N., ENJU, A., GOLDSMITH, A. D., GURJAL, M., HANSEN, N. F., HAYASHIZAKI, Y., JOHNSON-HOPSON, C., HSUAN, V. W., IIDA, K., KARNES, M., KHAN, S., KOESEMA, E., ISHIDA, J., JIANG, P. X., JONES, T., KAWAI, J., KAMIYA, A., MEYERS, C., NAKAJIMA, M., NARUSAKA, M., SEKI, M., SAKURAI, T., SATOU, M., TAMSE, R., VAYSBERG, M., WALLENDER, E. K., WONG, C., YAMAMURA, Y., YUAN, S., SHINOZAKI, K., DAVIS, R. W., THEOLOGIS, A. & ECKER, J. R. (2003) Empirical Analysis of Transcriptional Activity in the Arabidopsis Genome. *Science*, 302**,** 842-846.

YEHUDAI-RESHEFF, S., HIRSH, M. & SCHUSTER, G. (2001) Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts. *Molecular Cellular Biology*, 21**,** 5408-5416.

YUAN, J., REED, A., CHEN, F. & STEWART, C. N. (2006) Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7**,** 85.

ZHANG, W., CARRIQUIRY, A., NETTLETON, D. & DEKKERS, J. C. M. (2007) Pooling mRNA in microarray experiments and its effect on power. *Bioinformatics*, 23**,** 1217-1224.

2.7 Supplementary Tables and Figures

Table S2.1:

Illumina GAII DGE reads that mapped to the *EgCesA2* transcript in the unbent control library. For each sequenced tag is given the nucleotide sequence, the sense or antisense strand orientation, the nucleotide position to which the tag mapped and the corresponding count.

Tag	Sense /Antisense	Nucleotide position on transcript	Count
GATCGTTCGGAGAATGGGGAC	Sense	458	1
GATCGAGAAGTGGAAAATCA	Sense	583	1
CCTTGTCGTGGATTCTTGATC	Antisense	861	3
GATCAATTCCCTACGTGGAA	Sense	878	10
GATCAATTCCCTAAGCGGAAC	Sense	878	2
GATCAATTCCCTAAGTGGAAC	Sense	878	5
GATCAATGCATTGGTGGAAAA	Sense	1312	1
GATCCAGATTTATTTGGGGA	Sense	1417	2
GATCCAGGTTTATTTGGTAA	Sense	1417	4
CTATGTGTTTTTTAATGGATC	Antisense	1641	2
GATCGACATGACAGATATTC	Sense	1718	7
GATCGACATGACAGATATGCT	Sense	1718	7
GGCATTGTATGGGTATGATC	Antisense	1834	2
GATCCTCCAGTGTCCCAAAA	Sense	1850	3
GATCCTCCAGTGTCCCAAAAG	Sense	1850	2
GTCTGGACCTGTCTTTGATC	Antisense	2032	1
GTCTGGACCTGTCTTGGATC	Antisense	2032	6
GATCTAGAAGACATTAAAGAA	Sense	2048	11
GATCTAGAAGACATTGAATAA	Sense	2048	48
AGACCAGCTTTCAAGGGATC	Antisense	2372	1
AGACCAGCTTTAAAGGGATC	Antisense	2372	3
GATCAGCACCTATAAATCTG	Sense	2388	1
GATCAGCACCTATAAATCTGT	Sense	2388	1
GATCAACTCCATCAAGTTCT	Sense	2411	1
GATCGACTCCATNAAGTTCT	Sense	2411	1
GATCGNCTCCATCAAGTTCT	Sense	2411	1
GATCGACTCCATCACGTTCT	Sense	2411	4
GATCGACTCCATCAAGTTCTG	Sense	2411	1
GATCGACTCCATCAAGTTCTC	Sense	2411	133
ATAAACAAAGGAGACGGATC	Antisense	2969	2
ATAAACAATGGATGCGGATC	Antisense	2969	70
TTCTTTGCTTTTTGGGTGATC	Antisense	3014	2
TTCTTCGCCTTTTGGGTGATC	Antisense	3014	6
TTCTTGGCCTTTTGGGTGATC	Antisense	3014	27
TCTTTGCATTTTGGGTGATC	Antisense	3015	5
GATCGTCCATCTCTACCCTTT	Sense	3031	4

Tag	Sense /Antisense	Nucleotide position on transcript	Count
CAGAGCAGGACACGCACGATC	Antisense	3074	2
GAGAACAGGACACCCACGATC	Antisense	3074	2
AAAACAGGACACCCACGATC	Antisense	3075	1
AGAACAGGACACCCGCGATC	Antisense	3075	54
GATCGTGGCCCTTTTGTCCGT	Sense	3091	1
GATCGTGGTGCTTTGGTCCGT	Sense	3091	1
GATCGTTGTCCTTTGGTCCGT	Sense	3091	1
GATCGTGGTCCTTTGGTCCGC	Sense	3091	45
GATCGTGGTCCTTTGGTCNGT	Sense	3091	351
TGGTCTGGGGCCGGCTCGATC	Antisense	3135	23
GGTCTGGGTCCGGATCGATC	Antisense	3136	1
GGTCAGGGTCCGGCTCGATC	Antisense	3136	4
GATCGATCCGCTCCTGCCGA	Sense	3148	1
GATCGATCCGTTCCTGCCCA	Sense	3148	2
GATCGATCCGTTCCTTTCGA	Sense	3148	2
GATCGATCNGTTCCTGCCGA	Sense	3148	1
GATCGCTCCGTTCCTGCCGC	Sense	3148	1
GATCGNTCCGTTCCTGCCGA	Sense	3148	1
GATCAATCCGTTCCTGCCGA	Sense	3148	11
GATCGATCCGTTCCTGCCCAA	Sense	3148	2
GATCGATCCGTTCCTGCNGAA	Sense	3148	1
GATCGCTCCGTTCCTGCCGCA	Sense	3148	1
GATCGATCCGTTCCTGCCGGA	Sense	3148	4
GATCGATCCGTTCCTTCCGAA	Sense	3148	3
GATCGCTCCGTTCCTGCCGAA	Sense	3148	3
GATCGATCCGTTCCTGCCGAA	Sense	3148	609
GATCGATCCGTTCCTGCCGAC	Sense	3148	498
GATCGATCCGTTCCTGCCGAN	Sense	3148	21
GATCGATCCGTTCCTGCCGCA	Sense	3148	13
GATCGATCCGTTCCTGCCGCC	Sense	3148	17
GATCAGTTCCTGCCGAAGCA	Sense	3152	1
GATCCGCTCCTGCCGAAGCA	Sense	3152	1
GATCCGTACCTGCCGAAGCA	Sense	3152	2
GATCCGTTACTGCCGAAGCA	Sense	3152	1
GATCCGTTCCTGCCGAAGCC	Sense	3152	2
GATCCGTTCCTGCCGAAGCN	Sense	3152	1
GATCCGTTCCTGCCGAAGGA	Sense	3152	1
GATCCGTTCCTGCCGAAGTA	Sense	3152	2
GATCCGTTCCTGCCGAANCA	Sense	3152	2
GATCCGTTCCTGCCGACGAA	Sense	3152	1
GATCCGTTCCTGCCGACGCC	Sense	3152	1
GATCCGTTCCTGCCGANGCA	Sense	3152	1
GATCCGTTCCTGCCGATGCA	Sense	3152	1
GATCCGTTCCTGCCGATGCT	Sense	3152	2
GATCCGTTCCTGCCGGAGCC	Sense	3152	1

Tag	Sense /Antisense	Nucleotide position on transcript	Count
GATCCGTTCCTGCCGTAGCA	Sense	3152	2
GATCCGTTCCTGCNGAAGCA	Sense	3152	2
GATCCGTTTCTGCCGAAGCA	Sense	3152	1
GATCCTTTCCTGCCGAAGCA	Sense	3152	1
GATCCGATCCTGCCGAAGCA	Sense	3152	4
GATCCGTTACTGCCGAAGTA	Sense	3152	4
GATCCGTTCCCGCCGAAGCA	Sense	3152	3
GATCCGTTCCTGCAGAAGCA	Sense	3152	3
GATCCGTTCCTGCCAAAGCA	Sense	3152	4
GATCCGTTCCTGCCACAGCA	Sense	3152	3
GATCCGTTCCTGCCGAAACA	Sense	3152	4
GATCCGTTCCTGCCGAACCN	Sense	3152	7
GATCCGTTCCTGCCGAAGCG	Sense	3152	3
GATCCGTTCCTGCCGAATCA	Sense	3152	3
GATCCGTTCCTGCCGCAGCC	Sense	3152	4
GATCCGTTCCTGCCGCATCA	Sense	3152	10
GATCCGTTCCTGCTGAAGCA	Sense	3152	4
GATCCGTTCCTGCTGAATCA	Sense	3152	5
GATCCGTTCCTTCCGAAGAA	Sense	3152	3
GATCCGTTCTTGCCGAAGCA	Sense	3152	4
GATCGGCTCCTGCCGAAGCA	Sense	3152	3
GATCNGTTCCTGCCGAAGCA	Sense	3152	7
GATCTGCTCCTGCCGAAGCA	Sense	3152	5
GATCCGTTCCTGCCGAAGCA	Sense	3152	19
GATCCGTTCCTGCCGAGGCA	Sense	3152	13
GATCCGTTCCTGCCGGAGAA	Sense	3152	42
GATCCGGTCCTGCCGAAGCAA	Sense	3152	1
GATCCGTACCTGCCGAAGCAA	Sense	3152	1
GATCCGTTCCNGCCGAAGCAA	Sense	3152	2
GATCCGTTCCTGCCGTAGCCA	Sense	3152	2
GATCCGTTCCTTCCGAAGCAA	Sense	3152	2
GATCCGTTCTTGCCGAAGCAA	Sense	3152	2
GATCCGTTTCTGCCGAAGCAA	Sense	3152	2
GATCCGATCCTGCCGAAGCAA	Sense	3152	4
GATCCGTCCCTGCCGAAGCAA	Sense	3152	5
GATCCGTTCCCGCCGCAGCAA	Sense	3152	5
GATCCGTTCCTGCCGAATCAA	Sense	3152	7
GATCCGTTCCTGCCGACGCAC	Sense	3152	3
GATCAGTTCCTGCCGAAGCAA	Sense	3152	39
GATCCGCTCCTGCCGAAGCAA	Sense	3152	17
GATCCGTTCCTGCCGAAGCAA	Sense	3152	5685
GATCCGTTCCTGCCGAAGCAC	Sense	3152	225
GATCCGTTCCTGCCGAAGCCA	Sense	3152	59
GATCCGTTCCTGCCGACGCAA	Sense	3152	103
GATCCGTTCCTGCCGCAGCAA	Sense	3152	95

Tag	Sense /Antisense	Nucleotide position on transcript	Count
GATCTGTTCCTGCCGAAGCAA	Sense	3152	22
GATCAAACCGTGTGGGGTGG	Sense	3184	2
GATCGAAATCTCTTGATTAG	Sense	3336	2

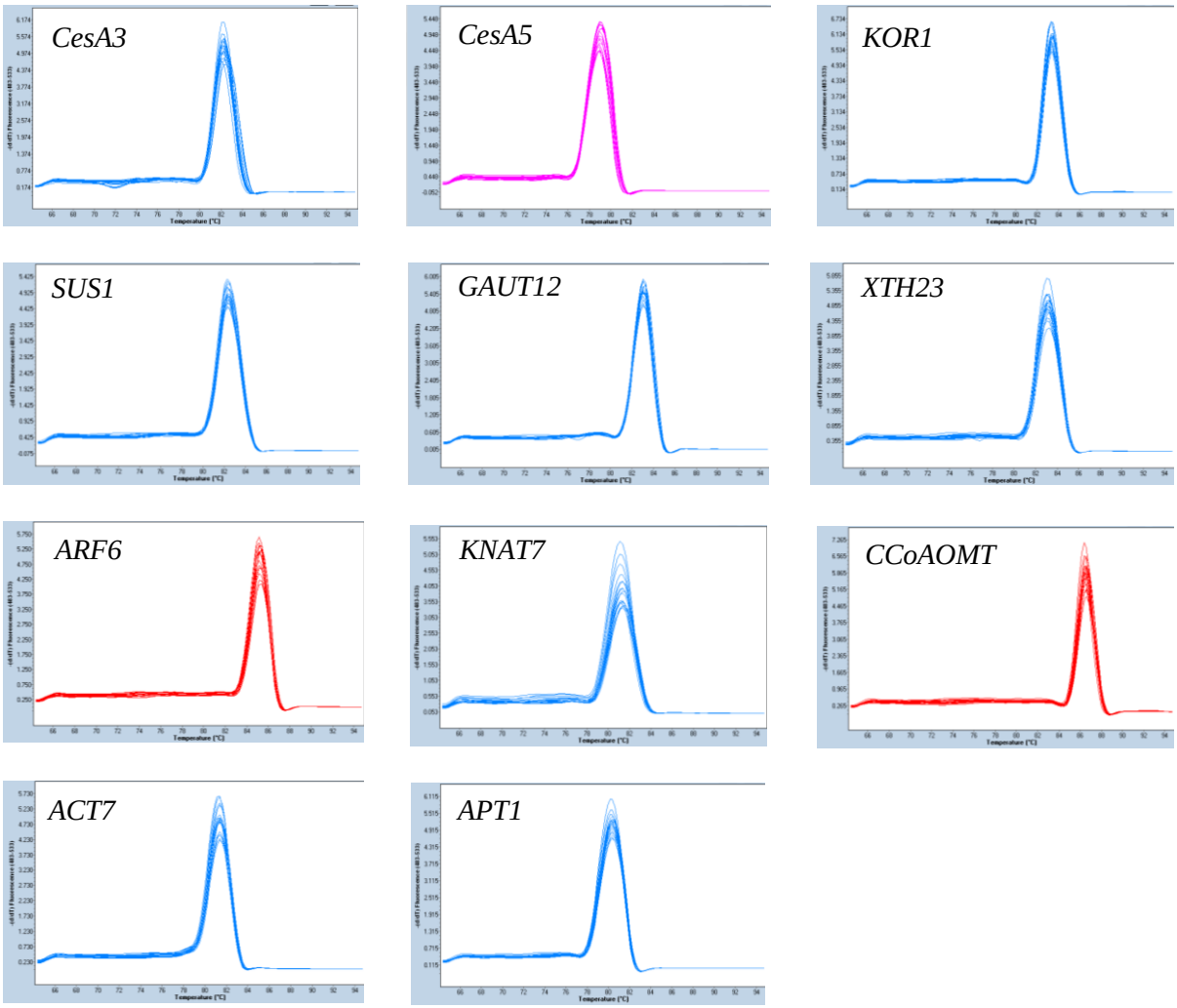


Figure S2.1: Melting curves of the RT-qPCR products for each gene-specific primer pair used for RT-qPCR. The y-axis shows rate of change (d/dT) in fluorescence and the x-axis shows temperature (°C).

Chapter 3

DEEP DIGITAL PROFILING OF GENE EXPRESSION DURING TENSION WOOD INDUCTION AND TENSION WOOD FORMATION IN *EUCALYPTUS*

J.F. Silberbauer[1], E. Mizrachi[1], C. Hefer[2], F. Joubert[2], D.K. Berger[3] and A.A. Myburg[1]

[1]Department of Genetics, [3]Department of Plant Science, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 2000, South Africa

[2]Bioinformatics and Computational Biology Unit, Department of Biochemistry, University of Pretoria, Pretoria, 2000, South Africa

Author contributions

I performed the wet laboratory work which includes the RNA extractions, DGE library preparation and RT-qPCR verification. I performed the data analysis assisted by E. Mizrachi who also guided me with the planning of the work. C. Hefer performed the bioinformatic programming for the data analysis. Prof. A.A. Myburg, Prof. D.K Berger and Prof F. Joubert provided advice, direction and supervision in the planning of the project, and performed critical review of the written chapter. Prof. Myburg conceived the main concept for the study and obtained funding for the research. All other technical assistance is acknowledged at the beginning of the thesis.

3.1 Introduction

The morphology of a tree is controlled in part by the formation of a specialised type of wood called reaction wood (Jiang et al., 2008). The reaction wood formed by Angiosperm genera such as *Eucalyptus* or *Populus* is termed tension wood as this type of wood functions in the generation of tensile maturation stress needed to stabilise branches and stems at specific angles (Alméras and Fournier, 2009; Bamber, 2001). Phototropic, gravitropic and mechanical stimuli contribute as cues to the induction of tension wood formation (Yamamoto et al., 2002). Adaptive organ reorientation has also been observed in cases where restoration or switching of apical dominance is required to replace a lost leader stem (Fisher and Stevenson, 1981, Wilson, 2000).

Compared to normal wood, tension wood exhibits altered physical, chemical and anatomical characteristics. During tension wood formation, increased cambial division and rapid growth results in asymmetrical thickening of the upper side of a branch or bent stem (Figure 3.1, Moyle et al., 2002). Rapid growth also results in an increased number of fibres and vessels with an increase in fibre cell length (Jourez et al., 2001). The secondary cell walls of the fibres are characterised by an increase in crystalline cellulose (Andersson-Gunnerås et al., 2006, Joseleau et al., 2004). Some hemicellulose and noncellulosic polysaccharide constituents such as the xylan content are decreased, while others show an increase (Bowling and Vaughn, 2008, Pilate et al., 2004a). The hemicellulose xyloglucan shows a large increase in tension wood tissues and is hypothesised to be one of the key players in the mechanism of action in tension wood (Baba et al., 2009). The crystalline cellulose formed by tension wood display a highly decreased cellulose microfibril angle (Baba et al., 2000). The fibre cells of some Angiosperm species including *Eucalyptus* and poplar have been shown to develop a specialised tension wood cell wall layer, the G-layer (Goswami et al., 2008). The tension wood mechanism of action, currently known as the growth stress hypothesis (Clair et al., 2006a), is mediated through the G-layer which replaces most of the S_2 and S_3 layers of a normal fibre wood cell (Fang et al., 2008). The G-layer is loosely linked to the other secondary cell wall layers and the altered chemical composition results in changed water absorption abilities needed for the tension wood mechanism of action (Gorišek and Straže, 2006). The formation of a G-layer is not essential for the formation and

action of tension wood as eucalypt tension wood does not always present with G-layer formation (Clair et al., 2006b, Fisher and Stevenson, 1981, Prodhan et al., 1995).

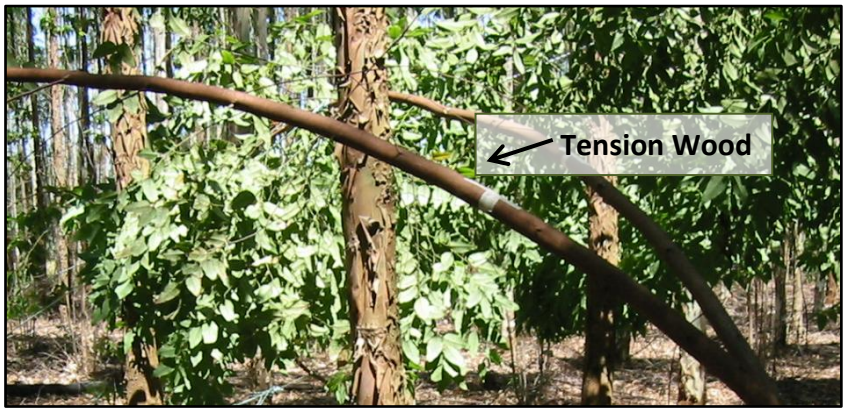


Figure 3.1: Position of tension wood formation. Rapid growth during tension wood formation results in asymmetrical thickening of the upper side of a bent stem with fibres that contain an increased content of crystalline cellulose and xyloglucan (Photo by Forest and Molecular Genetics Programme, University of Pretoria).

These altered characteristics of tension wood are the end result of transcriptome wide differential gene expression changes. A few key studies have examined the transcriptional changes during *Eucalyptus* and *Populus* tension wood formation (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007, Paux et al., 2005, Qiu et al., 2008). Gene expression changes are observed as early as six hours after tension wood induction (Paux et al., 2005). In *Eucalyptus*, a slight decrease in lignification is observed as the first phenotypic sign of immature tension wood formation one week after induction, however tension wood fibre cells at this stage are still not clearly distinguished from normal wood fibre cells (Baba et al., 2000). Changes in protein expression are clearly discernible two weeks after induction (Baba et al., 2000). The formation of mature tension wood is visible in fibre cells two – three weeks after induction (Baba et al., 2000, Paux et al., 2005). Transcriptional studies have found genes involved in cellulose biosynthesis and energy metabolism to be increased during tension wood formation. This includes sucrose synthase transcripts and transcripts involved in fructose and starch metabolism (Andersson-Gunnerås et al., 2006). A consistent response to tension

wood induction has not yet been recorded for the *CesA* genes. They have been observed to be strongly up-regulated, down-regulated and unaffected during poplar and eucalypt tension wood induction (Andersson-Gunnerås et al., 2006, Bhandari et al., 2006, Djerbi et al., 2004, Paux et al., 2005, Qiu et al., 2008). The transcriptional activity of lignin biosynthesis genes are reduced in accordance with the reduced lignin content of tension wood (Andersson-Gunnerås et al., 2006, Kaku et al., 2009, Paux et al., 2005). This includes genes involved in the early and late lignin biosynthesis pathways. The involvement of specific hormone signalling pathways in tension wood formation have also been investigated (Du and Yamamoto, 2003, Funada et al., 2008, Jiang et al., 2009, Moyle et al., 2002). Rapid growth during at the bent location of the stem tension wood formation is accompanied by a decrease in the flux through auxin signalling pathways and an increase in the flux through ethylene signalling pathways (Du and Yamamoto, 2003, Jiang et al., 2009, Moyle et al., 2002). The altered hemicellulose and non-cellulosic polysaccharide content have also been shown to be reflected in the tension wood transcriptome. An increased number of studies are linking fasciclin-like arabinogalactan (FLA) proteins and xyloglucan endotransglucosylase (*XTH/XET* genes) enzymes to important structural roles during tension wood formation (Baba et al., 2000, 2009; Qiu et al., 2008, Wasteneys and Fujita, 2006). Expression profiles of the *FLA* and *XTH* genes are among the highest up-regulated transcripts during tension wood formation (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007, Qiu et al., 2008).

Tension wood formation, compared to normal wood formation, gives the opportunity to associate gene expression changes with phenotypic properties so that the underlying molecular mechanisms may be clarified. Paux et al. (2005) profiled the *Eucalyptus* transcriptome during early tension wood formation up to one week after induction. Other transcriptome studies profiled gene expression during tension wood formation in *Eucalyptus* and *Populus* at one time point, three weeks after induction (Andersson-Gunnerås et al., 2006) or investigated naturally occurring branch tension wood (Qiu et al., 2008). Currently no tension wood induction trials spanning both early and late tension wood formation have been recorded. Second generation DNA sequencing-based transcript profiling applications are open ended platforms able to produce transcriptome-wide expression

profiles generating millions of gene-specific sequence tags that represent a digital read-out of the expression of each gene (Hanriot et al., 2008). Digital Gene Expression (DGE) profiling gives the opportunity to study the development of tension wood at an unprecedented depth. The richness and depth of information provided by such a study will enable the simultaneous investigation of a variety of cellular processes linked to tension wood formation.

The overall aim of this study was to generate an in-depth tension wood expression profiling data set for the investigation of tension wood induction and tension wood formation. The aim was to identify genes involved in tension wood induction and tension wood formation in *Eucalyptus grandis x Eucalyptus urophylla* trees and to elucidate the global transcriptional regulation of these processes. The first objective was to identify genes significantly differentially regulated during tension wood formation. Expression profiles of genes significantly differentially expressed from the unbent control were investigated as a preliminary exploratory inspection of tension wood expression profiles. This chapter is presented as a small pilot study of early (tension wood induction) and late tension wood formation. The final objective was to validate the tension wood digital gene expression profiles. This validation was performed using RT-qPCR on a selected subset of genes involved in secondary cell wall biosynthesis. Future work includes examining transcriptome wide early and late tension wood responses in *Eucalyptus grandis x Eucalyptus urophylla*.

3.2 Materials and Methods

The data of interest presented in this Chapter was derived from the generation and processing of six DGE libraries described in Chapter 2. These include the unbent control, 6 hours, 24 hours, 1 week, 2 weeks and 6 months libraries. The tension wood induction trial and sampling of plant material was described in sections 2.3.1 and 2.3.2. RNA was isolated using the CTAB method by (Chang et al., 1993) and assessed for quality and purity as described in section 2.3.3. The six DGE libraries were prepared from the pooled RNA samples using Illumina's DGE: Tag Profiling for DpnII Sample Prep kit according to the manufacturer's instructions and assessed for purity and yield as described in

section 2.3.4. Sequencing by synthesis was performed for each of the libraries using the Illumina Genome Analyser II system at the University of Western Cape, Cape Town, South Africa according to the manufacturer's protocols. Image analysis and base calling were performed using the Illumina data analysis pipelines associated with the instrument.

3.2.1 Mapping, processing and analysis of DGE data

The processing and mapping of the data was performed as described in Chapter 2 in section 2.2.4.1. Briefly, tags were mapped to the annotated assembled contigs of the *Eucalyptus grandis x E.urophylla* gene catalogue (Mizrachi et al., 2010) with Bowtie (Langmead et al., 2009) allowing for two mismatches per read. The automated processing of DGE data as described in Chapter 2 was performed with a customised Python script developed to extract the positions and counts for each tag mapped to each contig from the resulting SAM formatted file (Li et al., 2009). Tags that overlapped by a minimum of 1 bp were joined into a single count for the position, and only positions with a tag count of 10 or more were considered in the downstream analysis. The number of tags that mapped in the sense and antisense direction on the transcript were calculated separately and converted to parts per million. The expression profiles were extracted to Microsoft Office Excel for investigation. Expression profiles for 58 genes of interest (based on previous studies), named based on the *Arabidopsis thaliana* ortholog descriptions, were analysed with Microsoft Office Excel and the KMC application of TIGR_Mev v2.2. This was done by converting the tension wood parts per million (ppm) expression profiles to a ratio to the unbent control and converting these ratios to a Log_2 value. The significance of differential expression was tested using the Audic and Claverie (1997) test with Bonferroni correction through the program IDEG6 (Romualdi et al., 2003) and by employing a 2 fold change cut off in either direction.

3.2.2 Quantitative real-time reverse transcriptase PCR (RT-qPCR) analysis

Eight targets were selected for RT-qPCR validation (Table 2.1). All of the genes, except the *EgCesA* genes were named based on *Arabidopsis thaliana* ortholog descriptions. These targets include primary cell wall cellulose synthase 5 (*EgCesA5*; AT4G32410.1), secondary cell wall cellulose synthase 3 (*EgCesA3*; AT5G17420.1), korrigan (*KOR1*; AT5G49720.1), sucrose synthase (*SUS1*; AT3G43190.1), xyloglucan endotranscglucosylase (*XTH23*; AT4G25810.1), galacturonosyltransferase (*GAUT12*; AT5G54690.1), a knotted-like homeobox of *Arabidopsis thaliana* transcription factor (*KNAT7*; AT1G62990.1) and Caffeoyl-CoA 3-O-Methyltransferase (*CCoAOMT*; AT4G34050.1). Two genes with constitutive expression in the DGE data set were selected as control genes. These were Actin (*ACT7*; AT5G09810.1) and Adenine phosphoribosyl transferase (*APT1*; AT1G27450.1). Five genes were represented by primer pairs from previous studies in our research group. These were *CesA5*, *CesA3*, *KOR1*, *SUS1* and *CCoAOMT*. Gene specific primer pairs were designed for the remaining genes using Primer Designer 4 (Scientific and Educational Software, Cary, NC). All primer sequences are recorded in Section 2.8, Table 2.1.

RT-qPCR validation was performed as described in section 2.3.5 using the LightCycler® 480 instrument and LightCycler® 480 SYBR green I Master kit (Roche Diagnostics GmbH, Mannhein, Germany) according to the manufacturer's instructions. cDNA synthesis was carried out in technical duplicate for each sample and pooled. A total of 3 cDNA samples (pooled technical replicates) were synthesised as described in section 2.3.3 from the pooled unbent control, 24 hours and 2 weeks RNA samples. Melting curve analysis with agarose gel electrophoresis of the amplified product was performed. The qBase software was used for raw data quality control, PCR amplification efficiency correction and multiple reference gene normalisation (Hellemans et al., 2007). The results table was exported and graphs were produced using Microsoft Excel (Section 2.2.5).

3.3 Results

3.3.1 Bending trial and sampling of material

To investigate early tension wood induction and late tension wood formation in eucalypts, 18 month old *Eucalyptus grandis x Eucalyptus urophylla* trees (ramets of a single clonal genotype) were bent for 6 hours, 24 hours, 1 week, 2 weeks, and 6 months. Trees bent for 6 months responded by thickening two or three branches in a vertical orientation on the upper side of the bent stem (Figure 3.2 A). This was not observed at the earlier time points (up to 2 weeks). Discs cut from the region bent at 45° on the main bent stem showed a clear tension wood forming zone and asymmetrical growth (Figure 3.2B).

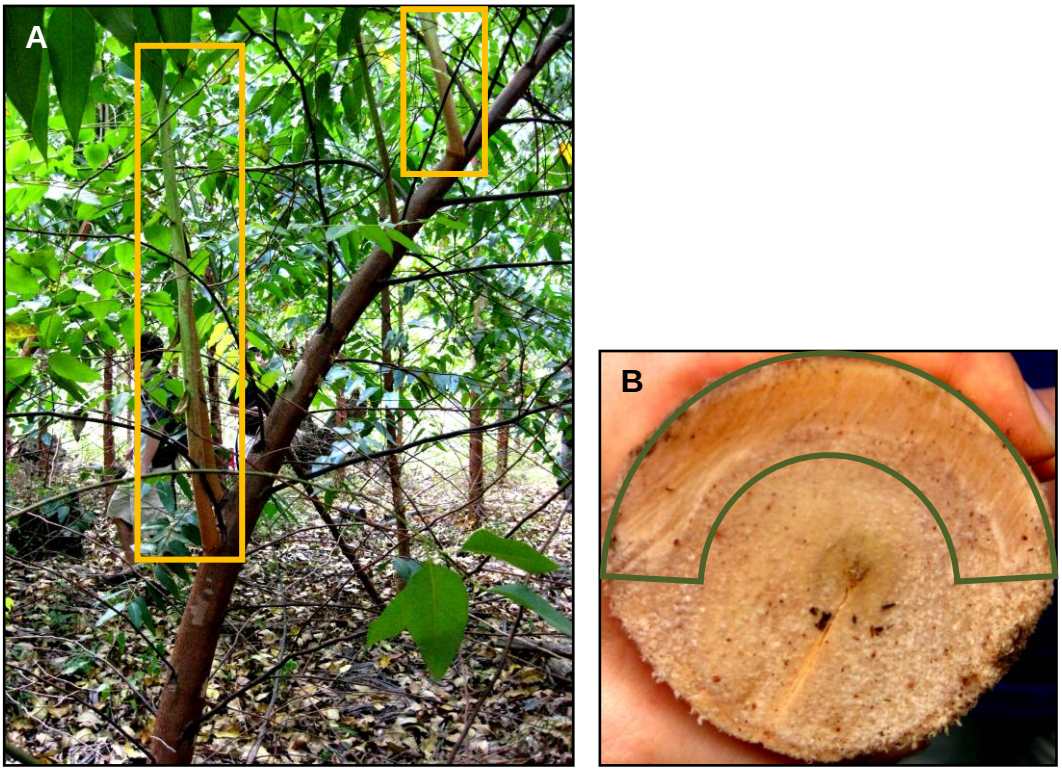


Figure 3.2: (A) Trees bent for 6 months responded by thickening 2-3 branches orientated into a vertical position. (B) A disc cut from a *Eucalyptus* tree bent at 45° for 6 months. A clear region of tension wood formation is visible (highlighted) by the asymmetrical growth on the upper side of the disc and at the formation of a hard, darkened half moon in this zone.

3.3.2 DGE library preparation and data processing

Six DGE libraries were prepared from good quality RNA extracted and pooled from three biological replicates of *Eucalyptus grandis x Eucalyptus urophylla* xylem tissues (Chapter 2, section 2.3.1 – 2.3.2). These libraries were prepared from tissues sampled from trees bent for 6 hours, 24 hours, 1 week, 2 weeks and 6 months to induce tension wood formation. The DGE sequencing statistics for these libraries are presented in Chapter 2, section 2.3.3. The DGE expression profiles were obtained by using the data processing guidelines developed in Chapter 2, section 2.3.5. Briefly a gene was considered to be expressed in a tissue sample when characterised by at least one DGE tag position with a count of at least 10. The expression levels were subsequently converted to parts per million (ppm). This procedure resulted in the detection of 11,711 genes expressed during tension wood formation (Chapter 2, Table 2.7).

3.3.3 Identification of genes significantly differentially expressed during tension wood formation

A few published DGE studies have assessed the significance of differentially expressed genes without separate, biologically replicated DGE libraries (David et al., 2010, Hegedus et al., 2009, Liu et al., 2010, Pewzner-Jung et al., 2010, Wang et al., 2010, Xiao et al., 2010). Four studies made use of the Audic and Claverie (1997) test which is adapted to tag sampling data (Liu et al., 2010, Pewzner-Jung et al., 2010, Wang et al., 2010, Xiao et al., 2010). This test assesses the probability that the expression of a gene in two different libraries was observed at different frequencies. In addition to a P-value the DGE study by David et al. (2010) also used a two-fold change in expression as a second criterion for significant differential expression. Therefore, the confidence that a gene is differentially expressed in the unbent control and a tension wood sample was represented by a P-value (Audic and Claverie, 1997, Romualdi et al., 2003) and by employing a two-fold change cut off in either direction. As would be expected, the number of significantly differentially expressed genes increased from the 6 hours tension wood library to the 6 months tension wood library and ranged from 431 (4% of genes

for which transcription was detected) to 2,654 (26% of genes for which transcription was detected, Table 3.1). The most responsive genes in the 6 hours and 24 hours libraries were up- and down-regulated by 9 fold (Figure 3.3 A and B). In the 1 week tension wood library the most responsive genes were differentially expressed by 9 fold (Figure 3.3 C). Differential expression increased even more in the 2 weeks and 6 months libraries to a maximum of 13 fold differential regulation (Figure 3.3 D and E).

Table 3.1: The number of genes significantly differentially expressed in each tension wood library compared to the unbent control.

DGE library	Up-regulated	Down-regulated	Total[a]	% of detected genes[b]
6 hours	282	149	431	4%
24 hours	350	127	477	4%
1 week	971	762	1733	16%
2 weeks	759	670	1429	15%
6 months	1130	1524	2654	26%

[a]Criteria for significant differential expression were a P-value $p<0.05$ (Audic and Claverie, 1997) and a two-fold change in expression.

[b] The total numbers of significantly differentially expressed genes are given as a percentage of the total number of genes for which expression was recorded (at least one tag position with a minimum count of 10) (Chapter 2, Table 2.7).

Thirty genes that showed significant differential up-regulation in each library are listed in Tables 3.2 – 3.6, and thirty genes that showed significant down-regulation in each library are listed in Tables 3.7-3.11. Genes significantly differentially regulated in libraries that represent the early tension wood (6 hours and 24 hours) response mostly included transcription factors and genes linked to abiotic and biotic stress responses. Genes involved in gravitropic and hormone signalling pathways were also

significantly differentially regulated. Genes involved in secondary cell wall biosynthesis pathways started to respond to the bending treatment in the 1 week library. Secondary cell wall biosynthetic genes showed strong significant responses to the bending in the late tension wood libraries (2 weeks and 6 months). Many stress response genes continued to show a response throughout the tension wood series.

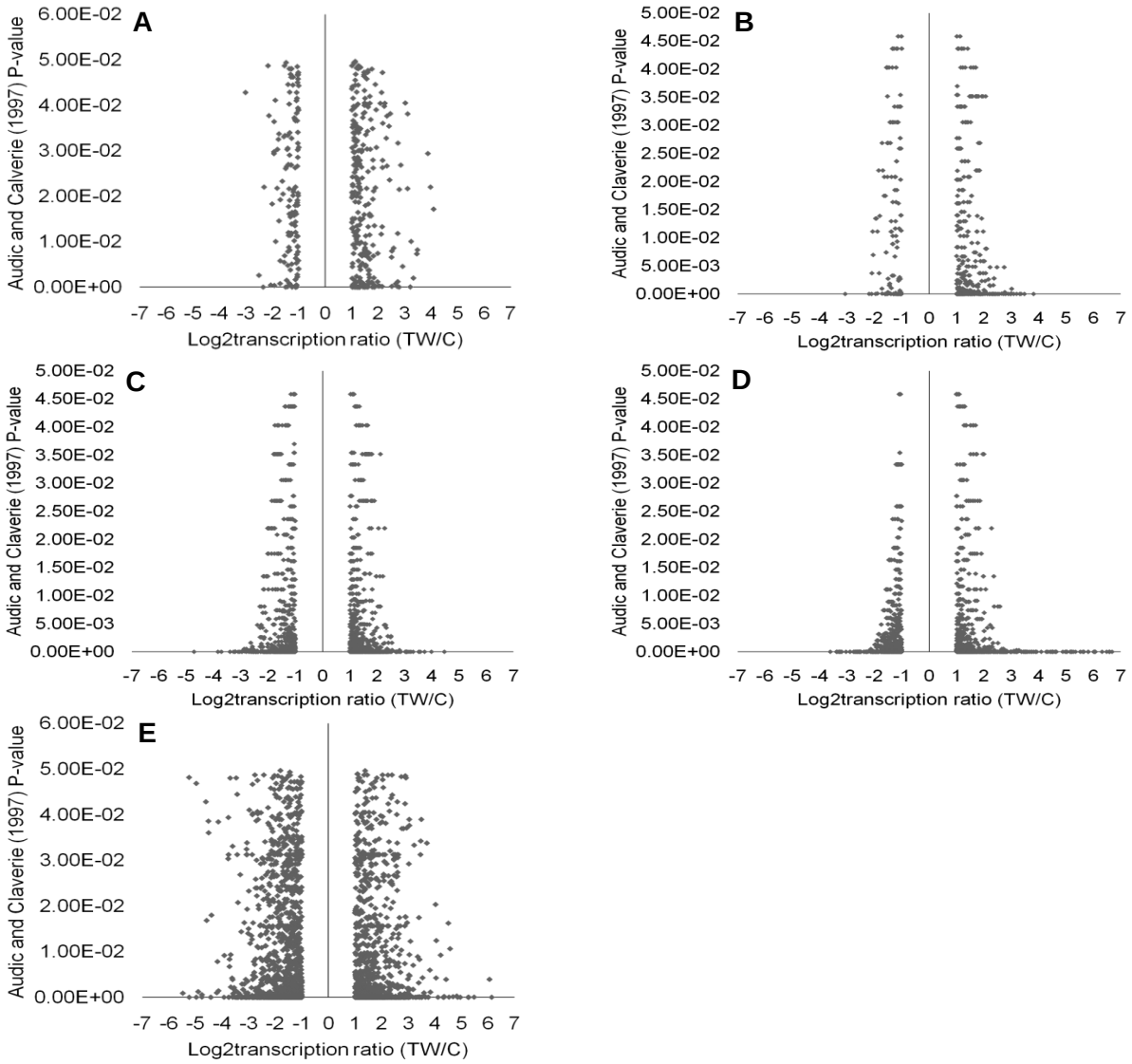


Figure 3.3: Fold change (Log_2) and significance of transcripts significantly differentially regulated during the tension wood formation time trial compared to the unbent control. Significant differential transcription is defined by Log_2(tension wood/unbent control) ratios and P-values as calculated by Audic and Claverie (1997). Only the genes which were differentially up- or down-regulated at a two-fold threshold with significance of $P < 0.05$ are considered significant and shown. (A) 6 hours, (B) 24 hours, (C) 1 week, (D) 2 weeks and (E) 6 months.

Table 3.2: Thirty genes significantly up-regulated at 6 hours after tension wood induction compared to the unbent control.

AT number	Description	Abbreviation	Fold up-regulation
AT3G12500.1	chitinase	*PR3*	9.13
AT2G23290.1	myb domain protein 70, transcription factor	*AtMYB70*	7.93
AT3G45140.1	lipoxygenase 2,	*LOX2*	7.48
AT2G47730.1	Glutathione s-transferase	*GSTF8*	6.50
AT4G17900.1	zinc-binding family protein	-	6.47
AT3G22400.1	lipoxygenase 5	*LOX5*	6.47
AT1G73500.1	map kinase kinase 9	*MKK9*	6.41
AT4G02380.2	senescence-associated gene 21	*SAG21*	6.04
AT1G60420.1	DC1 domain-containing protein	-	5.11
AT4G38400.1	*Arabidopsis thaliana* expansin-like A2	*EXPL2*	5.00
AT4G16563.1	aspartyl protease family protein	-	4.95
AT3G10190.1	calmodulin, putative	-	4.73
AT4G21970.1	expressed Protein	-	4.49
AT4G25810.1	xyloglucan endotransglycosylase 6	*XTR6/XTH23*	4.38
AT1G13260.1	transcription factor	*RAV1*	4.31
AT1G70810.1	C2 domain-containing protein	-	4.31
AT1G19180.1	jasmonate-zim-domain protein 1	*JAZ1*	4.29
AT4G21870.1	26.5 kDa class P-related heat shock protein	*HSP26.5-P*	4.29
AT4G20780.1	calcium-binding protein, putative	-	4.24
AT4G25470.1	C-repeat/DRE binding factor 2	*CBF2*	4.11
AT3G15660.2	glutaredoxin 4	*GRX4*	4.11
AT4G14130.1	xyloglucan endotransglycosylase 7	*XTR7*	3.62
AT3G16350.1	myb family transcription factor	-	3.29
AT3G23150.1	ethylene response 2	*ETR2*	2.97
AT5G13180.1	*Arabidopsis* NAC domain containing protein 83	*ANAC083*	2.66
AT1G74840.1	myb family transcription factor	-	2.49
AT1G09960.1	sucrose transporter 4	*SUT4*	2.42
AT2G24570.1	transcription factor	*WRKY17*	2.37
AT3G17860.1	jasmonate-zim-domain protein 3	*JAZ3*	2.25
AT2G25490.1	EIN3-binding F box protein 1	*EBF1*	2.10

Table 3.3: Thirty genes significantly up-regulated at 24 hours after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold up-regulation
AT4G25470.1	C-repeat/DRE binding factor 2	*CBF2*	6.96
AT2G29500.1	17.6 kDa class I small heat shock protein B-CI	*HSP17.6B*	6.03
AT3G09640.2	ascorbate peroxidase 2	*APX2*	5.98
AT4G34410.1	redox responsive transcription factor 1	*RRTF1*	5.53
AT2G40140.2	transcription factor	*CZF1*	5.43
AT5G63160.1	BTB and TAZ domain protein 1	*BT1*	4.96
AT4G11280.1	ACC synthase 6	*ACS6*	4.83
AT4G14130.1	xyloglucan endotransglycosylase 7	*XTR7*	4.81
AT5G42380.1	calmodulin like 37	*CML37*	4.75
AT4G02380.2	senescence-associated gene 21	*SAG21*	4.75
AT5G15410.1	Defense no death 1; calcium channel	*DND1*	4.45
AT4G23810.1	transcription factor	*WRKY53*	4.40
AT5G11260.1	elongated hypocotyl 5	*HY5*	4.13
AT5G14920.1	gibberellin-regulated family protein	-	3.95
AT5G03170.1	fasciclin-like arabinogalactan protein 11	*FLA11*	3.80
AT4G25810.1	xyloglucan endotransglycosylase 6	*XTR6/XTH23*	3.79
AT2G01850.1	xyloglucan:xyloglucosyl transferase	*XTH27*	3.56
AT3G16350.1	myb family transcription factor	-	3.56
AT1G33060.2	no apical meristem (NAM) family protein	*ANAC014*	3.51
AT2G38080.1	irregular xylem 12, laccase	*IRX12/LAC4*	3.61
AT5G13200.1	GRAM domain-containing protein / ABA-responsive protein-related	-	3.59
AT3G16350.1	myb family transcription factor	-	3.56
AT5G59845.1	gibberellin-regulated family protein	-	3.06
AT1G17870.1	ethylene-dependent gravitropism-deficient and yellow-green-like 3	*EGY3*	2.47
AT5G13180.1	*Arabidopsis* NAC domain containing protein 83	*ANAC083*	2.28
AT5G59845.1	gibberellin-regulated family protein	-	2.22
AT3G23150.1	ethylene response 2, receptor	*ETR2*	2.21
AT2G23340.1	AP2 domain-containing transcription factor, putative	-	2.14
AT5G06390.1	fasciclin-like arabinogalactan protein 17	*FLA17*	2.06
AT2G45470.1	fasciclin-like arabinogalactan protein 8	*FLA8*	2.04

Table 3.4: Thirty genes significantly up-regulated at 1 week after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold up-regulation
AT4G02380.2	senescence-associated gene 21	*SAG21*	8.92
AT2G01850.1	xyloglucan:xyloglucosyl transferase	*XTH27*	7.16
AT4G38680.1	glycine rich protein 2	*GRP2*	6.19
AT1G51730.1	RWD domain-containing protein	-	6.16
AT3G15010.2	RNA recognition motif (RRM)-containing protein	-	6.11
AT5G49010.1	Synthetic lethality with DPB11-1 5	*SLD5*	5.62
AT1G55740.1	*Arabidopsis thaliana* seed imbibition 1	*AtSIP1*	5.50
AT3G03330.1	short-chain dehydrogenase/reductase family protein	*SDR*	5.41
AT5G23120.1	protein binding protein	*HCF136*	5.41
AT5G61030.1	glycine-rich RNA-binding protein 3	*GR-RBP3*	5.37
AT5G14920.1	gibberellin-regulated family protein	-	5.10
AT1G10370.1	Early-responsive to dehydration 9	*ERD9*	5.10
AT1G55110.1	*Arabidopsis thaliana* Indeterminate(ID)-Domain 7	*AtIDD7*	5.08
AT3G12580.1	heat shock protein 70	*HSP70*	5.04
AT2G45470.1	fasciclin-like arabinogalactan protein 8	*FLA8*	4.79
AT3G51160.1	murus 1	*MUR1*	4.70
AT5G10270.1	cyclin-dependent kinase C;1	*CDKC1*	4.62
AT5G11260.1	elongated hypocotyl 5	*HY5*	4.24
AT3G19280.1	fucosyl transferase 11	*FUT11*	4.09
AT5G03170.1	fasciclin-like arabinogalactan protein 11	*FLA11*	3.86
AT1G22690.1	gibberellin-responsive protein, putative	-	3.81
AT3G05970.1	long-chain acyl-CoA synthetase 6	*LACS6*	3.46
AT1G62990.1	Knotted-like homeobox of *Arabidopsis thaliana* 7	*KNAT7*	3.45
AT2G38080.1	irregular xylem 12, laccase	*IRX12/LAC4*	3.16
AT2G35840.2	sucrose-phosphatase 1	*SPP1*	2.82
AT3G07230.1	wound-responsive protein-related	-	2.56
AT2G01060.1	myb family transcription factor	-	2.29
AT3G54050.1	fructose-1,6-bisphosphatase, putative	-	2.10
AT4G25810.1	xyloglucan endotransglycosylase 6	*XTR6/XTH23*	2.06
AT5G22290.1	*Arabidopsis* NAC domain containing protein 89	*ANAC089*	2.03

Table 3.5: Thirty genes significantly up-regulated at 2 weeks after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold up-regulation
AT3G09640.2	ascorbate peroxidase 2	*APX2*	13.20
AT5G12030.1	heat shock protein 17.6A	*HSP17.6*	12.98
AT2G29500.1	17.6 kDa class I small heat shock protein B-Cl	*HSP17.6B*	12.64
AT1G17870.1	ethylene-dependent gravitropism-deficient and yellow-green-like 3	*EGY3*	12.14
AT2G46240.1	BCL-2- associated athanogene 6	*BAG6*	12.03
AT5G59720.1	heat shock protein 18.2	*HSP18.2*	11.63
AT3G27330.1	zinc finger (C3HC4-type RING finger) family protein	-	10.53
AT3G46230.1	heat shock Protein	*HSP17.4*	10.36
AT5G40390.1	seed imbibition 1-like	*SIP1*	10.08
AT4G02380.2	senescence-associated gene 21	*SAG21*	10.04
AT5G59720.1	heat shock protein 18.2	*HSP18.2B*	9.77
AT1G12060.1	BCL-2-associated athanogene 5	*BAG5*	9.04
AT2G20900.4	diacylglycerol kinase, putative	*DGK5*	8.56
AT5G14920.1	gibberellin-regulated family protein	-	7.46
AT2G01850.1	xyloglucan:xyloglucosyl transferase	*XTH27*	7.08
AT3G19280.1	fucosyl transferase 11	*FUT11*	6.96
AT5G22290.1	NAC domain containing protein 89	*ANAC089*	6.76
AT5G03170.1	fasciclin-like arabinogalactan protein 11	*FLA11*	6.75
AT3G04720.1	pathogenesis-related 4	*PR4*	5.92
AT5G18370.1	disease resistance protein (TIR-NBS-LRR class), putative	-	5.56
AT4G25810.1	xyloglucan endotransglycosylase 6	*XTR6/XTH23*	5.35
AT5G25190.1	ethylene-responsive element-binding protein, putative	-	2.33
AT2G26980.4	CBL-interacting protein kinase 3	*CIPK3*	2.24
AT2G45140.1	plant VAP homolog 12	*PVA12*	2.15
AT3G22400.1	lipoxygenase	*LOX5*	2.14
AT3G23150.1	ethylene response 2	*ETR2*	2.13
AT1G47128.1	responsive to dehydration 21	*RD21*	2.11
AT1G06040.1	Salt tolerance, transcription factor	*STO*	2.10
AT5G13870.1	endoxyloglucan transferase A4	*EXGT-A4*	2.04
AT4G39350.1	Cellulose synthase A2	*AtCESA2*	2.03

Table 3.6: Thirty genes significantly up-regulated at 6 months after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold up-regulation
AT1G75750.1	GAST1 protein homolog 1	*GASA1*	12.29
AT2G37870.1	lipid transfer protein family protein	*LTP*	10.56
AT4G02380.2	senescence-associated gene 21	*SAG21*	10.52
AT4G17900.1	zinc-binding family protein	-	9.83
AT5G65660.1	hydroxyproline-rich glycoprotein family protein	-	9.08
AT3G12500.1	chitinase	*PR3*	9.03
AT5G13200.1	GRAM domain-containing protein, ABA-responsive protein-related	-	8.63
AT4G14130.1	xyloglucan endotransglycosylase 7	*XTR7*	8.59
AT3G27330.1	zinc finger (C3HC4-type RING finger) family protein	-	8.45
AT5G03170.1	fasciclin-like arabinogalactan protein 11	*FLA11*	8.12
AT5G12030.1	heat shock protein 17.6A	*HSP17.6*	8.10
AT4G25470.1	C-repeat/DRE binding factor 2	*CBF2*	8.07
AT1G73500.1	Map kinase kinase 9	*MKK9*	7.19
AT3G04720.1	pathogensis-related 4	*PR4*	6.87
AT4G23810.1	transcription factor	*WRKY53*	6.82
AT5G04760.1	myb family transcription factor	-	6.80
AT5G60490.1	fasciclin-like arabinogalactan protein 12	*FLA12*	6.06
AT5G44030.1	cellulose synthase A4	*AtCESA4*	5.59
AT1G19180.1	jasmonate-zim-domain protein 1	*JAZ1*	5.51
AT1G33060.2	no apical meristem (NAM) family protein	*ANAC014*	5.32
AT1G05710.3	ethylene-responsive protein, putative	-	4.81
AT5G60360.1	Arabidopsis aleurain-like protease	*SAG2*	4.61
AT2G23290.1	myb domain protein 70	*MYB70*	4.26
AT5G17420.1	cellulose synthase	*AtCESA7*	4.13
AT5G59845.1	gibberellin-regulated family protein	-	3.54
AT5G20900.1	jasmonate-zim-domain protein 12	*JAZ12*	3.37
AT3G14230.3	transcription factor	*RAP2.2*	3.34
AT3G16770.1	transcription factor	*RAP2.3*	2.99
AT1G62990.1	Knotted-like homeobox of *Arabidopsis thaliana* 7	*KNAT7*	2.95
AT2G38080.1	laccase 4	*LAC4*	2.77

Table 3.7: Thirty genes significantly down-regulated at 6 hours after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold down-regulation
AT5G43760.1	3-ketoacyl-CoA synthase 20	*KCS20*	6.37
AT3G53980.2	lipid transfer protein family protein	*LTP*	4.95
AT5G09995.2	expressed Protein	-	4.64
AT2G42770.1	peroxisomal membrane 22 kDa family protein	-	4.37
AT5G64030.1	dehydration-responsive protein-related	-	4.24
AT1G15740.1	leucine-rich repeat family protein	-	4.24
AT5G67620.1	expressed Protein	-	4.00
AT5G24350.1	expressed Protein	-	3.89
AT5G58800.2	quinone reductase family protein	-	3.89
AT4G27670.1	heat shock protein 21	*HSP21*	3.78
AT1G04940.1	Translocon at the inner envelope membrane of chloroplasts 20	*TIC20*	3.78
AT2G41460.1	DNA-(apurinic or apyrimidinic site) lyase	*ARP*	3.78
AT5G56730.1	peptidase M16 family protein	-	3.78
AT2G14750.1	ATP binding / adenylylsulfate kinase	*APK*	3.67
AT3G02340.1	zinc finger (C3HC4-type RING finger) family protein	-	3.57
AT2G18050.1	histone H1-3	*HIS1.3*	3.47
AT2G41835.1	zinc finger (C2H2 type, AN1-like) family protein	-	3.38
AT3G11410.1	protein phosphatase 2CA	*PP2CA*	3.20
AT4G12020.3	transcription factor	*WRKY19*	3.20
AT4G34040.1	zinc finger (C3HC4-type RING finger) family protein	-	3.11
AT5G47530.1	auxin-responsive protein, putative	-	3.03
AT3G13540.1	myb domain protein 5	*MYB5*	2.64
AT1G58370.1	endo-1,4-beta-xylanase	*RXF12/A*	2.30
AT3G12120.2	fatty acid desaturase 2	*FAD2*	2.30
AT1G49430.1	Long-chain acyl-CoA synthetase 2	*LACS2*	2.06
AT4G02290.1	glycosyl hydrolase 9B13	*GH9B13*	2.06
AT5G24760.1	alcohol dehydrogenase, putative	-	2.00
AT1G19300.1	polygalacturonate 4-alpha-galacturonosyltransferase	*PARVUS/GATL1*	2.00
AT3G51680.1	short-chain dehydrogenase/reductase family protein	*SDR*	2.00
AT4G04320.2	malonyl-CoA decarboxylase family protein		2.00

Table 3.8: Thirty genes significantly down-regulated at 24 hours after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold down-regulation
AT3G12120.2	fatty acid desaturase 2	*FAD2*	4.24
AT5G25560.1	zinc finger (C3HC4-type RING finger) family protein	-	3.89
AT3G43810.1	calmodulin 7	*CAM7*	3.78
AT3G63010.1	GA insensitive dwarf1B	*GIDB1*	3.67
AT3G12580.1	heat shock protein 70	*HSP70*	3.67
AT3G44110.1	protein binding protein	*ATJ3*	3.47
AT2G19860.1	hexokinase 2	*HXK2*	3.38
AT1G80670.1	transducin family protein, WD-40 repeat family protein	-	3.29
AT5G18525.1	ATP binding, protein kinase	-	3.29
AT4G01850.2	S-adenosylmethionine synthetase 2	*SAM-2*	3.20
AT5G42190.1	*Arabidopsis* SKP1-like 2, ubiquitin-protein ligase	*ASK2*	3.20
AT3G53260.1	phenylalanine ammonia-lyase	*PAL2*	3.11
AT2G40980.1	Expressed Protein	-	3.11
AT2G36740.1	Transcription factor	*SWC2*	3.11
AT3G55580.1	regulator of chromosome condensation family protein	*RCC1*	3.03
AT1G69545.1	leucine-rich repeat family protein	-	3.03
AT5G43600.1	ureidoglycolate amidohydrolase	*UAH*	3.03
AT3G54700.1	carbohydrate transmembrane transporter	-	3.03
AT4G12020.3	transcription factor	*WRKY19*	2.95
AT3G03150.1	Expressed Protein	-	2.87
AT1G28260.2	Expressed Protein	-	2.87
AT1G50430.1	dwarf 5	*DWF5*	2.87
AT4G36830.1	membrane family protein	*SUR4*	2.30
AT2G20680.1	glycosyl hydrolase family 5 protein, cellulase family protein	-	2.30
AT1G08650.1	phosphoenolpyruvate carboxylase kinase	*PPCK1*	2.24
AT1G65820.1	microsomal glutathione s-transferase, putative	-	2.18
AT1G17020.1	Senescence-related gene 1	*SRG1*	2.18
AT1G19300.1	polygalacturonate 4-alpha-galacturonosyltransferase	*PARVUS/GATL1*	2.12
AT5G20150.1	SPX domain gene 1	*SPX1*	2.06
AT1G11910.1	aspartyl protease family protein	-	2.00

Table 3.9: Thirty genes significantly down-regulated at 1 week after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold down-regulation
AT3G26810.1	Auxin signalling F-BOX 2	*AFB2*	6.37
AT5G43760.1	3-ketoacyl-CoA synthase 20	*KCS20*	6.37
AT4G01850.2	S-adenosylmethionine synthetase 2	*SAM-2*	6.12
AT2G19810.1	zinc finger (CCCH-type) family protein	-	5.47
AT2G35980.1	Yellow-leaf-specific gene 9	*YLS9*	5.11
AT1G19300.1	polygalacturonate 4-alpha-galacturonosyltransferase	*PARVUS/GATL1*	5.11
AT4G22920.1	non-yellowing 1	*NYE1*	4.95
AT2G23320.1	transcription factor	*WRKY15*	4.79
AT4G12980.1	auxin-responsive protein, putative	-	4.79
AT2G29500.1	17.6 kDa class I small heat shock protein	*HSP17.6B*	4.64
AT3G21670.1	nitrate transporter	*NTP3*	4.24
AT3G12120.2	fatty acid desaturase 2	*FAD2*	4.12
AT4G29230.1	*Arabidopsis* NAC domain containing protein 75	*ANAC075*	4.00
AT2G44480.1	beta glucosidase 17	*BGLU17*	4.00
AT1G05170.1	galactosyltransferase family protein	-	3.89
AT5G16530.1	pin-formed 5, auxin:hydrogen symporter	*PIN5*	3.57
AT4G12020.3	transcription factor	*WRKY19*	3.47
AT3G25290.2	auxin-responsive family protein	-	3.29
AT2G30490.1	cinnamate-4-hydroxylase	*C4H*	3.03
AT1G20925.1	auxin efflux carrier family protein	-	2.87
AT3G54820.1	Plasma membrane intrinsic protein 2.5	*PIP2.5*	2.79
AT3G53260.1	phenylalanine ammonia-lyase	*PAL2*	2.79
AT1G17420.1	Lipoxygenase	*LOX3*	2.79
AT1G65060.1	4-coumarate-CoA ligase 3	*4CL3*	2.72
AT5G36880.2	acetyl-CoA synthetase, putative	-	2.72
AT5G54160.1	caffeate O-methyltransferase	*OMT1*	2.72
AT1G63910.1	myb domain protein 103	*MYB103*	2.57
AT1G19300.1	polygalacturonate 4-alpha-galacturonosyltransferase	*PARVUS/GATL1*	2.50
AT5G01990.1	auxin efflux carrier family protein	-	2.30
AT4G12350.1	myb domain protein 42	*MYB42*	2.30

Table 3.10: Thirty genes significantly down-regulated at 2 weeks after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold down-regulation
AT4G25100.2	Fe superoxide dismutase 1	*FSD1*	7.29
AT4G12980.1	auxin-responsive protein, putative	-	6.12
AT3G11410.1	Protein phosphatase 2CA	*PP2CA*	5.67
AT3G09390.1	metallothionein 2A	*MT2A*	5.47
AT2G19810.1	zinc finger (CCCH-type) family protein	-	5.29
AT5G13720.1	expressed protein	-	5.11
AT1G19300.1	polygalacturonate 4-alpha-galacturonosyltransferase	*PARVUS/GATL1*	4.95
AT4G22920.1	non-yellowing 1	*NYE1*	4.50
AT2G02160.1	zinc finger (CCCH-type) family protein	-	4.50
AT3G53260.1	Phenylalanine ammonia-lyase 2	*PAL2*	4.00
AT5G54160.1	O-methyltransferase 1	*OMT1*	3.78
AT3G25290.2	auxin-responsive family protein	-	3.67
AT5G03760.1	mannan synthase	*CSLA9*	3.67
AT5G14700.1	cinnamoyl-CoA reductase-related	*CCR*	3.38
AT2G30490.1	Cinnamate-4-hydroxylase	*C4H*	3.38
AT5G60020.1	laccase 17	*LAC17*	3.38
AT4G22680.1	myb domain protein 85	*MYB85*	3.03
AT2G23910.1	cinnamoyl-CoA reductase-related	*CCR*	3.03
AT5G01580.1	OAS high accumulation 1	*OSH1*	3.03
AT4G30080.1	auxin response factor 16	*ARF16*	2.95
AT1G72770.1	Homology to ABI1	*HAB1*	2.87
AT1G20925.1	auxin efflux carrier family protein	-	2.87
AT5G44790.1	Responsive-to-antagonist 1	*RAN1*	2.72
AT3G23920.1	Beta-amylase 1	*BAM1/BMY7*	2.50
AT1G30330.2	auxin response factor 6	*ARF6*	2.44
AT1G34190.1	*Arabidopsis* NAC domain containing protein 17	*ANAC017*	2.44
AT2G38120.1	auxin resistant 1	*AUX1/WAV5*	2.37
AT4G23980.1	auxin response factor 9	*ARF9*	2.06
AT4G00880.1	auxin-responsive family protein	-	2.06
AT5G03650.1	starch branching enzyme 2.2	*SBE2.2*	2.00

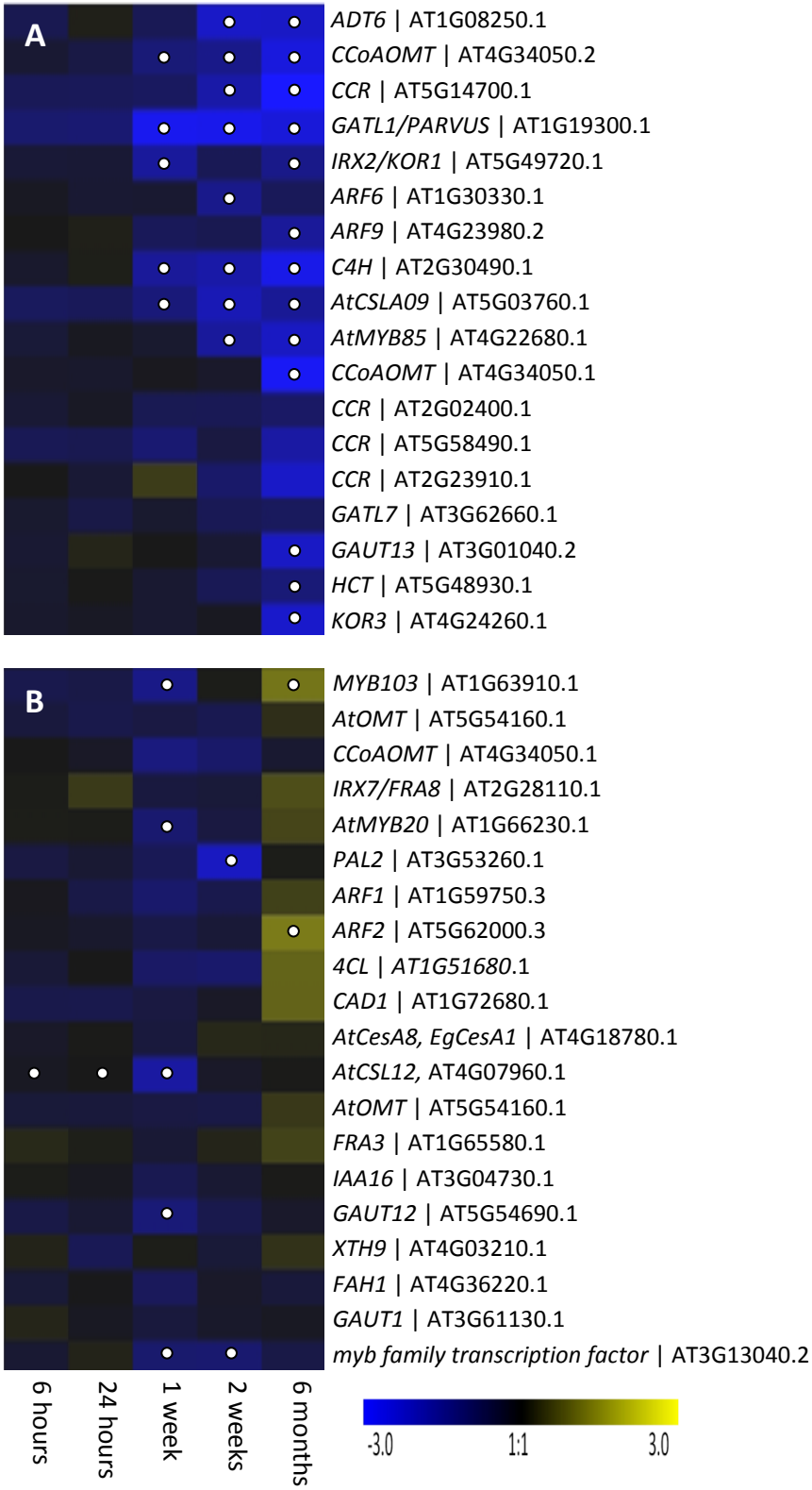
Table 3.11: Thirty genes significantly down-regulated at 6 months after tension wood induction compared to the unbent control.

At number	Description	Abbreviation	Fold down-regulation
AT3G18280.1	lipid transfer protein family protein	*LTP*	11.29
AT4G25100.2	Fe superoxide dismutase 1	*FSD1*	9.29
AT1G60950.1	iron-sulfur cluster binding protein	*FD2*	7.29
AT3G26810.1	auxin signalling F-BOX 2	*AFB2*	7.29
AT3G09320.1	zinc finger (DHHC type) family protein	-	7.29
AT4G12350.1	myb domain protein 42	*MYB42*	6.12
AT5G01930.1	(1-4)-beta-mannan endohydrolase, putative	-	6.12
AT4G12980.1	auxin-responsive protein, putative	-	5.47
AT2G17840.1	early-responsive to dehydration 7	*ERD7*	5.47
AT4G34050.1	caffeoyl-CoA 3-O-methyltransferase, putative	*CCoAOMT*	5.29
AT4G38840.1	auxin-responsive protein, putative	-	5.11
AT5G64570.1	xylan 1,4-beta-xylosidase	*XYL4/BXL4*	5.11
AT2G30490.1	cinnamate-4-hydroxylase	*C4H*	4.95
AT1G19300.1	polygalacturonate 4-alpha-galacturonosyltransferase	*PARVUS/GATL1*	4.50
AT2G38120.1	guxin resistant 1,	*AUX1/WAV5*	4.37
AT1G13250.1	galacturonosyltransferase-like 3	*GATL3*	4.24
AT4G24260.1	*Arabidopsis thaliana* glycosyl hydrolase 9A3	*KOR3*	4.24
AT3G01040.2	galacturonosyltransferase 13	*GAUT13*	4.00
AT5G60020.1	laccase 17	*LAC17*	3.78
AT2G36210.1	auxin-responsive family protein	-	3.67
AT4G36890.1	irregular xylem 14	*IRX14*	3.38
AT5G58490.1	cinnamoyl-CoA reductase family	*CCR*	3.29
AT5G60450.1	auxin response factor 4	*ARF4*	3.11
AT4G23980.2	auxin response factor 9	*ARF9*	3.03
AT1G15950.1	Cinnamoyl CoA reductase 1	*CCR1*	3.03
AT5G03760.1	mannan synthase	*CSLA09*	2.95
AT5G20520.1	Wavy growth 2	*WAV2*	2.79
AT4G34230.1	Cinnamyl alcohol dehydrogenase 5	*CAD5*	2.12
AT1G20925.1	auxin efflux carrier family protein	-	2.06
AT5G45300.1	beta-amylase 2	*BMY2/BAM8*	2.00

3.3.4 Verification of tension wood induction

Previous studies in *Eucalyptus* (Paux et al., 2005, Qiu et al., 2008) and two in *Populus* (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007) were used to select genes expected to be differentially regulated during tension wood development. Genes involved in the biosynthesis of cellulose, hemicellulose and lignin were selected along with genes coding for transcription factors involved in the regulation of hormonal pathways and secondary cell wall biosynthesis. Transcription factors such as *KNAT7* and NAC domain transcription factors hypothesised to be involved in the regulation of secondary cell wall biosynthesis, but for which strict roles have not yet been determined were included (Zhong et al., 2010, Zhong et al., 2007). Mutant phenotypes of *Arabidopsis* plants carrying mutations in GATL and GAUT gene family members present with decreased amounts of secondary cell wall hemicelluloses xylan (Brown et al., 2007, Kong et al., 2009, Persson et al., 2007). Two GATL (*GATL1/PARVUS* and *GATL7*) and two GAUT (*GAUT12* and *GAUT13*) genes were therefore also included in this subset. In total, the 58 genes selected had a wide range of expression levels, ranging in expression from less than 20 ppm for *KNAT7* to 19 234 ppm for *FLA12*. To inspect general trends in expression and for ease of visualisation, these genes were clustered according to their Log_2(tension wood expression ratio to the unbent control) values using the K-means application of TIGR MeV v2.2 (Ben-Dor et al., 1999). The maximum number of clusters was set to four to represent genes either up-, mostly up-, mostly down- or down-regulated (Figure 3.4). This analysis focused on genes involved in later stages of tension wood induction and tension wood formation.

Clusters representing genes down-regulated include lignin and hemicellulose biosynthesis genes and transcription factors involved in the auxin hormone signalling pathway (Figure 3.4 A and B). Some of these genes were significantly down-regulated later in the tension wood time series from 1 week to 6 months after tension wood induction (Figure 3.4 A). One gene *AtCSL12*, encoding a glycosyltransferase, was significantly down-regulated early in the tension wood time series from 6 hours to 1 week after tension wood induction (Figure 3.4 B).



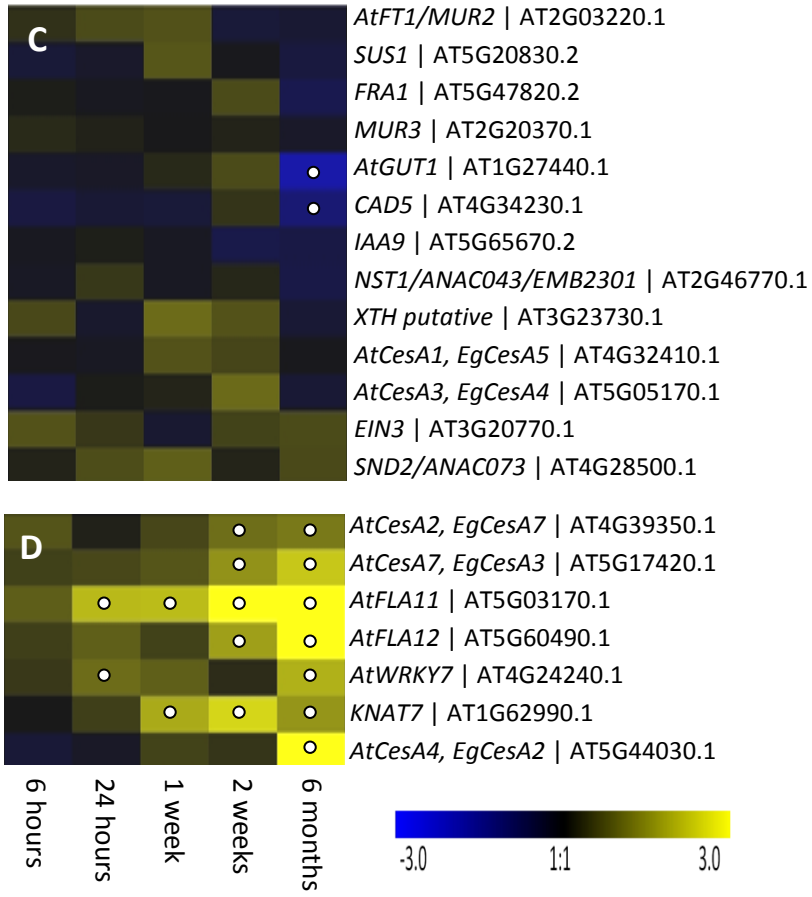


Figure 3.4: (continued from page 124)

K-means clustering of tension wood to unbent control expression ratios of 58 genes involved in cellulose, hemicellulose or lignin biosynthesis. White dots indicate genes significantly differentially regulated at that time point. Genes previously shown to be differentially regulated during tension wood formation were clustered into 4 main groups (A) down-regulated up to 6 months, (B) down-regulated at 1-2 weeks, (C) up-regulated at 1-2 weeks, and (D) up-regulated up to 6 months. The expression levels are Log2-transformed expression ratios of the expression value (ppm) in tension wood time points to the control respectively. Yellow signifies Log_2(TW/C) transcripts that are up-regulated and blue signifies Log_2(TW/C) transcripts that are down-regulated. Genes mostly down-regulated include lignin and xylan biosynthesis genes and transcription factors involved in the auxin hormone signalling pathway (A-B). Genes mostly up-regulated include cellulose and xylan biosynthesis, structural cell wall genes and transcription factors involved in the control of cellulose biosynthesis (C-D).

Figure 3.4: (continued from page 125)

(A) *ADT6*, arogenate dehydratase 6; *CCoAOMT*, caffeoyl-CoA 3-O-methyltransferase; *CCR*, cinnamoyl-CoA reductase-related; *GATL1/PARVUS*, GalUAT-like 1; *IRX2/KOR1*, *Arabidopsis thaliana* glycosyl hydrolase 9A1; *ARF6*, auxin response factor 6; *ARF9*, auxin response factor 9; *C4H*, cinnamate-4-hydroxylase; *AtCSLA09*, mannan synthase/transferase; *AtMYB85*, myb domain protein 85; *CCoAOMT*, caffeoyl-CoA 3-O-methyltransferase; *CCR*, cinnamoyl-CoA reductase family; *CCR*, cinnamoyl-CoA reductase family; *CCR*, cinnamoyl-CoA reductase family; *GATL7*, *AT3G62660.1* GalUAT-like 7; *GAUT13*, galacturonosyltransferase 13; *HCT*, hydroxycinnamoyl-CoA shikimate; *KOR3*, *Arabidopsis thaliana* glycosyl hydrolase 9A3

(B) *MYB103*, myb domain protein *103; AtOMT*, caffeate O-methyltransferase*; CCoAOMT*, caffeoyl-CoA 3-O-methyltransferase; *IRX7/FRA8*, glucuronosyltransferase; *AtMYB20*, myb domain protein 20; *PAL2*, phenylalanine ammonia-lyase 2; *ARF1*, auxin response factor 1; *ARF2*, auxin response factor 2; *4CL*, 4-coumarate:CoA ligase 1; *CAD1*, cinnamyl-alcohol dehydrogenase; *AtCesA8/EgCesA1*, *Eucalyptus grandis* cellulose synthase 1; *AtCSL12*, cellulose synthase like c12; *AtOMT*, O-methyltransferase 1; *FRA3*, phosphatidylinositol phosphatise; *IAA16*, indole-3-acetic acid indicible; *GAUT12*, galacturonosyltransferase 12; *XTH9*, xyloglucan endotransglucosylase 9; *FAH1*, ferulic acid 5-hydroxylase; *GAUT1*, galacturonosyltransferase 1; myb family transcription factor

(C) *AtFT1/MUR2*, fucosyltransferase 1; *SUS1*, sucrose synthase 1; *FRA1*, fragile fiber 1; *MUR3*, murus 3; *AtGUT1*, glucuronosyltransferase 1; *CAD5*, cinnamyl alcohol dehydrogenase 5; *IAA9*, indole-3-acetic acid inducible; *NST1/ANAC043/EMB2301*, embryo defective 2301; *XTH putative*, xyloglucan endotransglucosylase; *AtCesA1/ EgCesA5*, *Eucalyptus grandis* cellulose synthase 5; *AtCesA3/ EgCesA4*, *Eucalyptus grandis* cellulose synthase 4; *EIN3*, ethylene-insensitive 3; *SND2/ANAC073*, *Arabidopsis* NAC domain containing protein 73

(D) *AtCesA2/ EgCesA6*, *Eucalyptus grandis* cellulose synthase 6; *AtCesA7/ EgCesA3*, *Eucalyptus grandis* cellulose synthase 3; *AtFLA11*, fasciclin-like arabinogalactan protein 11; *AtFLA12*, fasciclin-like arabinogalactan protein 12; *AtWRKY7*, transcription factor; *KNAT7*, knotted-like homeobox of *Arabidopsis thaliana* 7; *AtCesA4/ EgCesA2*, *Eucalyptus grandis* cellulose synthase 2.

Genes mostly down-regulated (Figure 3.4 B) and genes mostly up-regulated (Figure 3.4 C) showed interesting opposing expression patterns. These genes which were generally up- or down-regulated in the 6 hours, 24 hours, 1 week and 2 weeks libraries and then reversed expression to be down- or up-regulated respectively in the 6 months library. However, the expression of only two of these genes, *MYB103* and *ARF2*, was significantly up-regulated in the 6 months library (Figure 3.4 B) and one gene, *AtGUT1* showed a significant change to down-regulation in the 6 months library (Figure 3.4 C). Genes up-regulated later during the tension wood time series included cellulose and xyloglucan biosynthesis, structural cell wall genes and transcription factors involved in the control of cellulose biosynthesis which clustered as up-regulated or mostly up-regulated (Figure 3.4 C and D).

Tables 3.12 and 3.13 summarise the expression profiles of the 24 hours and 2 weeks libraries for the genes in Figure 3.5 (line graphs of genes grouped as down-regulated) and the genes in Figure 3.6 (line graphs of genes grouped as up-regulated) and compare the expression profiles obtained in this study to that of previous tension wood expression profiling studies. Only one previous study profiled genes during early tension wood development at 6 hours, 24 hours and 1 week after induction (Paux et al., 2005). The 24 hours tension wood DGE library was used to represent early tension wood induction or formation and for comparison against Paux et al. (2005) (Table 3.12). The other tension wood studies which were used for comparison investigated tension wood formation at 19 days (Nishikubo et al., 2007), 3 weeks (Andersson-Gunnerås et al., 2006) and in a stable tension wood forming branch (Qiu et al., 2008) and were considered as representatives of late tension wood formation. As the sequencing of the 3 weeks tension wood library failed, the 2 weeks library was used as representative of late tension wood induction and for comparison to the other tension wood expression profiling studies (Table 3.13).

Of the 10 genes shown in Figure 3.5, nine were significantly down-regulated in at least one DGE library. Three lignin biosynthetic genes, *PAL2*, *CCR1* and *CCoAOMT* showed a decrease in expression (Figure 3.5A). Down-regulation of *PAL2* was supported by literature (Andersson-Gunnerås et al., 2006, Paux et al., 2005) and the 4-fold-decreased expression level in the 2 weeks DGE library was significant (Figure 3.5 A).

Table 3.12: Comparison of the DGE and RT-qPCR results in this study to the response of selected genes 24 hours after tension wood induction as observed in a previous *Eucalyptus* tension wood studies. Arrows indicate an increase or a decrease in expression respectively. (-) = gene not profiled.

AT Number	Gene	Pathway/Function	Reference	DGE	RT-qPCR
AT5G17420.1	*EgCesA3*	Cellulose biosynthesis (secondary cell wall)	-	↑	↑
AT4G32410.1	*EgCesA5*	Cellulose biosynthesis (primary cell wall)	-	↓	↓
AT3G43190.1	*SUS1*	Carbohydrate metabolism	-	↑	↑
AT5G49720.1	*KOR1*	Possible involvement in cellulose biosynthesis	-	↓	↓
AT4G25810.1	*XTH23*	Cell wall modification of hemicelluloses.	-	↑	↑
AT5G03170.1	*FLA11*	Cell wall structural protein	-	↑	-
AT5G60490.1	*FLA12*	Cell wall structural protein	-	↑	-
AT5G03760.1	*AtCSLA09*	Hemicellulose biosynthesis	-	↓	-
AT5G54690.1	*GAUT12*	Hemicellulose biosynthesis (xylan)	- (a)	↓	↑
AT1G19300.1	*GATL1*	Hemicellulose biosynthesis (xylan)	- (a)	↓	-
AT4G34050.1	*CCoAOMT*	Lignin biosynthesis	↑ (b)	↓	↑
AT1G15950.1	*CCR1*	Lignin biosynthesis	↑ (b)	↑	-
AT3G53260.1	*PAL2*	Lignin biosynthesis	↓ (b)	↓	-
AT3G20770.1	*EIN3*	Ethylene hormone pathway related transcription factor	-	↑	-
AT1G30330.1	*ARF6*	Auxin hormone pathway related transcription factor	↓ (b)	↓	-
AT4G24240.1	*WRKY7*	Transcription factor. Stress response	-	↑	-
AT1G63910.1	*MYB103*	Transcription factor	↑ (b)	↓	-
AT1G62990.1	*KNAT7*	Transcription factor	-	↑	↑

[a] Mutant phenotypes show decreased xylan content (Brown et al., 2007).

[b] A *Eucalyptus* tension wood induction study profiling genes at 6 hours, 24 hours and 1 week (Paux et al., 2005).

Table 3.13: Comparison of the DGE and RT-qPCR results in this study to the response of selected genes 2 weeks after *Eucalyptus* tension wood induction as observed in literature investigating tension wood formation later than 1 week after induction. Arrows indicate an increase or a decrease in expression respectively. (-) = gene not profiled.

AT number	Gene	Pathway/Function	Reference	DGE	RT-qPCR
AT5G17420.1	*EgCesA3*	Cellulose biosynthesis	No response (a)	↑	↑
AT4G32410.1	*EgCesA5*	Cellulose biosynthesis	↑ (a)	↑	↑
AT3G43190.1	*SUS1*	Carbohydrate metabolism	↑ (a)	↓	↑
AT5G49720.1	*KOR1*	Possible involvement in cellulose biosynthesis	↓ (a)	↓	↑
AT4G25810.1	*XTH23*	Cell wall modification of hemicelluloses.	↑ (a), (b)	↑	↑
AT5G03170.1	*FLA11*	Cell wall structural protein	↑ (a), (c)	↑	-
AT5G60490.1	*FLA12*	Cell wall structural protein	↑ (a), (c)	↑	-
AT5G03760.1	*AtCSLA09*	Hemicellulose biosynthesis	↓ (a)	↓	-
AT5G54690.1	*GAUT12*	Hemicellulose biosynthesis (xylan)	- (d)	↓	↑
AT1G19300.1	*GATL1*	Hemicellulose biosynthesis (xylan)	- (d)	↓	-
AT4G34050.1	*CCoAOMT*	Lignin biosynthesis	↓ (a)	↓	↓
AT1G15950.1	*CCR1*	Lignin biosynthesis	↓ (a)	↓	-
AT3G53260.1	*PAL2*	Lignin biosynthesis	↓ (a)	↓	-
AT3G20770.1	*EIN3*	Ethylene hormone pathway related transcription factor	↑ (a)	↑	-
AT1G30330.1	*ARF6*	Auxin hormone pathway related transcription factor	↓ (a)	↓	-
AT4G24240.1	*WRKY7*	Transcription factor. Stress response	↑ (a)	↑	-
AT1G63910.1	*MYB103*	Transcription factor	↑ (a)	↑	-
AT1G62990.1	*KNAT7*	Transcription factor	-	↑	↑

[a] A *Populus* study profiling genes expressed in tension wood at 3 weeks after induction using microarray (Andersson-Gunnerås et al., 2006).

[b] A macroarray study investigating the activity of *XTH23* genes in *Populus* trees induced to form tension wood for 19 days (Nishikubo et al., 2007).

[c] A microarray based investigation of branch tension wood in *Eucalyptus* (Qiu et al., 2008*)*.

[d] Mutant phenotypes show decreased xylan content (Brown et al., 2007).

The other two genes involved in monolignol biosynthesis, *CCoAOMT* and *CCR1* both showed strong decreases in the 6 months library of almost 5 fold and 3 fold, respectively (Figure 3.5 A). The increased expression of *CCR1* at 24 hours (not significant) and decreased expression of *CCR1* and *CCoAOMT* at 2 weeks was supported by previous studies (Tables 3.12 and 3.13, Andersson-Gunnerås et al., 2006, Paux et al., 2005). The down-regulated expression of *CCoAOMT* at 24 hours was not supported by Paux et al. (2005).

IRX8 and PARVUS3 mutants show large decreases in hemicellulose xylan content (Brown et al., 2007). The mutagenesis of two genes, *GAUT12* and *GATL1*, resulted in these phenotypes, respectively. Both of these transcripts showed an according decrease during *Eucalyptus* tension wood formation, however it was the significant decrease in expression up to 5 fold of *GATL1* in the 1 week, 2 weeks and 6 months libraries that best reflected the decreased xylan content of tension wood (Figure 3.5 B). A significant decrease in the expression of mannan synthase (*AtCSLA09*) in the 1 week, 2 weeks and 6 months libraries (Figure 3.5 B) also reflected a decrease in hemicellulose and the redirection of carbon metabolism towards cellulose synthesis. This was supported by tension wood expression profiling in poplar (Table 3.13, Andersson-Gunnerås et al., 2006).

KOR1 showed a decrease in all of the tension wood induction libraries, and a significant down-regulation of approximately 3 fold in the 1 week and 6 months libraries (Figure 3.5 C). In contrast to what was found by Andersson-Gunneras et al. (2006), the expression of sucrose synthase *SUS1* was not significantly differentially regulated (Figure 3.5 C). The down-regulation of the transcription factor *ARF6* (Figure 3.5 D), an auxin responsive protein was supported by literature (Andersson-Gunnerås et al., 2006, Paux et al., 2005) and was found to be significant in 2 weeks DGE library. *MYB103*, hypothesised to be involved in the regulation of secondary cell wall biosynthesis showed a complex expression pattern (Figure 3.5 D). It was significantly down-regulated in the 1 week tension wood library and significantly up-regulated in the 6 months tension wood library. Only the up-regulation of *MYB103* was supported by literature (Andersson-Gunnerås et al., 2006, Paux et al., 2005).

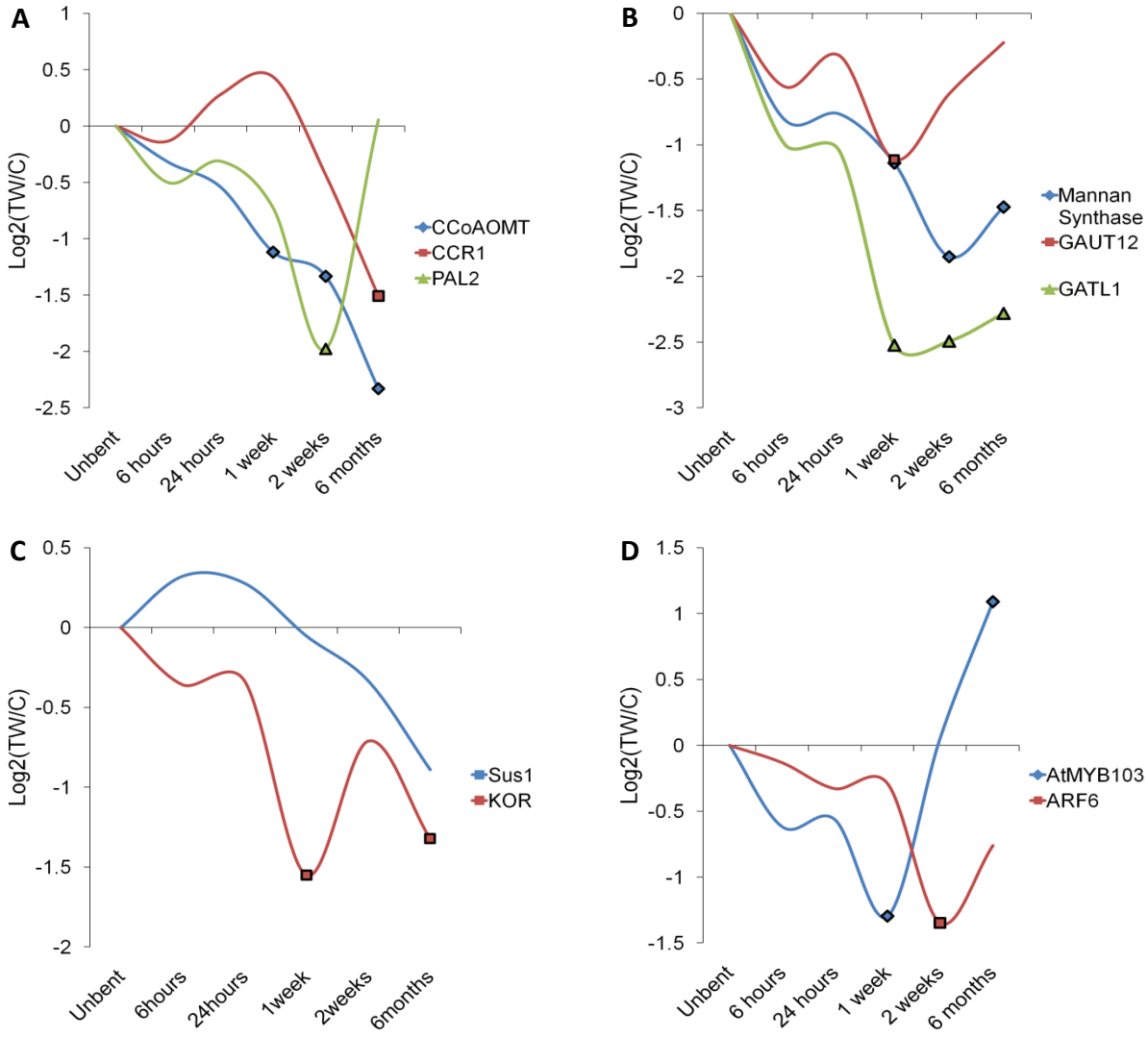


Figure 3.5: Expression profiles of genes mostly down-regulated during tension wood formation in *Eucalyptus grandis x Eucalyptus urophylla* hybrid trees. The expression profiles are represented as Log_2 transformed expression ratios of the expression values (ppm) tension wood libraries to the unbent control (TW/C). Symbols of the time points with significant differential expression are shown (symbols at non-significant time points were omitted). (A) Lignin biosynthesis genes caffeoyl-CoA 3-O-methyltransferase (*CCoAOMT*), cinnamoyl-CoA reductase 1 (*CCR1*) and phenylalanine ammonia-lyase 2 (*PAL2*) (B) Hemicellulose genes mannan synthase (*AtCSLA09*), galacturonosyltransferase 12 (*GAUT12*) and GalUAT-like 1 (*GATL1*) (C) sucrose synthase 1 (*SUS1*) and korriganA (*KOR1*) (D) myb domain transcription factor (*MYB103*) and auxin response factor 6 (*ARF6*).

Of the 13 genes shown in Figure 3.6 nine were significantly up-regulated in at least one library. *EgCesA1* has previously been shown to be up-regulated in *Eucalyptus* 1 week after tension wood induction (Paux et al., 2005). Poplar tension wood induction expression profiling studies found the ortholog *PtrCesA1* to be up-regulated too (Andersson-Gunnerås et al., 2006, Bhandari et al., 2006, Djerbi et al., 2004). Although highly expressed, the expression of *Eucalyptus EgCesA1* was not significantly altered and failed to reach a change in expression of even 1 fold. *EgCesA2*, like *EgCesA1* showed little change in expression except at 6 months after induction where the expression was up-regulated to 5.5 fold. *EgCesA3* was previously thought to show no response to tension wood induction (Qiu et al., 2008). In this study, it was the only secondary cell wall *EgCesA* that showed a significant increase in expression in more than one DGE library and reached a 4 fold up-regulation 6 months after tension wood induction (Figure 3.6 A). The primary cell wall cellulose synthase genes *EgCesA5-7* showed no clear pattern of expression and expression changes were weak throughout most of the tension wood series (Figure 3.6 B). *EgCesA6* and *EgCesA7* were significantly up-regulated 2 fold in the 2 weeks and 6 months DGE libraries respectively. Even though the differential expression of *EgCesA5* was not significant, the direction of change for the 2 weeks library was supported by Andersson-Gunneras et al. (2006).

With the exception of the 6 months library, the structural cell wall *XTH23* enzyme was significantly up-regulated in every tension wood library (Figure 3.6 C). Up-regulation of *XTH* enzymes during tension wood formation was also observed in previous *Eucalyptus* and *Populus* studies (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007). The FLA transcripts *FLA11* and *FLA12* were found to be up-regulated throughout the tension wood time series (Figure 3.6 C). *FLA11* was significantly up-regulated in four of the five libraries, with the exception of the 6 hours library. *FLA12* was significantly up-regulated in the 2 weeks and 6 months libraries. These observations were supported by two previous tension wood studies, one in eucalypts (Qiu et al., 2008) and one in poplar (Andersson-Gunnerås et al., 2006).

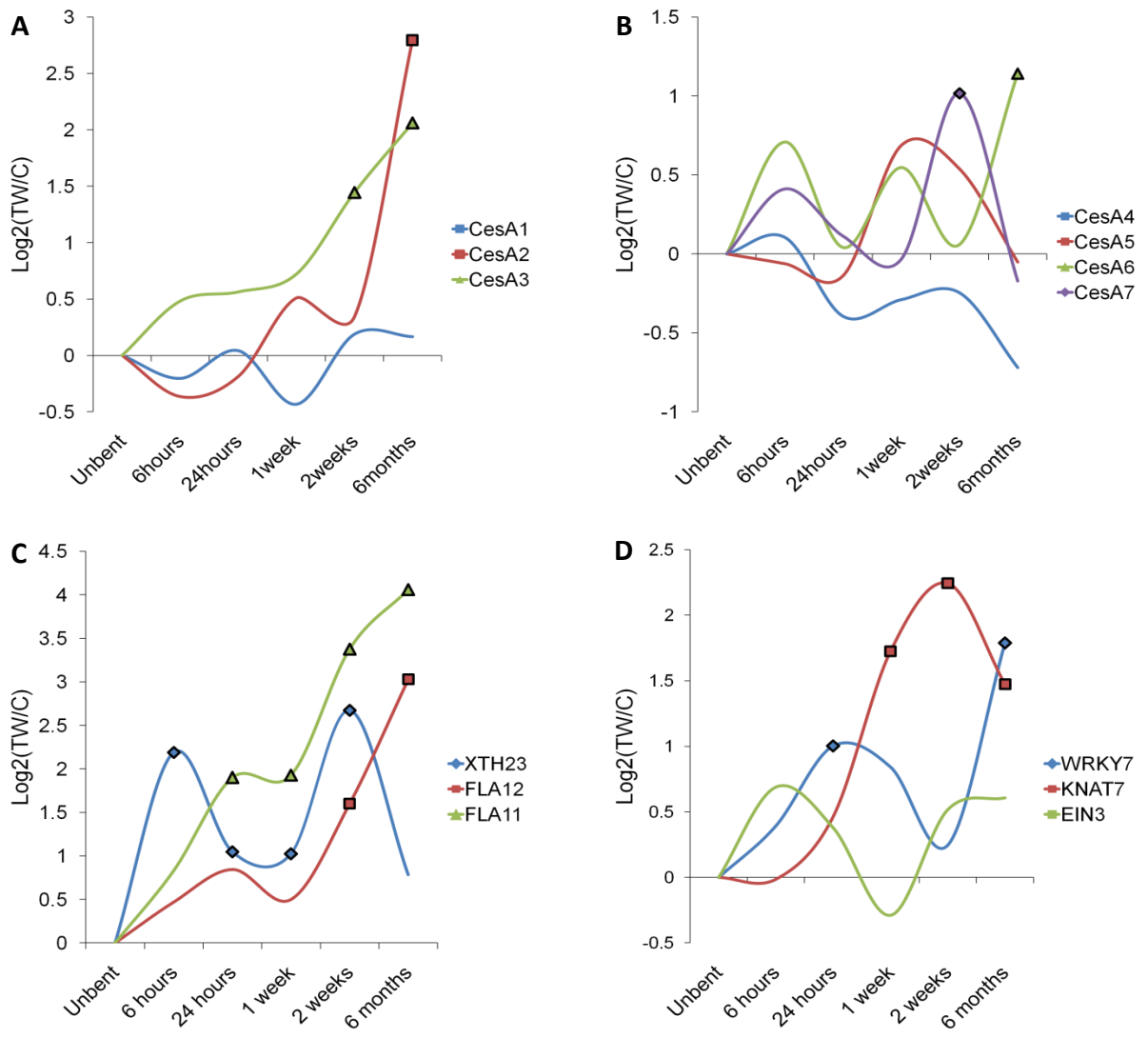


Figure 3.6: Expression profiles genes mostly up-regulated and *EgCesA4* which is down-regulated during tension wood formation in *Eucalyptus grandis x Eucalyptus urophylla* hybrid trees. The expression profiles are represented as Log_2 transformed expression ratios of the expression values (ppm) tension wood libraries to the unbent control (TW/C). Symbols of the time points with significant differential expression are shown (symbols at non-significant time points were omitted). (A) secondary cell wall cellulose synthase genes *EgCesA1*, *EgCesA2*, *EgCesA3;* (B) primary cell wall cellulose synthase genes *EgCesA4, EgCesA5, EgCesA6* and *EgCesA7* (C) xyloglucan endotransglucosylase (*XTH23*) and fasciclin-like arabinogalactan (*FLA11* and *FLA12*) (D) stress related transcription factor (*AtWRKY7*), secondary cell wall biosynthesis transcription factor knotted-like homeobox of *Arabidopsis thaliana* 7 (*KNAT7*) and ethylene hormone related transcription factor ethylene-insensitive 3 (*EIN3*).

WRKY7, a stress response transcription factor was up-regulated during tension wood formation and significantly up-regulated in the 24 hours and 6 months libraries (Figure 3.6 D). The up-regulation of this transcription factor was also observed in the tension wood study by Andersson-Gunneras et al. (2006). The ethylene induced transcription factor, *EIN3* showed weak up-regulation which was not significant (Figure 3.6 D). *KNAT7*, a transcription factor involved in the regulation of secondary cell wall biosynthesis (Zhong et al., 2008) was significantly up-regulated in the 1 week, 2 weeks and 6 months libraries (Figure 3.6 D).

3.3.5 RT-qPCR validation

RT-qPCR was performed on eight selected genes to validate the DGE expression profiles. Samples profiled were 24 hours to represent early tension wood induction, 2 weeks to represent late tension wood induction and the unbent control. These time points were also selected to supplement the literature verification of the DGE profiles. The genes profiled include primary cell wall cellulose synthase 5 (*EgCesA5*; AT4G32410.1), secondary cell wall cellulose synthase 3 (*EgCesA3*; AT5G17420.1), korrigan (*KOR1*; AT5G49720.1), sucrose synthase (*SUS1*; AT3G43190.1), xyloglucan endotranscglucosylase (*XTH23*; AT4G25810.1), galacturonosyltransferase (*GAUT12*; AT5G54690.1), a knotted-like homeobox of *Arabidopsis thaliana* transcription factor (*KNAT7*; AT1G62990.1) and caffeoyl-CoA 3-O-Methyltransferase (*CCoAOMT*; AT4G34050.1) (Table 2.1). The correlation between the Log_2(TW/C) expression ratios of DGE and RT-qPCR for the 24 hours time point was 0.65 (Figure 3.7 A) and the correlation between the Log_2(TW/C) expression values of DGE and RT-qPCR for the 2 weeks time point was 0.88 (Figure 3.7 B). Profiled by RT-qPCR, five of the eight genes, *EgCesA3*, *EgCesA5*, *KOR1*, *XTH23* and *KNAT7* showed the same direction of change in gene expression in response to tension wood induction at 24 hours as DGE (Table 3.12). In contrast to down-regulation as profiled by DGE, up-regulation at 24 hours of *CCoAOMT* is supported by RT-qPCR and by Paux et al. (2005). For the 2 weeks time point, five genes which showed the same change in expression in response to tension wood induction included *EgCesA3*, *EgCesA5*,

XTH23, *KNAT7* and *CCoAOMT* (Table 3.13). The change in expression observed with DGE at 2 weeks after tension induction for the three genes *EgCesA5*, *XTH23* and *CCoAOMT* is supported by RT-qPCR and previous tension wood studies (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007). Lastly the down-regulation of *KOR1* is supported by DGE and Andersson-Gunneras et al. (2006), but contrasted by RT-qPCR.

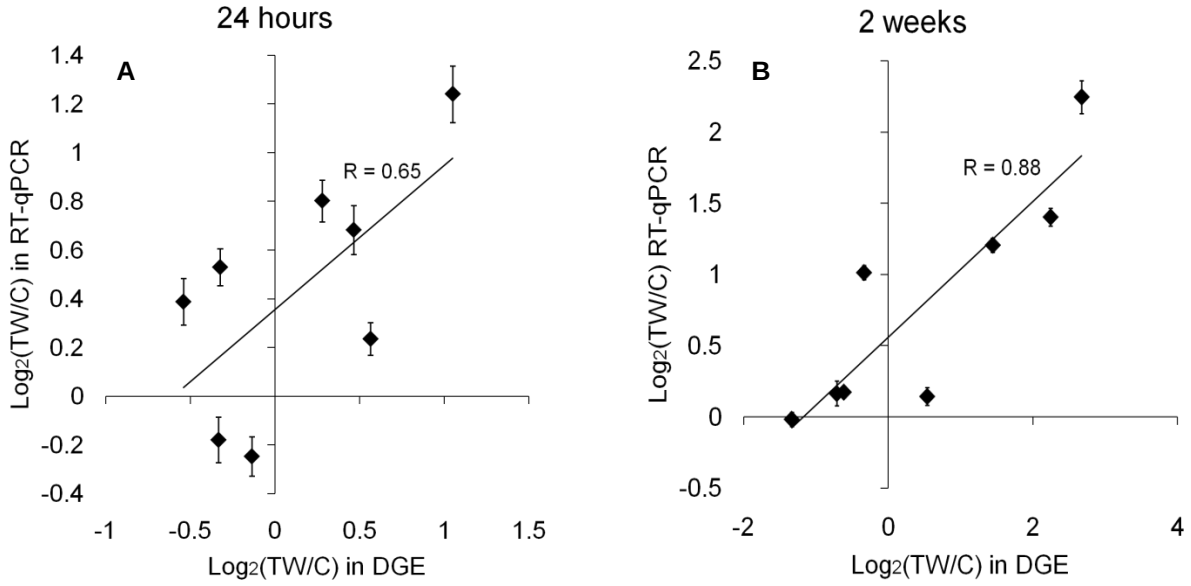


Figure 3.7: Correlation between the Log_2(tension wood to control ratio) values of RT-qPCR and DGE expression profiles of genes affected by tension wood formation. R represents the Pearson correlation coefficient and the error bars represent the technical standard error for RT-qPCR, n=3 (A) Correlation of the gene expression values at 24 hours after tension wood induction (B) Correlation of the gene expression values at 2 weeks after tension wood induction.

3.4 Discussion

The overall aim of this study was to assess the global change in gene expression in *Eucalyptus* tension wood forming tissues in early and late time points by bending *Eucalyptus grandis x Eucalyptus urophylla* F1 hybrid trees (ramets of a single clonal genotype) for 6 hours, 24 hours, 1 week, 2 weeks, 3 weeks and 6 months. Six DGE libraries were successfully constructed using Illumina DGE expression profiling kits and sequenced with the Genome Analyser (GAII). These libraries included a library from each time point and an unbent control pooled from three biological replicates. Sequencing of the 3 weeks tension wood library was unsuccessful and was therefore excluded from further analyses. The induction and formation of tension wood was confirmed by comparing the expression profiles of genes previously profiled in *Eucalyptus* (Paux et al., 2005, Qiu et al., 2008) and *Populus* (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007) tension wood formation and by validation with RT-qPCR. Significantly differentially regulated genes were identified (Audic and Claverie, 1997) and a pilot investigation of these genes revealed that early tension wood induction and formation are marked by a strong stress response to bending followed by a response in secondary cell wall biosynthetic genes.

3.4.1 Validation of tension wood induction in *Eucalyptus grandis x Eucalyptus urophylla* trees

Genes previously profiled during tension wood induction studies were selected to confirm tension wood-specific changes in gene expression. In general, lignin and xylan genes were down-regulated (Figure 3.4 A and B), while cellulose biosynthesis and xyloglucan hemicellulose biosynthesis genes were up-regulated (Figure 3.4 C and D). This reflects the chemical compositional changes observed during tension wood formation; an increase in crystalline cellulose accompanied by a decrease in xylan content, but an increase in the content of the hemicellulose xyloglucan and a decrease in lignin content (Joseleau et al., 2004, Nishikubo et al., 2007).

Down-regulation of xylan biosynthesis pathways are demonstrated by the decrease of *GAUT12* and *GATL1*. The *GAUT12* and *GATL1* genes have not previously been profiled during tension wood formation, but mutant plants of these genes exhibit a decreased amount of xylan (Brown et al., 2007) which is also a characteristic of tension wood (Bowling and Vaughn, 2008, Pilate et al., 2004b). Even though the xylan content of tension wood is decreased, other hemicelluloses such as xyloglucan (Mellerowicz et al., 2008), rhamnogalacturonan (Bowling and Vaughn, 2008) and arabinogalactan (Bowling and Vaughn, 2008, Lafarguette et al., 2004) have been shown to be increased refuting the long standing belief that the G-layer is composed of pure crystalline cellulose (Norberg and Meier, 1966). In accord with this, transcripts of the xyloglucan endotransglucosylases (*XTH/XET* genes) have been shown to be up-regulated during tension wood (Andersson-Gunneras et al., 2006). In this study, *XTH23* showed strong significant up-regulation during early and late tension wood induction (Figure 3.6 C). *XTH* enzymes cut the xyloglucan backbone so that a new glycosidic bond can be formed with another xyloglucan chain. Xyloglucan is hypothesised to play an important structural role in tension wood by linking the G-layer to the S_2-layer. Linking the G-layer to the rest of the fibre cell wall is essential for the transfer of tensile stress resulting in the corrective bending moment (Arend, 2008, Baba et al., 2009, Mellerowicz et al., 2008). As a result, the expression of *XTH* transcripts remains up-regulated during the production of mature tension wood formation (Nishikubo et al., 2007).

As mentioned above, another protein found in high quantities in tension wood is arabinogalactan protein (Bowling and Vaughn, 2008, Lafarguette et al., 2004). Fasciclin-like arabinogalactan proteins (*FLAs*) are among the most highly expressed transcripts in tension wood (Andersson-Gunneråas et al., 2006, Lafarguette et al., 2004). The *Eucalyptus* orthologs of *AtFLA11* and *AtFLA12* were also the most highly expressed transcripts in this study and highly up-regulated during late tension wood formation (Figure 3.6C). These subgroup A *FLA* genes (Johnson et al., 2003) were significantly up-regulated in the 6 months library to 6 and 8 fold respectively. FLA proteins are thought to aid in the production of a hydrophilic environment within the tension wood tissue needed for optimal *XTH* enzyme functioning (Mellerowicz et al., 2008). Highly expressed *FLA*

transcripts are consistent with previous observations of a large increase of *Eucalyptus FLA* gene expression and were linked to a difference in wood fiber properties and in particular to a decrease in cellulose microfibril angle (Lafarguette et al., 2004, Qiu et al., 2008). *Arabidopsis* T-DNA *FLA* knockout studies yielded plants with altered biomechanics, an increase in microfibril angle and a decrease in cellulose content (MacMillan et al., 2010). The *FLA* knockout lines exhibited characteristics opposite to that of the tension wood. MacMillan et al. (2010) suggested a direct role for FLAs in stem biomechanics including the strength of a tree stem through involvement in cellulose biosynthesis and cellulose deposition.

Genes in both the early and later stages of lignin biosynthesis have been observed to be decreased during poplar and eucalypt tension wood formation (Andersson-Gunnerås et al., 2006, Paux et al., 2005). The early phenylpropanoid pathway phenylalanine-lyase gene *PttPAL1* is decreased significantly during poplar tension wood formation (Andersson-Gunnerås et al., 2006). Genes involved later in lignin biosynthesis specifically from the monolignol specific branch were observed to be consistently decreased (Paux et al., 2005). The same held true for tension wood formation in this study (Figure 3.5 A). Even though the down-regulation of *CCoAOMT* at 24 hours was not supported by literature (Paux et al., 2005), it should be noted that *CCoAOMT* was expressed at low levels at this time point and that the significant down-regulation of *CCoAOMT* at 1 week was supported by down-regulation at one week observed in the study of Paux et al. (2005).

The increased cellulose content of tension wood was reflected by the significant up-regulation of *EgCesA2* and *EgCesA3*, two secondary cell wall related *CesA* genes. However, the expression patterns overall still did not clarify previous observations with regards to the response of the secondary cell wall *CesA* genes during tension wood formation (Figure 3.6 A). In this study, *EgCesA3* showed a strong response, *EgCesA1* showed no significant differential regulation, while the poplar orthologs for these genes showed no response and up-regulation respectively (Andersson-Gunnerås et al., 2006). This may be a species specific response, and may support the hypothesis that different CesA isoforms are utilised during tension wood formation (Andersson-Gunnerås et al., 2006, Qiu et al., 2008). Although the expression of *SUS1* was up-regulated in the 6 hours and 24 hours

libraries and down-regulated in the remaining libraries it was not significantly differentially expressed during tension wood formation (Figure 3.5 C). An increase in *Susy* transcription has been linked to the channelling of carbon towards cellulose synthesis, however an increase in *Susy* expression may not be necessary for an increase in cellulose content as is demonstrated by the activation of alternate carbon channelling pathways during poplar tension wood formation (Andersson-Gunnerås et al., 2006). Overall, the weak expression of *EgCesA5-7* point to their involvement in primary cell wall synthesis and not secondary cell wall synthesis.

Three genes, *KOR1*, and the transcription factors *KNAT7* and *MYB103* putatively involved in cellulose formation, were differentially expressed during tension wood formation. *KOR1*, a korrigan endoglucanase hypothesised to play an important role in cellulose biosynthesis (Bhandari et al., 2006), showed a decrease in all of the tension wood induction libraries, reaching a 3 fold decrease one week after induction (Figure 3.5 C). The poplar ortholog too was down-regulated during poplar tension wood formation (Andersson-Gunnerås et al., 2006). Korrigan is required for normal xylem development and RNAi-suppressed poplar lines deposit a more highly crystalline cellulose in the secondary cell wall (Maloney and Mansfield, 2010). The cellulose deposited in tension wood is more crystalline than the cellulose deposited during normal wood formation (Nishikubo et al., 2007). Secondary cell wall NAC proteins activate a cascade of downstream transcription factors, which in turn, lead to the activation of secondary cell wall biosynthetic genes. Among these are *KNAT7* and *MYB103* (Zhong et al., 2006, Zhong et al., 2008). Studies still have to confirm the specific roles of these transcription factors during secondary cell wall development and their respective roles in the lignin and cellulose biosynthetic pathways. *KNAT7* was significantly up-regulated confirming its important role in secondary cell wall formation and perhaps pointing towards positive regulation of cellulose biosynthesis (Figure 3.7D). The expression pattern of *MYB103* was less clear as it was significantly down-regulated at one week after tension wood induction, but also significantly up-regulated at 6 months. However, the importance of this gene in cellulose biosynthesis is supported by its responsiveness in this study and by its up-regulation in two other tension wood studies (Andersson-Gunnerås et al., 2006, Paux et al., 2005).

ARF6 and *EIN3* are two genes representative of the auxin and ethylene hormone signalling pathways respectively. Down-regulation of the auxin hormone levels is indicative of cambial differentiation (Benjamins and Scheres, 2008) which is conferred through down-regulation of *ARF6.* Ethylene acts antagonistically to auxin to positively stimulate cambial differentiation (Love et al., 2009) and the induction of ethylene has previously been associated with tension wood formation (Andersson-Gunnerås et al., 2006, Du and Yamamoto, 2003). However ethylene has been thought to rather play an indirect role in tension wood formation (Jiang et al., 2009).

WRKY transcription factors are known to mediate both biotic and abiotic stress responses (Jiang and Deyholos, 2009). The up-regulation of the *WRKY7* transcription factor coincided with the expected response to mechanical stress that bent eucalypt trees experience (Table 3.12 and 3.13). Evidence of down-regulation of auxin signalling through *ARF6* coincides well with observations that the over-expression of *OsWRKY* transcription factors in *Arabidopsis* interfered with the normal auxin signalling pathway and induced an enhanced gravitropism response (Song et al., 2010).

In summary, for early tension wood formation (24 hours after induction) three of the five genes showed similar expression patterns compared to Paux et al. (2005, Table 3.12) and for late tension wood induction 13 of the 15 genes selected from the 2 weeks library showed similar expression patterns to those reported in literature (Andersson-Gunnerås et al., 2006, Nishikubo et al., 2007, Qiu et al., 2008, Table 3.13). Verification of the DGE profiles by RT-qPCR yielded a correlation of 0.65 for the 24 hours time point (Figure 3.7 A) and 0.88 for the 2 weeks time point (Figure 3.7 B). Therefore the expression profiles of the subset of genes sampled from literature and discussed above most likely reflect the transcriptional and cell wall chemistry changes that occur during *Eucalyptus* and *Populus* tension wood formation.

3.4.2 Early tension wood induction and formation

Very early responses to stem bending involve mechanisms such as mechanosensing and transduction of the perceived stimuli into two types of responses; thigmomorphogenic (induced by a short

stimulus, Jaffe, 1980) or gravitropic (induced by longer lasting stimuli, Hejnowicz, 1997). In a natural environment, plants are continuously bombarded with dynamic abiotic stimuli and the distinction between the mechanosensing and gravitropic processes is dim (Coutand, 2010). Both of these processes are mediated by the ethylene signalling pathway and may be induced by applied ethylene (Braam, 2005, Kagaya and Hattori, 2009). Thus a chronic stimulus, such as continued bending, will result in strong activation of cell signalling cascades activating mechanosensing, gravitropic and hormonal signalling responses during the early phases of tension wood induction, which later results in transcriptional activation of secondary cell wall related genes for a tension wood formation response (Love et al., 2009).

Many of the genes significantly differentially regulated during the early tension wood response (6 hours and 24 hours) of *Eucalyptus grandis x Eucalyptus urophylla* hybrid trees were transcription factors (Tables 3.2 and 3.3). The transcription factor *RAV1*, specifically induced by mechanical stress (Kagaya and Hattori, 2009) was up-regulated 4.3 fold after 6 hours. *RAV1* has also been shown to be up-regulated in *Arabidopsis* mutant plants that showed increased abiotic stress tolerance (Osakabe et al., 2009). Other transcription factors generally implicated in abiotic stress signalling was also up-regulated such as the cold stress and wound inducible *CZF1* (Table 3.3, Matsukura et al., 2010) which is co-regulated with the transcription factor *CBF2* (Table 3.2, 3.3, Sreenivasulu et al., 2007), and the transcription factor *BT1* previously shown to be up-regulated during salt stress (Table 3.3, Fujita et al., 2007, Mandadi et al., 2009). Abscisic acid (ABA) is one of the major hormones involved in abiotic stress signalling in plants (Takahashi et al., 2004), and transcription factors such as the MYBs (Abe et al., 2003, Yanhui et al., 2006), some NACs (Fujita et al., 2004, Gao et al., 2010) and AP2/EREBP (Dietz et al., 2010), which were up-regulated in these early tension wood libraries, are also ABA response mediators. Two other plant hormones, ethylene and gibberellic acid (GA) are known to be associated with the up-regulation of transcripts in tension wood formation (Andersson-Gunnerås et al., 2006, Funada et al., 2008, Jiang et al., 2008). *HY5* transcripts, involved in ABA signalling and GA biosynthesis were up-regulated 4 fold after 24 hours (Weller et al., 2009). In accordance with previous observations, ethylene related transcripts, *ETR2*,

EIN3, *ACS06*, and gibberellin-regulated family protein transcripts were up-regulated after 6 and 24 hours. The stress related map kinase kinase 9 (*MKK9*), which functions as a signalling molecule in the ethylene pathway and in ethylene biosynthesis (Hahn and Harter, 2009, Xu et al., 2008, Yoo et al., 2008) was highly up-regulated (9 fold) in the 6 hours library. The up-regulation of transcripts with multiple roles in stress response may result in the cross-talk of biotic and abiotic stress response pathways (Glombitza et al., 2004). Among others, the biotic stress pathway leads to the up-regulation of salicylic acid, jasmonic acid (JA) and ethylene hormone signalling pathways and the up-regulation of reactive oxygen species (ROS) transcripts (Mittler et al., 2006, Thatcher et al., 2005). A number of the transcripts up-regulated during the early tension wood formation libraries such as the *DC1* domain containing protein (Table 3.2) and *DND1* transcripts (Table 3.3) are biotic stress induced transcripts (Govrin and Levine, 2000, Mohr and Cahill, 2007) and the ROS defence transcript, peroxidase 2 (*APX2*) also induced by abiotic stresses, were up-regulated by 6 fold (Mittler et al., 2006). Wound responsive biotic transcripts of the *JAZ* and *LOX2* genes result in JA signalling and were significantly up-regulated after 6 hours (Bu et al., 2008, Glauser et al., 2009, Robson et al., 2010). A chitinase, *PR3*, normally up-regulated during fungal infections to destroy chitin, a major component of fungal cell walls (Wang et al., 2009), was the transcript with the highest up-regulation in the 6 hours library at 23 fold. Many *WRKY* transcription factors were up-regulated during tension wood formation and are known to be involved in both biotic and abiotic stress responses (Song et al., 2010, Yamamoto et al., 2004). Despite biotic and abiotic cross-talk, it should be noted that even though the *Eucalyptus* trees sampled for this study showed no evidence of disease, the trees were field grown and biotic defence responses co-occurring with mechanical stress responses may be an expected occurrence.

The number of genes significantly up-regulated in the 6 hours and 24 hours libraries was double the number of genes significantly down-regulated. This is perhaps an indication that the early tension wood response largely involves activation of genes by the mechanosensing and gravity sensing pathways. As discussed above, auxin responsive proteins were found to be down-regulated during early tension wood formation. Lipid anabolism has been shown to be down-regulated for energy conservation in *Arabidopsis* plants undergoing oxidative stress (van Dongen et al., 2009).

Many anabolic lipid genes such as *LTP* family proteins (Baxter et al., 2007), *KCS20* (Lee et al., 2009), *FAD2* (Kajiwara et al., 1996), and *LACS2* (Lü et al., 2009) were significantly down-regulated during tension wood induction which may reflect a stress induced effort to conserve energy.

3.4.3 Late tension wood induction and tension wood formation

It is interesting to note that the number of genes that responded to tension wood induction and the magnitude of the response increased in each successive time point from 6 hours to 6 months after tension wood induction (Table 3.1 and Figure 3.3). As mentioned above, this may reflect the activation of an initial stress response which is later followed by a response in the secondary cell wall biosynthetic pathways. With the exception of a few *FLA* and *XTH* genes, genes other than those involved in stress responses and signalling cascades started to show significant responses after one week. The first phenotypic sign of tension wood formation (decreased lignin content) has previously been reported to take place as early as one week after tension wood induction (Baba et al., 2000) and it is clear from this study also that some secondary cell wall biosynthesis genes started to show a response at one week (Figure 3.4). Some of these genes were the focus of previous tension wood studies and were already discussed above (section 3.4.1). These included cellulose, hemicellulose and lignin biosynthesis and hormone signalling related genes. Additional genes previously not profiled during tension wood formation, but linked to hormone signalling pathways were found to be significantly differentially regulated during late tension wood induction in this study. The *SAG21* gene, which is not directly involved in senescence, but is rather a marker of the stresses that accompany senescence (Weaver et al., 1998), has been shown to be induced by various abiotic stresses (Hundertmark and Hincha, 2008). This gene has been shown to respond rapidly within a few hours of JA treatment (Jung et al., 2007) and was up-regulated by 5 fold and 8 fold in the 6 hours and 24 hours libraries, respectively (Table 3.2 and 3.3). *SAG21* gene continued to respond strongly in late tension wood induction libraries as the highest up-regulated gene in the one week library and reached an up-regulation of 10.52 fold (Table 3.6) in the 6 months library. The *EGY3* gene, a metalloprotease

enzyme and a homolog of *EGY1* (Chen et al., 2005), was up-regulated by 12.14 fold (Table 3.5) in the 2 weeks library. Mutant *egy1 Arabidopsis* lines show defective ethylene dependent gravitropism responses due to the gene's indirect involvement in gravity signalling (Guo et al., 2008). The transcription factor RAP2.2 is an ethylene induced transcription factor which has been found to be expressed in *Arabidopsis* roots and shoots (Hinz et al., 2010). It is a stress related transcription factor which is specifically up-regulated during hypoxic stress (Licausi et al., 2010). *RAN1*, a GTPase which was down-regulated (Table 3.10), is involved in the auxin signal transduction pathway. Overexpression of wheat *RAN1, TaRAN1* in *Arabidopsis* and rice resulted in hypersensitivity to auxin and arrested cell cycle development (Wang et al., 2006). Over-expression of the transcription factor *OSH1* has been shown to depress GA metabolism and signal transduction in tobacco (Kusaba et al., 1998) and kiwi fruit (Kita et al., 2006). Here *OSH1* was down-regulated (Table 3.10) which was consistent with GA signalling involved in tension wood formation. Lastly, the roots of mutant *Arabidopsis wav2* plants bend with greater root curvatures (Mochizuki et al., 2005). *WAV2*, an enzyme which belongs to the *BUDEMERGENCE 46* (*BEM46*) gene family and which negatively regulates root hydrotropism and gravitropism by inhibiting root curvature, was also significantly down-regulated in this study (Mochizuki et al., 2005, Takahashi et al., 2002).

In general, the formation of tension wood is thought to serve two main functions: to either re-orientate a bent stem to an upright position or to maintain tree architecture by stabilizing branches (Bamber, 2001). The formation of reaction wood may also be induced in branches of trees by removing the apical control of dominant branches (Wilson, 2000). In addition, some tree species never re-orientate their stems to an upright position, but instead transfer apical dominance to resprout a new leading stem (Nzunda et al., 2007). Auxin has been shown to exert apical dominance by being involved in the mechanism that acts to represses the outgrowth of buds (Brewer et al., 2009). In contrast, cytokinin phytohormones (CK) stimulate bud outgrowth. Expression of a key enzyme in CK biosynthesis, adenosine phosphate-isopentenyltransferase *PsIPT2*, has been shown to be up-regulated after removing the apical shoots of pea plants (Tanaka et al., 2006). In the present study, the expression of *AtIPT2* was up-regulated almost 2 fold in the 6 months tension wood library and

reduced auxin sensitivity was demonstrated by the down-regulation of auxin related transcripts. Also, in previous studies the expression of Knotted-like homeobox proteins induce the expression of cytokinin biosynthesis (Frugis et al., 2001, Sakamoto et al., 2006), which is in accordance with the strong up-regulation of *KNAT7* expression during *Eucalyptus* tension wood formation. The thickening of 2 to 3 branches that have been turned to a vertical position by bending the *Eucalyptus* stems to induce tension wood formation may therefore have been cases of transferred apical dominance to resprout a new leading stem (Figure 3.2). The thickened branches may compete to replace the lost leader stem and the formation of tension wood may serve to stabilize the bent stem.

A question that remains is whether after the bent stem has been stabilised, *Eucalyptus* continues to form tension wood. Of the 58 genes presented in Figure 3.4 more than half (31) showed an opposite response in the 6 months library compared to the rest of the tension wood series (Figure 3.4 B and C). Taken together with the thickening of a new branch to replace the old stem (Figure 3.2), it is therefore possible that gene expression levels return closer to that of the control once the stem is stabilised. However, genes proven to be of importance to the formation of tension wood continued to show a strong significant response in the 6 months library (Figure 3.4 A and D). These included the two secondary cell wall *CesA* genes *EgCesA2* and *EgCesA3*, the *FLA11* and *FLA12* transcripts, and a number of *XTH* and lignin genes. Nishikubo et al. (2007) showed that high levels of *XTH* expression and enzymatic activity remained in the tension wood of field grown poplar trees for up to three years. Therefore the expression of certain transcripts such as the *XTH* transcripts may continue to be upregulated as these gene products may be needed for the continued mechanism of action of the already formed tension wood (Nishikubo et al., 2007).

3.5 References

ABE, H., URAO, T., ITO, T., SEKI, M., SHINOZAKI, K. & YAMAGUCHI-SHINOZAKI, K. (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell,* 15**,** 63-78.

ALMÉRAS, T. & FOURNIER, M. (2009) Biomechanical design and long-term stability of trees: Morphological and wood traits involved in the balance between weight increase and the gravitropic reaction. *Journal of Theoretical Biology,* 256**,** 370-381.

ANDERSSON-GUNNERÅS, S., MELLEROWICZ, E. J., LOVE, J., SEGERMAN, B., OHMIYA, Y., COUTINHO, P. M., NILSSON, P., HENRISSAT, B., MORITZ, T. & SUNDBERG, B. (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *The Plant Journal,* 45**,** 144-165.

AREND, M. (2008) Immunolocalization of (1,4)-beta-galactan in tension wood fibers of poplar. *Tree Physiology,* 28**,** 1263-1267.

AUDIC, S. & CLAVERIE, J.-M. (1997) The significance of digital gene expression profiles. *Genome research,* 7**,** 986-995.

BABA, K. I., ASADA, T. & HAYASHI, T. (2000) Relation between developmental changes on anatomical structure and on protein pattern in differentiating xylem of tension wood. *Journal of Wood Science,* 46**,** 1-7.

BABA, K. I., PARK, Y. W., KAKU, T., KAIDA, R., TAKEUCHI, M., YOSHIDA, M., HOSOO, Y., OJIO, Y., OKUYAMA, T., TANIGUCHI, T., OHMIYA, Y., KONDO, T., SHANI, Z., SHOSEYOV, O., AWANO, T., SERADA, S., NORIOKA, N., NORIOKA, S. & HAYASHI, T. (2009) Xyloglucan for generating tensile stress to bend tree stem. *Molecular Plant,* 2**,** 893-903.

BAMBER, R. K. (2001) A general theory for the origin of growth stresses in reaction wood: How trees stay upright. *IAWA Journal,* 22**,** 205-212.

BAXTER, C. J., REDESTIG, H., SCHAUER, N., REPSILBER, D., PATIL, K. R., NIELSEN, J., SELBIG, J., LIU, J., FERNIE, A. R. & SWEETLOVE, L. J. (2007) The metabolic response of heterotrophic arabidopsis cells to oxidative stress. *Plant Physiology,* 143**,** 312-325.

BEN-DOR, A., SHAMIR, R. & YAKHINI, Z. (1999) Clustering gene expression patterns. *Journal of Computational Biology,* 6**,** 281-297.

BENJAMINS, R. & SCHERES, B. (2008) Auxin: the looping star in plant development. *Annual Review of Plant Biology,* 59**,** 443-465.

BHANDARI, S., FUJINO, T., THAMMANAGOWDA, S., ZHANG, D., XU, F. & JOSHI, C. (2006) Xylem-specific and tension stress-responsive coexpression of KORRIGAN endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. *Planta,* 224**,** 828-837.

BOWLING, A. J. & VAUGHN, K. C. (2008) Immunocytochemical characterization of tension wood: gelatinous fibers contain more than just cellulose. *American Journal of Botany,* 95**,** 655-663.

BRAAM, J. (2005) In touch: plant responses to mechanical stimuli. *New Phytologist,* 165**,** 373-389.

BREWER, P. B., DUN, E. A., FERGUSON, B. J., RAMEAU, C. & BEVERIDGE, C. A. (2009) Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and *Arabidopsis*. *Plant Physiology,* 150**,** 482-493.

BROWN, D. M., GOUBET, F., WONG, V. W., GOODACRE, R., STEPHENS, E., DUPREE, P. & TURNER, S. R. (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *The Plant Journal,* 52**,** 1154-1168.

BU, Q., JIANG, H., LI, C.-B., ZHAI, Q., ZHANG, J., WU, X., SUN, J., XIE, Q. & LI, C. (2008) Role of the *Arabidopsis thaliana* NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. *Cell Res,* 18**,** 756-767.

CHANG, S., PURYEAR, J. & CAIRNEY, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter,* 11**,** 113-116.

CHEN, G., BI, Y. R. & LI, N. (2005) EGY1 encodes a membrane-associated and ATP-independent metalloprotease that is required for chloroplast development. *The Plant Journal,* 41**,** 364-375.

CLAIR, B., ALMÉRAS, T., YAMAMOTO, H., OKUYAMA, T. & SUGIYAMA, J. (2006a) Mechanical behavior of cellulose microfibrils in tension wood, in relation with maturation stress generation. *Biophysical Journal,* 91**,** 1128-1135.

CLAIR, B., RUELLE, J., BEAUCHENE, J., PREVOST, M. F. & FOURNIER, M. (2006b) Tension wood and opposite wood in 21 tropical rain forest species occurence and efficiency of the G-layer. *IAWA Journal,* 27**,** 329-338.

COUTAND, C. (2010) Mechanosensing and thigmomorphogenesis, a physiological and biomechanical point of view. *Plant Science,* 179**,** 168-182.

DAVID, J.-P., COISSAC, E., MELODELIMA, C., POUPARDIN, R., RIAZ, M., CHANDOR-PROUST, A. & REYNAUD, S. (2010) Transcriptome response to pollutants and insecticides in the dengue vector *Aedes aegypti* using next-generation sequencing technology. *BMC Genomics,* 11**,** 216.

DIETZ, K.-J., VOGEL, M. & VIEHHAUSER, A. (2010) AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma,* 245**,** 3-14.

DJERBI, S., ASPEBORG, H., NILSSON, P., SUNDBERG, B., MELLEROWICZ, E., BLOMQVIST, K. & TEERI, T. T. (2004) Identification and expression analysis of genes encoding putative cellulose synthases (CesA) in the hybrid aspen, *Populus tremula* (L.) × *P. tremuloides* (Michx.). *Cellulose,* 11**,** 301-312.

DU, S. & YAMAMOTO, F. (2003) Ethylene evolution changes in the stems of *Metasequoia glyptostroboides* and *Aesculus turbinata* seedlings in relation to gravity-induced reaction wood formation. *Trees - Structure and Function,* 17**,** 522-528.

FANG, C., CLAIR, B., GRILL, J. & LIU, S. (2008) Growth stresses are highly controlled by the amount of G-layer in poplar tension wood. *IAWA Journal,* 29**,** 237-246.

FISHER, J. B. & STEVENSON, J. W. (1981) Occurrence of reaction wood in branches of dicotyledons and its role in tree architecture. *Botanical Gazette,* 142**,** 82-95.

FRUGIS, G., GIANNINO, D., MELE, G., NICOLODI, C., CHIAPPETTA, A., BITONTI, M. B., INNOCENTI, A. M., DEWITTE, W., VAN ONCKELEN, H. & MARIOTTI, D. (2001)

Overexpression of KNAT1 in lettuce shifts leaf determinate growth to a shoot-like indeterminate growth associated with an accumulation of isopentenyl-type cytokinins. *Plant Physiology*, 126**,** 1370-1380.

FUJITA, M., FUJITA, Y., MARUYAMA, K., SEKI, M., HIRATSU, K., OHME-TAKAGI, M., TRAN, L.-S. P., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *The Plant Journal*, 39**,** 863-876.

FUJITA, M., MIZUKADO, S., FUJITA, Y., ICHIKAWA, T., NAKAZAWA, M., SEKI, M., MATSUI, M., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. (2007) Identification of stress-tolerance-related transcription-factor genes via mini-scale full-length cDNA over-expressor (FOX) gene hunting system. *Biochemical and Biophysical Research Communications*, 364**,** 250-257.

FUNADA, R., MIURA, T., SHIMIZU, Y., KINASE, T., NAKABA, S., KUBO, T. & SANO, Y. (2008) Gibberellin-induced formation of tension wood in angiosperm trees. *Planta*, 227**,** 1409-1414.

GAO, F., XIONG, A., PENG, R., JIN, X., XU, J., ZHU, B., CHEN, J. & YAO, Q. (2010) *OsNAC52*, a rice NAC transcription factor, potentially responds to ABA and confers drought tolerance in transgenic plants. *Plant Cell, Tissue and Organ Culture*, 100**,** 255-262.

GLAUSER, G., DUBUGNON, L., MOUSAVI, S. A. R., RUDAZ, S., WOLFENDER, J.-L. & FARMER, E. E. (2009) Velocity estimates for signal propagation leading to systemic jasmonic acid accumulation in wounded *Arabidopsis*. *Journal of Biological Chemistry*, 284**,** 34506-34513.

GLOMBITZA, S., DUBUIS, P.-H., THULKE, O., WELZL, G., BOVET, L., GÖTZ, M., AFFENZELLER, M., GEIST, B., HEHN, A., ASNAGHI, C., ERNST, D., SEIDLITZ, H., GUNDLACH, H., MAYER, K., MARTINOIA, E., WERCK-REICHHART, D., MAUCH, F. & SCHÄFFNER, A. (2004) Crosstalk and differential response to abiotic and biotic stressors reflected at the transcriptional level of effector genes from secondary metabolism. *Plant Molecular Biology*, 54**,** 817-835.

GORIŠEK, Ž. & STRAŽE, A. (2006) Sorption and swelling characteristics of normal and tension beech wood (*Fagus sylvatica* L.). *Wood Structure and Properties,* 1**,** 227-231.

GOSWAMI, L., DUNLOP, J. W. C., JUNGNIKL, K., EDER, M., GIERLINGER, N., COUTAND, C., JERONIMIDIS, G., FRATZL, P. & BURGERT, I. (2008) Stress generation in the tension wood of poplar is based on the lateral swelling power of the G-layer. *The Plant Journal,* 56**,** 531-538.

GOVRIN, E. M. & LEVINE, A. (2000) The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. *Current Biology,* 10**,** 751-757.

GUO, D., GAO, X., LI, H., ZHANG, T., CHEN, G., HUANG, P., AN, L. & LI, N. (2008) *EGY1* plays a role in regulation of endodermal plastid size and number that are involved in ethylene-dependent gravitropism of light-grown *Arabidopsis* hypocotyls. *Plant Molecular Biology,* 66**,** 345-360.

HAHN, A. & HARTER, K. (2009) Mitogen-activated protein kinase cascades and ethylene: signaling, biosynthesis, or both? *Plant Physiology,* 149**,** 1207-1210.

HANRIOT, L., KEIME, C., GAY, N., FAURE, C., DOSSAT, C., WINCKER, P., SCOTE-BLACHON, C., PEYRON, C. & GANDRILLON, O. (2008) A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. *BMC Genomics,* 9**,** 418.

HEGEDUS, Z., ZAKRZEWSKA, A., ÁGOSTON, V. C., ORDAS, A., RÁCZ, P., MINK, M., SPAINK, H. P. & MEIJER, A. H. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Molecular Immunology,* 46**,** 2918-2930.

HEJNOWICZ, Z. (1997) Graviresponses in herbs and trees: a major role for the redistribution of tissue and growth stresses. *Planta,* 203**,** S136-S146.

HELLEMANS, J., MORTIER, G., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology,* 8**,** R19.

HINZ, M., WILSON, I. W., YANG, J., BUERSTENBINDER, K., LLEWELLYN, D., DENNIS, E. S., SAUTER, M. & DOLFERUS, R. (2010) *Arabidopsis* RAP2.2: an ethylene response transcription factor that is important for hypoxia survival. *Plant Physiology,* 153**,** 757-772.

HUNDERTMARK, M. & HINCHA, D. (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics,* 9**,** 118.

JAFFE, M. J. (1980) Morphogenic responses of plants to mechanical stimuli or stress. *BioScience,* 30**,** 239-243.

JIANG, S., XU, K., WANG, Y.-Z., REN, Y.-P. & GU, S. (2008) Role of GA3, GA4 and uniconazole-p in controlling gravitropism and tension wood formation in *Fraxinus mandshurica* Rupr. var. *japonica* Maxim. seedlings. *Journal of Integrative Plant Biology,* 50**,** 19-28.

JIANG, S., XU, K., ZHAO, N., ZHENG, S.-X., REN, Y.-P., GAO, Y.-B. & GU, S. (2009) Ethylene evolution changes in tilted *Fraxinus mandshurica* Rupr. var. *japonica* Maxim. seedlings in relation to tension wood formation. *Journal of Integrative Plant Biology,* 51**,** 707-713.

JIANG, Y. & DEYHOLOS, M. (2009) Functional characterization of *Arabidopsis* NaCl-inducible *WRKY25* and *WRKY33* transcription factors in abiotic stresses. *Plant Molecular Biology,* 69**,** 91-105.

JOHNSON, K. L., JONES, B. J., BACIC, A. & SCHULTZ, C. J. (2003) The fasciclin-like arabinogalactan proteins of *Arabidopsis*. A multigene family of putative cell adhesion molecules. *Plant Physiology,* 133**,** 1911-1925.

JOSELEAU, J.-P., IMAI, T., KURODA, K. & RUEL, K. (2004) Detection in situ and characterization of lignin in the *G*-layer of tension wood fibres of *Populus deltoides*. *Planta,* 219**,** 338-345.

JOUREZ, B., RIBOUX, A. & LECLERCQ, A. (2001) Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* cv 'Ghoy'). *International Association of Wood Anatomists,* 22**,** 133-157.

JUNG, C., LYOU, S., YEU, S., KIM, M., RHEE, S., KIM, M., LEE, J., CHOI, Y. & CHEONG, J.-J. (2007) Microarray-based screening of jasmonate-responsive genes in *Arabidopsis thaliana*. *Plant Cell Reports,* 26**,** 1053-1063.

KAGAYA, Y. & HATTORI, T. (2009) *Arabidopsis* transcription factors, RAV1 and RAV2, are regulated by touch-related stimuli in a dose-dependent and biphasic manner. *Genes & Genetic Systems,* 84**,** 95-99.

KAJIWARA, S., SHIRAI, A., FUJII, T., TOGURI, T., NAKAMURA, K. & OHTAGUCHI, K. (1996) Polyunsaturated fatty acid biosynthesis in *Saccharomyces cerevisiae*: expression of ethanol tolerance and the *FAD2* gene from *Arabidopsis thaliana*. *Applied Environmental Microbioogy.,* 62**,** 4309-4313.

KAKU, T., SERADA, S., BABA, K. I., TANAKA, F. & HAYASHI, T. (2009) Proteomic analysis of the G-layer in poplar tension wood. *Journal of Wood Science,* 55**,** 250-257.

KITA, M., HONDA, C., KOMATSU, S., KUSABA, S., FUJII, Y. & MORIGUCHI, T. (2006) Protein profile in the transgenic kiwifruit overexpressing a transcription factor gene, *OSH1*. *Biologia Plantarum,* 50**,** 759-762.

KONG, Y., ZHOU, G., AVCI, U., GU, X., JONES, C., YIN, Y., XU, Y. & HAHN, M. G. (2009) Two poplar glycosyltransferase genes, *PdGATL1.1* and *PdGATL1.2,* are functional orthologs to *PARVUS/AtGATL1* in *Arabidopsis*. *Molecular Plant,* 2**,** 1040-1050.

KUSABA, S., FUKUMOTO, M., HONDA, C., YAMAGUCHI, I., SAKAMOTO, T. & KANO-MURAKAMI, Y. (1998) Decreased GA1 content caused by the overexpression of OSH1 is accompanied by suppression of GA 20-oxidase gene expression. *Plant Physiology,* 117**,** 1179-1184.

LAFARGUETTE, F., LEPLÉ, J.-C., DÉJARDIN, A., LAURANS, F., COSTA, G., LESAGE-DESCAUSES, M.-C. & PILATE, G. (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytologist,* 164**,** 107-121.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* 10**,** R25.

LEE, S.-B., JUNG, S.-J., GO, Y.-S., KIM, H.-U., KIM, J.-K., CHO, H.-J., PARK, O. K. & SUH, M.-C. (2009) Two *Arabidopsis* 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are

functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *The Plant Journal,* 60**,** 462-475.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LICAUSI, F., VAN DONGEN, J. T., GIUNTOLI, B., NOVI, G., SANTANIELLO, A., GEIGENBERGER, P. & PERATA, P. (2010) HRE1 and HRE2, two hypoxia-inducible ethylene response factors, affect anaerobic responses in *Arabidopsis thaliana*. *The Plant Journal,* 62**,** 302-315.

LIU, X., JIANG, J., SHAO, J., YIN, Y. & MA, Z. (2010) Gene transcription profiling of *Fusarium graminearum* treated with an azole fungicide tebuconazole. *Applied Microbiology and Biotechnology,* 85**,** 1105-1114.

LOVE, J., BJÖRKLUND, S., VAHALA, J., HERTZBERG, M., KANGASJÄRVI, J. & SUNDBERG, B. (2009) Ethylene is an endogenous stimulator of cell division in the cambial meristem of *Populus*. *Proceedings of the National Academy of Sciences,* 106**,** 5984-5989.

LÜ, S., SONG, T., KOSMA, D. K., PARSONS, E. P., ROWLAND, O. & JENKS, M. A. (2009) *Arabidopsis CER8* encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. *The Plant Journal,* 59**,** 553-564.

MACMILLAN, C. P., MANSFIELD, S. D., STACHURSKI, Z. H., EVANS, R. & SOUTHERTON, S. G. (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus*. *The Plant Journal,* 62**,** 689-703.

MALONEY, V. J. & MANSFIELD, S. D. (2010) Characterization and varied expression of a membrane-bound endo-β-1,4-glucanase in hybrid poplar. *Plant Biotechnology Journal,* 8**,** 294-307.

MANDADI, K. K., MISRA, A., REN, S. & MCKNIGHT, T. D. (2009) BT2, a BTB protein, mediates multiple responses to nutrients, stresses, and hormones in *Arabidopsis*. *Plant Physiology,* 150**,** 1930-1939.

MATSUKURA, S., MIZOI, J., YOSHIDA, T., TODAKA, D., ITO, Y., MARUYAMA, K., SHINOZAKI, K. & YAMAGUCHI-SHINOZAKI, K. (2010) Comprehensive analysis of rice *DREB2*-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Molecular Genetics and Genomics,* 283**,** 185-196.

MELLEROWICZ, E. J., IMMERZEEL, P. & HAYASHI, T. (2008) Xyloglucan: The molecular muscle of trees. *Annals of Botany*, 102**,** 659-665.

MITTLER, R., KIM, Y., SONG, L., COUTU, J., COUTU, A., CIFTCI-YILMAZ, S., LEE, H., STEVENSON, B. & ZHU, J.-K. (2006) Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. *FEBS Letters,* 580**,** 6537-6542.

MIZRACHI, E., HEFER, C., RANIK, M., JOUBERT, F. & MYBURG, A. (2010) De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics*, 11**,** 681.

MOCHIZUKI, S., HARADA, A., INADA, S., SUGIMOTO-SHIRASU, K., STACEY, N., WADA, T., ISHIGURO, S., OKADA, K. & SAKAI, T. (2005) The *Arabidopsis* WAVY GROWTH 2 protein modulates root bending in response to environmental stimuli. *Plant Cell,* 17**,** 537-547.

MOHR, P. & CAHILL, D. (2007) Suppression by ABA of salicylic acid and lignin accumulation and the expression of multiple genes, in *Arabidopsis* infected with *Pseudomonas syringae* pv. *tomato*. *Functional & Integrative Genomics*, 7**,** 181-191.

MOYLE, R., SCHRADER, J., STENBERG, A., OLSSON, O., SAXENA, S., SANDBERG, G. & BHALERAO, R. P. (2002) Environmental and auxin regulation of wood formation involves members of the Aux/IAA gene family in hybrid aspen. *The Plant Journal,* 31**,** 675-685.

NISHIKUBO, N., AWANO, T., BANASIAK, A., BOURQUIN, V., IBATULLIN, F., FUNADA, R., BRUMER, H., TEERI, T. T., HAYASHI, T., SUNDBERG, B. R. & MELLEROWICZ, E. J. (2007) Xyloglucan endo-transglycosylase (XET) functions in gelatinous layers of tension wood fibers in poplar - a glimpse into the mechanism of the balancing act of trees. *Plant and Cell Physiology*, 48**,** 843-855.

NORBERG, P. H. & MEIER, H. (1966) Physical and chemical properties of the gelatinous layer in tension wood fibres of aspen (*Populus tremula L.*). *Holzforschung*, 20**,** 174-178.

NZUNDA, E. F., GRIFFITHS, M. E. & LAWES, M. J. (2007) Resprouting versus turning up of leaning trees in a subtropical coastal dune forest in South Africa. *Journal of Tropical Ecology,* 23**,** 289-296.

OSAKABE, Y., MIZUNO, S., TANAKA, H., MARUYAMA, K., OSAKABE, K., TODAKA, D., FUJITA, Y., KOBAYASHI, M., SHINOZAKI, K. & YAMAGUCHI-SHINOZAKI, K. (2009) Overproduction of the membrane-bound receptor-like protein kinase 1, RPK1, enhances abiotic stress tolerance in *Arabidopsis*. *Journal of Biological Chemistry,* 285**,** 9190-9201.

PAUX, E., CAROCHA, V., MARQUES, C., MENDES DE SOUSA, A., BORRALHO, N., SIVADON, P. & GRIMA-PETTENATI, J. (2005) Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytologist,* 167**,** 89-100.

PERSSON, S., CAFFALL, K. H., FRESHOUR, G., HILLEY, M. T., BAUER, S., POINDEXTER, P., HAHN, M. G., MOHNEN, D. & SOMERVILLE, C. (2007) The *Arabidopsis irregular xylem8* mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. *Plant Cell,* 19**,** 237-255.

PEWZNER-JUNG, Y., BRENNER, O., BRAUN, S., LAVIAD, E. L., BEN-DOR, S., FELDMESSER, E., HORN-SABAN, S., AMANN-ZALCENSTEIN, D., RAANAN, C., BERKUTZKI, T., EREZ-ROMAN, R., BEN-DAVID, O., LEVY, M., HOLZMAN, D., PARK, H., NYSKA, A., MERRILL, A. H. & FUTERMAN, A. H. (2010) A critical role for ceramide synthase 2 in liver homeostasis. *Journal of Biological Chemistry,* 285**,** 10911-10923.

PILATE, G., CHABBERT, B., CATHALA, B., YOSHINAGA, A., LEPLÉ, J.-C., LAURANS, F., LAPIERRE, C. & RUEL, K. (2004a) Lignification and tension wood. *Comptes Rendus Biologies,* 327**,** 889-901.

PILATE, G., DÉJARDIN, A., LAURANS, F. & LEPLÉ, J.-C. (2004b) Tension wood as a model for functional genomics of wood formation. *New Phytologist,* 164**,** 63-72.

PRODHAN, A. K. M. A., OHTANI, J., FUNADA, R., ABE, H. & FUKAZAWA, K. (1995) Ultrastructural investigation of tension wood fibre in *Fraxinus mandshurica* Rupr. var. *japonica* Maxim. *Annals of Botany,* 75**,** 311-317.

QIU, D., WILSON, I. W., GAN, S., WASHUSEN, R., MORAN, G. F. & SOUTHERTON, S. G. (2008) Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. *New Phytologist,* 179**,** 94-103.

ROBSON, F., OKAMOTO, H., PATRICK, E., HARRIS, S.-R., WASTERNACK, C., BREARLEY, C. & TURNER, J. G. (2010) Jasmonate and phytochrome A signaling in *Arabidopsis* wound and shade responses are integrated through JAZ1 stability. *Plant Cell,* 22**,** 1143-1160.

ROMUALDI, C., BORTOLUZZI, S., D'ALESSI, F. & DANIELI, G. (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol. Genomics,* 12**,** 159-162.

SAKAMOTO, T., SAKAKIBARA, H., KOJIMA, M., YAMAMOTO, Y., NAGASAKI, H., INUKAI, Y., SATO, Y. & MATSUOKA, M. (2006) Ectopic expression of KNOTTED1-like homeobox protein induces expression of cytokinin biosynthesis genes in rice. *Plant Physiol.,* 142**,** 54-62.

SONG, Y., CHEN, L., ZHANG, L. & YU, D. (2010) Overexpression of *OsWRKY72* gene interferes in the abscisic acid signal and auxin transport pathway of *Arabidopsis*. *Journal of Biosciences,* 35**,** 459-471.

SREENIVASULU, N., SOPORY, S. K. & KAVI KISHOR, P. B. (2007) Deciphering the regulatory mechanisms of abiotic stress tolerance in plants by genomic approaches. *Gene,* 388**,** 1-13.

TAKAHASHI, N., GOTO, N., OKADA, K. & TAKAHASHI, H. (2002) Hydrotropism in abscisic acid, wavy, and gravitropic mutants of *Arabidopsis thaliana*. *Planta,* 216**,** 203-211.

TAKAHASHI, S., SEKI, M., ISHIDA, J., SATOU, M., SAKURAI, T., NARUSAKA, M., KAMIYA, A., NAKAJIMA, M., ENJU, A., AKIYAMA, K., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. (2004) Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in *Arabidopsis* cell culture using a full-length cDNA microarray. *Plant Molecular Biology,* 56**,** 29-55.

TANAKA, M., TAKEI, K., KOJIMA, M., SAKAKIBARA, H. & MORI, H. (2006) Auxin controls local cytokinin biosynthesis in the nodal stem in apical dominance. *The Plant Journal,* 45**,** 1028-1036.

THATCHER, L. F., ANDERSON, J. P. & SINGH, K. B. (2005) Plant defense responses: what have we learnt from *Arabidopsis? Functional Plant Biology,* 32**,** 1-19.

VAN DONGEN, J. T., FROHLICH, A., RAMIREZ-AGUILAR, S. J., SCHAUER, N., FERNIE, A. R., ERBAN, A., KOPKA, J., CLARK, J., LANGER, A. & GEIGENBERGER, P. (2009) Transcript and metabolite profiling of the adaptive response to mild decreases in oxygen concentration in the roots of *Arabidopsis* plants. *Annals of Botany,* 103**,** 269-280.

WANG, J., TIAN, N., HUANG, X., CHEN, L., SCHLÄPPI, M. & XU, Z. (2009) The tall fescue turf grass class I chitinase gene *FaChit1* is activated by fungal elicitors, dehydration, ethylene, and mechanical wounding. *Plant Molecular Biology Reporter,* 27**,** 305-314.

WANG, X.-W., LUAN, J.-B., LI, J.-M., BAO, Y.-Y., ZHANG, C.-X. & LIU, S.-S. (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics,* 11**,** 400.

WANG, X., XU, Y., HAN, Y., BAO, S., DU, J., YUAN, M., XU, Z. & CHONG, K. (2006) Overexpression of *RAN1* in rice and *Arabidopsis* alters primordial meristem, mitotic progress, and sensitivity to auxin. *Plant Physiology,* 140**,** 91-101.

WASTENEYS, G. & FUJITA, M. (2006) Establishing and maintaining axial growth: wall mechanical properties and the cytoskeleton. *Journal of Plant Research,* 119**,** 5-10.

WEAVER, L. M., GAN, S., QUIRINO, B. & AMASINO, R. M. (1998) A comparison of the expression patterns of several senescence-associated genes in response to stress and hormone treatment. *Plant Molecular Biology,* 37**,** 455-469.

WELLER, J. L., HECHT, V., VANDER SCHOOR, J. K., DAVIDSON, S. E. & ROSS, J. J. (2009) Light regulation of gibberellin biosynthesis in pea is mediated through the COP1/HY5 pathway. *Plant Cell,* 21**,** 800-813.

WILSON, B. F. (2000) Apical control of branch growth and angle in woody plants. *American Journal of Botany.,* 87**,** 601-607.

XIAO, S., JIA, J., MO, D., WANG, Q., QIN, L., HE, Z., ZHAO, X., HUANG, Y., LI, A., YU, J., NIU, Y., LIU, X. & CHEN, Y. (2010) Understanding PRRSV infection in porcine lung based on genome-wide transcriptome response identified by deep sequencing. *PLoS ONE,* 5**,** e11377.

XU, J., LI, Y., WANG, Y., LIU, H., LEI, L., YANG, H., LIU, G. & REN, D. (2008) Activation of MAPK kinase 9 induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in *Arabidopsis*. *Journal of Biological Chemistry,* 283**,** 26996-27006.

YAMAMOTO, H., YOSHIDA, M. & OKUYAMA, T. (2002) Growth stress controls negative gravitropism in woody plant stems. *Planta,* 216**,** 280-292.

YAMAMOTO, S., NAKANO, T., SUZUKI, K. & SHINSHI, H. (2004) Elicitor-induced activation of transcription via W box-related cis-acting elements from a basic chitinase gene by WRKY transcription factors in tobacco. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1679**,** 279-287.

YANHUI, C., XIAOYUAN, Y., KUN, H., MEIHUA, L., JIGANG, L., ZHAOFENG, G., ZHIQIANG, L., YUNFEI, Z., XIAOXIAO, W., XIAOMING, Q., YUNPING, S., LI, Z., XIAOHUI, D., JINGCHU, L., XING-WANG, D., ZHANGLIANG, C., HONGYA, G. & LI-JIA, Q. (2006) The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology,* 60**,** 107-124.

YOO, S.-D., CHO, Y.-H., TENA, G., XIONG, Y. & SHEEN, J. (2008) Dual control of nuclear EIN3 by bifurcate MAPK cascades in C2H4 signalling. *Nature,* 451**,** 789-795.

ZHONG, R., DEMURA, T. & YE, Z.-H. (2006) SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell,* 18**,** 3158-3170.

ZHONG, R., LEE, C. & YE, Z. H. (2010) Functional characterisation of poplar wood-associated NAC domain transcription factors. *Plant Physiology,* 152**,** 1044-1055.

ZHONG, R., LEE, C., ZHOU, J., MCCARTHY, R. L. & YE, Z.-H. (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell,* 20**,** 2763-2782.

ZHONG, R., RICHARDSON, E. & YE, Z.-H. (2007) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta,* 225**,** 1603-1611.

CONCLUDING REMARKS

In this study, successful induction of tension wood formation and DGE analysis enabled a near transcriptome-wide assessment of gene expression profiles in response to bending stress. Expression profiling using Illumina's DGE tag profiling platform and data processing guidelines developed in Chapter 2 confirmed that tension wood formation is a complex process. An in-depth investigation of the nature of DGE data revealed a putative alternative splice variant of the *EgCesA3* gene expressed during tension wood formation. Early tension wood induction represented by the 6 hours and 24 hours libraries involved complex signalling networks through the activation of multiple stress response pathways. These include mechanosensing and gravisensing pathways along with various hormonal signalling pathways. Transcripts involved in JA, GA, ABA, and ethylene signalling were up-regulated and those involved in auxin signalling were down-regulated. Abiotic stress responses also activated biotic stress responses through "cross-talk" transcripts shared by both stress response divisions. In this study the one week library was categorised as an early tension wood formation library according to previous observations that the first phenotypic signs of tension wood are visible at one week after induction (Baba et al., 2000). Both late tension wood formation libraries (2 weeks and 6 months) continued to show stress related differential expression in addition to the differential regulation of secondary cell wall biosynthetic pathways. Down-regulation of transcripts related to lignin and xylan biosynthesis and up-regulation of transcripts related to cellulose biosynthesis (*EgCesA2* and *EgCesA3*) and xyloglucan reflected the chemical composition of tension wood. Lastly, increased *FLA* expression, transcripts that code for important structural cell wall proteins was highly up-regulated and were the highest expressed transcripts in the study confirming the important role these proteins play during tension wood formation.

Expression profiles inconsistent with literature were obtained for *SUS1* and *EgCesA1* as these cellulose biosynthesis related transcripts were not significantly differentially regulated. However, as shown before Andersson-Gunnerås et al. (2006) also found that cellulose synthase gene transcripts were generally not increased during poplar tension wood formation. In addition, it should be kept in mind that the regulation of the abundance of a particular transcript only partially reflects phenotypic

expression of the final gene product. Multiple levels of regulation need to be investigated to comprehend complex developmental processes such as tension wood formation.

Results suggest that the *Eucalyptus grandis x Eucalyptus urophylla* trees used in this study formed tension wood to stabilise the bent stem, while a new branch was thickened to replace the lost stem. A future study making use of quantitative measurements aimed at investigating the thickening of a side branch to replace the bent stem is needed to verify this observation. This is to our knowledge the first study that has profiled gene expression during *Eucalyptus* tension wood formation at 6 months after induction. Future work may investigate which biosynthetic pathways and gene networks as a whole are differentially regulated in these libraries. This may reveal which biosynthetic pathways in the 6 months library continued to respond to the bending stress to gain a more comprehensive overview of this process. This is essential as the small number of genes mentioned in this pilot study may only provide a partial representation of tension wood formation at 6 months.

References

ANDERSSON-GUNNERÅS, S., MELLEROWICZ, E. J., LOVE, J., SEGERMAN, B., OHMIYA, Y., COUTINHO, P. M., NILSSON, P., HENRISSAT, B., MORITZ, T. & SUNDBERG, B. (2006) Biosynthesis of cellulose-enriched tension wood in Populus: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *The Plant Journal,* 45**,** 144-165.

BABA, K. I., ASADA, T. & HAYASHI, T. (2000) Relation between developmental changes on anatomical structure and on protein pattern in differentiating xylem of tension wood. *Journal of Wood Science,* 46**,** 1-7.