This is to certify that the
dissertation entitled

GENOMIC APPROACHES TO HEARTWOOD FORMATION
IN HARDWOOD TREE SPECIES,
BLACK LOCUST (*ROBINIA PSEUDOACACIA* L.)

presented by

JAEMO YANG

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Plant Breeding and Genetic - Forestry

Major Professor's Signature

8/9/04

Date

GENOMIC APPROACHES TO HEARTWOOD FORMATION IN HARDWOOD TREE SPECIES, BLACK LOCUST (*ROBINIA PSEUDOACACIA* L.)

By

Jaemo Yang

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirement
For the degree of

DOCTOR OF PHILOSOPHY

Plant Breeding and Genetics - Forestry

2004

ABSTRACT

GENOMIC APPROACHES TO HEARTWOOD FORMATION IN HARDWOOD TREE SPECIES, BLACK LOCUST (*ROBINIA PSEUDOACACIA* L.)

By

Jaemo Yang

Trees comprise over 90% of the terrestrial biomass of the earth and serve as a primary feedstock for biofuel, fiber, solid wood products, and various natural compounds. Most commercially important tree crops produce heartwood. The presence of heartwood is the major determining factor for wood quality and influences the way in which specific woods are utilized. Understanding the molecular mechanisms of heartwood formation is of great commercial and keen scientific interest. Despite the long history of study on wood formation, our knowledge of this unique biological process is limited. It is difficult to experimentally observe the processes in sapwood that lead to heartwood formation. Forward genetic approaches are hampered by long generation times and poorly defined tree populations. Comparative molecular genetic studies have limited use because most model organisms do not undergo secondary woody growth. The changes during the transition from sapwood to heartwood are complex and involve integration of many metabolic processes in wood cells. A genomics-based approach provides a unique opportunity for understanding the molecular biology of heartwood and the processes by which its extractives are produced and stored.

In order to gain insight on the molecular regulations of heartwood and its extractive formation, I carried out global examination of gene expression profiles across the trunk wood of black locust *(Robinia pseudoacacia* L.) trees. Of the 2,915 expressed sequenced tags (ESTs) that were generated and analyzed, 55.3% showed no match to known

sequences. Cluster analysis of the ESTs identified a total of 2,278 unigene sets, which were used to construct cDNA microarrays. Microarray hybridization analyses were then performed to survey the changes in gene expression profiles of trunk wood. In addition to the regional gene expression profiles in trunk wood, seasonal gene expression changes were studied in the sapwood-heartwood transition zones that are in charge of heartwood formation. Furthermore, in this study, plant allantoinase genes involved in ureide pathway were investigated; in general, plants have the pathway, and these genes are differentially expressed in developmental and environmental conditions. Therefore, the impact of this study is expected to expand our knowledge of heartwood formation far beyond a single hardwood species.

DEDICATION

This dissertation is dedicated to my wife, Youngmi Kim, my parents, Seungsoon Yang and Sunsoon Lee, my parents in law, Dongho Kim and Soonja Lee, and my children, Jiwoo and Joanne.

ACKNOWLEDGEMENTS

I specially thank Dr. Kyung-Hwan Han, my advisor and professor, for his continuous support, expertise, and for the opportunity I had to pursue my graduate program under his guidance. His patience and generosity in terms of time and resources were invaluable and helped me throughout the study. I thank also to my other committee members, Dr. Amy Iezzoni, Dr. Wayne Loescher and Dr. Kenneth Sink for their helpful comments, encouragement, and positive criticism.

Special thanks go to Dr. Daniel Keathley and Dr. D. Pascal Kamdem, for their invaluable support, friendship, and help for my research. I extend my sincere thanks to Dr. Jeff Landgrof for his help with microarray printing, Paul Bloese for help with tree sampling, Dr. Jae-Heung Ko for help with valuable discussion, and Merilyn Ruthig for her help with manuscript editing. Many others deserve my acknowledgements and thanks including Dr. David MacFarlane, Sunchung Park, Costas Prassinos, and colleagues and staff members of the Department of Forestry at Michigan State University.

Finally, I would like to take the opportunity to say my special thank, love and gratitude to my wife, my parents, my parents in law, sisters, brothers and my children that have been there for me and always an continuous source of inspiration.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

INTRODUCTION 1

CHAPTER 1 13
Introduction 14
Materials and Methods 16
Results 22
Discussion 48
Literature Cited 55

CHAPTER 2 60
Introduction 61
Materials and Methods 64
Results 68
Discussion 85
Literature Cited 92

CHAPTER 3 97
Introduction 98
Materials and Methods 101
Results 108
Discussion 122
Literature Cited 127

CONCLUSION 131

LIST OF TABLES

Table 1-1. The functional classification of EST clones 26
Table 1-2. The redundancy of EST clones based on contig analysis 30
Table 1-3. Summary of up-regulated or down-regulated genes 34
Table 1-4. Up-regulated transcripts in the bark and cambial zone 37
Table 1-5. Up-regulated transcripts in the sapwood 40
Table 1-6. Up-regulated transcripts in the transition zone 41
Table 2-1. Summary of correlation coefficients of ratios from TZS versus TZF 69
Table 2-2. Functional classification of up-regulated clones in transition zones 73
Table 2-3. Categorized genes whose expression was up-regulated in summer (TZS). 75
Table 2-4. List of up-regulated genes in transition zones harvested in fall (TZF) 80

LIST OF FIGURES

Figure 1-1. Cross-section of a stem from a mature *Robinia pseudoacacia* tree 23
Figure 1-2. Functional classification of the ESTS from three libraries 27
Figure 1-3. Scatter plots of microarray hybridization results 32
Figure 1-4. Cluster analysis of expression ratios from SWS or TZS vs. BCS 44
Figure 1-5. Confirmation of microarray Data with antisense northern blot analysis 47
Figure 2-1. Distribution of mean ratios of expression from TZS and TZF samples 72
Figure 2-2. Metabolic pathway leading to phenylpropanoids 84
Figure 3-1. Catabolism of ureide in plants 100
Figure 3-2. Alignment of allantoinase amino acid sequences 109
Figure 3-3. Functional complementation of a yeast allantoinase mutant 111
Figure 3-4. Identification of T-DNA insertional mutants of Arabidopsis allantoinase .. 112
Figure 3-5. Dendrogram and similarity of allantoinases and dihydroorotases 115
Figure 3-6. Differential expression of black locust allantoinase 117
Figure 3-7. RpALN gene expression in trunk wood of black locust 120
Figure 3-8. Nitrogen influence on the expression of allantoinase genes 121

INTRODUCTION

General information.

Sapwood is gradually converted to inactive heartwood as its water columns of conducting vessels break due to freezing, wind vibration, tension, wood boring insects, and other factors (Mauseth, 1998). The ultimate fate of the cavity of the broken vessels affects different properties of wood. Trees adopt a mechanism to seal off the empty columns. The adjacent wood parenchyma cells undergo numerous metabolic changes to produce and accumulate in the vessels large quantity of heartwood extractives such as phenolic compounds, lignin, and aromatic substances. These substances inhibit microbial growth. In this process, which occurs in late fall in temperate zones, one annual ring is converted to heartwood each year (Mauseth, 1998).

Formation of heartwood is a form of senescence that is accompanied by a variety of alterations in metabolic conditions. Although the events of senescence have been studied at the molecular level during leaf senescence (Miller et al., 1999; Wingler et al., 1998), seed germination (Cercos et al., 1999), nodule development (Matamoros et al., 1999) and floral senescence (Breeze et al., 2004), the cell maturation and death events occurring during heartwood formation have been difficult to study because of the location and timing of the events. Therefore, analysis of global gene expression patterns during the transition may offer a powerful means of identifying the mechanisms controlling this process.

In practical sense, the presence of heartwood is a determining factor for wood quality and influences the utilization of wood in many different ways. It also adversely affects forest management decisions by making it difficult to predict the quality of wood

available for utilization from forest inventory. Understanding the biology of heartwood formation can provide to control the contributing factors for its formation and thus to make choices of the most suitable forestry practices. However, no study on the molecular biology of heartwood formation has been reported.

Timing of heartwood formation.

In the temperate zones, heartwood formation occurs in late summer to late fall and at the beginning of dormancy when temperatures are sufficiently high (above 5°C) for the required cellular reactions (Hillis, 1987). Earlier studies have indicated that heartwood formation occurs at the time of cambial dormancy in pine (Shain and Mackay, 1973a), walnut and cherry (Nelson, 1978). Studies of the cytology and coloration of the extractives suggested that heartwood formation commences in mid-summer and continues in the fall and winter seasons in sugi (Nobuchi et al., 1984a) and black locust (Nobuchi et al., 1984b).

Metabolic activity in transition zone.

During the transformation from sapwood to heartwood, cells undergo metabolic changes that result in increased synthesis of secondary products. These involve the consumption of storage carbohydrates and their conversion into heartwood substances such as phenolic compounds (Hillis, 1987; Magel et al., 1991; Magel et al., 1994). Observations lead to the suggestion that heartwood extractives are synthesized *in situ* in the sapwood-heartwood transition zone from the breakdown of starch or from soluble sugars (Magel et al., 1994; Magel and Hübner, 1997). The transition is tightly connected

with the degradation of storage lipids and the accumulation of hydrolysis products and intermediates (Hillinger et al., 1996a and b). Light and electron microscopic studies have found that during sapwood-heartwood transition, storage material such as starch is consumed (Datta and Kumar, 1987; Nair, 1988; Nobuchi et al., 1987). Starch hydrolyzed at the transition zone represents a primary source of hydroxycinnamic acid and flavonoid synthesis (Magel et al., 1994). A period of enhanced metabolic activity has been found at this zone in heartwood forming species such as *Acacia* (Baqui and Shah, 1985), black locust (Magel et al., 1991), oak (Ebermann and Stich, 1985), and walnut (Nelson et al., 1981). Maximum oxygen consumption was observed in the sapwood adjacent to the heartwood of black locust, suggesting increased metabolic activity (Höll and Lendzian, 1973). Also, the formation and accumulation of heartwood phenolics coincided with the transformation of sapwood to heartwood in black locust (Magel et al., 1994). Hillis and Hasegawa (1963) noted that, 19 days after labeled glucose was administered to the cambial region of *Eucalyptus sieberi*, it was converted to labeled extractives formed at the heartwood periphery. This conversion was also observed in the transition zone of Sugi (Higuchi et al., 1969).

Several enzymes have shown increased activity in the transition zone. Elevated levels of phenol oxidizing enzymes have been observed in the transition zone of *Eucalyptus polyanthemos* (Hillis and Yazaki, 1973). A marked peroxidase activity was observed in the transition zone of *Eucalyptus elaeophora* (Wardrop and Cronshaw, 1962) and *Fagus sylvatica* (Dietrichs, 1964). In addition, Baqui et al. (1979) found that succinate dehydrogenase was significantly active only in the transition zone of *Melia azedarach*. Adenosine triphosphatase, which is implicated in many energy-consuming

cellular processes, and lipase were also active in the transition zone. Other enzymes reported to be highly active in transition zone include catechol oxidase, glucose-6-phosphate dehydrogenase, malic dehydrogenase, maltase, and amylase (Hillis, 1987). Furthermore, Magel et al. (1991) have shown that two key enzymes for flavonoid biosynthesis (chalcone synthase and phenylalanine ammonia-lyase) are highly active in the sapwood-heartwood transition zone of black locust. They investigated the seasonal activities of phenylalanine ammonia-lyase (PAL) and chalcone synthase (CHS) in xylem of black locust. They found that PAL was active in the youngest wood near the cambium in April and September but that it was active in the transition zone in all seasons, while CHS was active only in the intermediate wood. From these results they suggested that PAL in the youngest woods is involved in lignin biosynthesis but in flavonoid synthesis in intermediate wood, while CHS is only involved in flavonoid synthesis. In addition, the stimulus for biosynthesis of robinetin, a flavonoid with strong antimicrobial activity, apparently occurs in the transition zone in *Instsia* species (Hillis, 1996).

There are three possible explanations for the increased enzyme activity in the transition zone of mostly dead cells (Hillis, 1987): 1) the enzymes are produced by the heartwood inhabiting microorganisms; 2) the enzymes are encoded by host parenchyma cells and remain active after the host cell death; or 3) the enzymes are made by living host parenchyma cells after heartwood formation. Shain and Mackay (1973b) provided indirect evidence indicating that two phenol-oxidizing enzymes were produced by host parenchyma cells and remain active after host cell necrosis.

Analysis of global gene expression pattern

Plant genomics changes and facilitates the way I acquire and utilize new knowledge on fundamental biological processes in plants (Walbot, 1999). DNA microarray technology provides a simple and economical way to explore gene expression patterns on a genomic scale (Brown and Botstein, 1999; DeRisi et al., 1996; Duggan et al., 1999; Heller et al., 1997; Schena et al., 1996, 1998). The use of DNA microarrays provide not only the global view of the changes, but also an unique opportunity to identify genes whose expression is altered during the formation of heartwood and its extractives. Furthermore, the genome-wide expression data obtained from DNA microarray hybridization can be analyzed using standard statistical algorithms to arrange genes according to similarity in gene expression pattern (Eisen et al., 1998). Such cluster analysis and display system facilitate the identification of transition zone-specific genes and the study of their expression patterns.

Black locust (Robinia pseudoacacia L.)

For this study, I choose to work with black locust (*Robinia pseudoacacia* L.). Although it is not the best species of economic importance, black locust represents one of the most extensively studied hardwood species with regard to heartwood formation. This multipurpose legume has great potential to become an economic species of choice, given biotechnological improvement can be made to alleviate a few inherent problems such as susceptibility to stem borer and stem crookedness. Black locust is the most favorable model species for the proposed research for several reasons: 1) it is a heartwood-forming species with clear distinction between distinct sapwood and heartwood. Poplar has been a popular model hardwood species for many forest biology studies. However, its utility as a

model species is limited for this project as poplar form less heartwood; 2) chemical and biochemical studies of heartwood formation have been extensively documented in black locust (Kamdem et al, 1996); 3) it produces and accumulates large amount of extractives in its heartwood, which renders characteristic durability of its wood. Black locust is currently being used as a model for the metabolic engineering study for decay resistance (Kamdem 1994); 4) black locust has relatively small genome size (637 Mbp) and enable genetic transformation (Han et al., 1999).

The ureid pathway

Allantoinase (allantoin amidohydrolase, EC 3.5.2.5), present in a wide variety of bacteria, fungi, and plants, as well as a few animals, such as fish and amphibians, catalyzes the hydrolysis of allantoin to allantoic acid in ureid metabolism (Noguchi et al., 1986, Vogels et al., 1966). Ureide metabolism serves different roles and is evolutionarily distinct in plants, animals, and microorganisms. The ureide pathway in animals functions primarily in salvage or excretion of nitrogen from purines (Campbell and Bishop, 1970; Stryer, 1988). In microorganisms, ureide degradation can provide nitrogen from a variety of sources in the external environment (Cooper, 1980). Ureides have been detected in various plants (Tracey 1961), and ureide metabolism has been studied on legume species (Lukaszewski et al., 1992; Bell and Webb, 1995). In addition, Allantoinase has been isolated from *Pseudomonas* (Jansenn et al., 1982), from mackerel and frog liver (Noguchi et al., 1986) and from soybean seed (Webb and Lindell, 1993), and the gene encoding allantoinase has been cloned from yeast (Buckholz and Cooper, 1991) and from bullfrog (Hayashi et al., 1994). Yet plant allantoinase gene has not been cloned.

My research hypothesis

1) Ray parenchyma cells at the sapwood-heartwood transition zone undergo a form of programmed cell daeth.

2) In the transition zone, reserve materials and phtotosynthesis products are converted to secondary metabolites, which may serve as defense chemicals.

3) Heartwood formation is orchestrated by a number of enzymes invloved in the breakdown of carbone sorces, secondary metabolite biosynthesis, and senescence.

4) In the transition zone, the genes encoding such enzymes are thought to be induced during the certain season.

5) Allantoinase gene involved in ureide pathway generally exists in plant kingdom.

6) Nitrogen regulates the expression of allantoinase genes.

Literature Cited

Baqui, S., Shah, J., Pandalai, R., and Kothari, I. (1979). Histochemical changes during transition from sapwood to heartwood in *Melia azedarach*. Indian J Exp Biol 17: 1032-1037.

Baqui, S., and Shah, J. (1985). Histoenzymatic studies in wood of *Acacia auriculiformis* Cunn. during heartwood formation. Holzforshchung 39: 311-320.

Bell, JA, and Webb, MA (1995). Immunoaffinity purification and comparison of allantoinases from soybean root nodules and cotyledons. Plant Physiol 107: 435-441.

Breeze, E., Wagstaff, C., Harrison, E., Bramke, I., Rogers, H., Stead, A., Thomas, B., and Buchanan-Wollaston, V. (2004). Gene expression patterns to define stages of post-harvest senescence in *Alstroemeria* petals. Plant Biotech J 2: 155-168.

Brown, P.O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nat Genet 21: 33-37.

Buckholz, R.G., and Cooper, T.G. (1991). The allantoinase (DAL1) gene of *Saccharomyces cerevisiae*. Yeast 7: 913-923.

Campbell, J.W., and Bishop, S.H. (1970). Nitrogen metabolism in molluscs. *In* JW Campbell, ed, Comparative Biochemistry of Nitrogen Metabolism, Vol 1. Academic Press, London, UK, pp 103-206.

Cercos, M., Santamaria, S., and Carbonell, J. (1999). Cloning and characterization of TPE4A, a thiol-protease gene induced during ovary senescence and seed germination in pea. Plant Physiol 119: 1341-1348.

Cooper T.G. (1980). Selective gene expression and intracellular compartment: two means of regulating nitrogen metabolism in yeast. Trands Biochem Sci 5: 332-334.

Datta, S., and Kumar, A. (1987). Histochemical studies of the transition from sapwood to heartwood in *Tectona grandis*. IAWA Bull 8: 363-368.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14: 457-460.

Dietrichs, H. (1964). Das Verhalten von Kohlenhydraten bei der. Holzforschung 18: 14-24.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. M. (1999). Expression profiling using cDNA microarrays. Nat Genet 21: 10-14.

Ebermann, R., and Stich, K. (1985). Distribution and seasonal variation of wood peroxydase activity in oak (*Quercus rubur*). Wood Fiber Sci 17: 391-396.

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

Han, K.-H., Gordon, M., and Keathley, D. (1999). Genetic transformation of black locust (*Robinia pseudoacacia* L.). *In* Biotechnology in agriculture and forestry (Y. Bajaj, ed.). Berlin Hiedelberg New York, Springer-Verlag, pp. 273-282.

Hayashi, S., Jain, S., Chu, R., Alvares, K., Xu, B., Erfurth, F., Usuda, N., Rao, M.S., Reddy, S.K., Noguchi, T., Reddy, J.K., and Yeldandi, A.Y. (1994). Amphibian allantoinase. Molecular cloning, tissue distribution, and functional expression. J Biol Chem 269: 12269-12276.

Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D. E., and Davis, R. W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci USA 94: 2150-2155.

Higuchi, T., Onda, Y., and Fujimoto, Y. (1969). Biochemical aspects of heartwood formation with special reference to the site of biogenesis of heartwood compounds. Wood Res 48: 15-30.

Hillinger, C., Holl, W., and Ziegler, H. (1996a). Lipids and lipolytic enzymes in the trunkwood of *Robinia pseudoacacia* L. during heartwood formation. I. Radial distribution of lipid classes. Trees 10: 366-375.

Hillinger, C., Holl, W., and Ziegler, H. (1996b). Lipids and lipolytic enzymes in the trunkwood of *Robinia pseudoacacia* L. during heartwood formation. II. Radial distribution of lipases and phospholipases. Trees 10: 376-381.

Hillis, W. (1987). Heartwood and tree exudates. Berlin, New York, Springer-Verlag.

Hillis, W., and Hasegawa, M. (1963). The formation of polyphenols in trees. I. Administration of 14C-glucose and subsequent distribution of radioactivity. Phytochem 2: 195-199.

Hillis, W., and Yazaki, Y. (1973). Wood polyphenols of Eucalyptus polyanthemos. Phytochem 12: 2969-2977.

Hillis, W. E. (1996). Formation of robinetin crystals in vessels of *Intsia* species. IAWA J 17: 405-419.

Höll, W., and Lendzian, K. (1973). Respiration in the sapwood and heartwood of *Robinia pseudoacacia*. Can J Bot 52: 727-734.

Janssen, D.B., Smits, R.A., and van der Drift, C. (1982). Allantoinase from *Pseudomonas aeruginosa*. Purification, properties and immunochemical characterization of its in vivo inactivation. Biochim Biophys Acta, 718: 212-219.

Kamdem, D., Dawson, A., and Wanli, M. (1996). Polyphenols content of black locust heartwood by HPLC. In Forest Products Society meeting. Chaleston, SC.

Kamdem, D. P. (1994). Fungal decay resistance of aspen blocks treated with heartwood extracts. For Prod J 44: 30-32.

Lukaszewski, K.M., Blevins, D.G., and Randall, D.D. (1992). Asparagine and boric acid cause allantoate accumulation in soybean leaves by inhibiting manganese-dependent Plant Physiol. 99: 1670-1676.

Magel, E., Drouet, A., Claudot, A., and Ziegler, H. (1991). Formation of heartwood substances in the stemwood of *Robinia pseudoacacia* L. I. Distribution of phenylalanine ammonium-lyase and chalcone synthase across the trunk. Trees 5: 203-207.

Magel, E., Jay-Allemand, C., and Ziegler, H. (1994). Formation of heartwood substances in the stemwood of Robinia pseudoacacia L.: II. Distribution of nonstructural carbohydrates and wood extractives across the trunk. Trees 8: 165-171.

Magel, E., and Hübner, B. (1997). Distribution of phenylalanine ammonia lyase and chalcone synthase within trunks of *Robinia pseudoacacia* L. Bot Acta 110: 314-322.

Matamoros, M. A., Baird, L. M., Escuredo, P. R., Dalton, D. A., Minchin, F. R., Iturbe-Ormaetxe, I., Rubio, M. C., Moran, J. F., Gordon, A. J., and Becana, M. (1999). Stress-induced legume root nodule senescence. Physiological, biochemical, and structural alterations. Plant Physiol 121: 97-112.

Mauseth, J. (1998). Botany: an introduction to plant biology. Jones and Bartlett Publishers. Sudbury, Massachusetts.

Miller, J. D., Arteca, R. N., and Pell, E. J. (1999). Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. Plant Physiol 120: 1015-1024.

Nair, M. (1988). Wood anatomy and heartwood formation in Neem (*Azadirachata indica* A. Juss.). Bot J Linn Soc 97: 79-90.

Nelson, N. (1978). Xylem ethylene, phenol-oxidising enzymes and nitrogen and heartwood formation in walnut and cherry. Can J Bot 56: 626-634.

Nelson, N., Rietveld, W.J., and Isebrands, J. (1981). Xyleme ethylene production in five black walnut families in the early stages of heartwood formation. For Sci 27: 537-543.

Nobuchi, T., Matsuno, H., and Harada, H. (1984). Relationship between heartwood phenols and cytological structure in the transition zone from sapwood to heartwood of sugi (*Cryptomeria japonica*). In Pacific Regional Wood Anatomy Conference (IAWA/IUFRO), Tsukuba, Japan, pp. 132-134.

Nobuchi, T., Sato, T., Iwata, R., and Harada, H. (1984). Season of heartwood formation and the related cytological structure of ray parenchyma cells in *Robinia pseudoacacia*. Mokuzai Gakkaishi 30: 628-636.

Noguchi, T., Fujiwara, S., and Hayashi, S. (1986). Evolution of allantoinase and allantoicase involved in urate degradation in liver peroxisomes. A rapid purification of amphibian allantoinase and allantoicase complex, its subunit locations of the 2 enzymes, and its comparison with fish allantoinase and allantoicase. J Biol Chem 261: 4221-4223.

Nobuchi, T., Tokuchi, N., and Harada, H. (1987). Variability of heartwood formation and cytological features in broadleaved trees. Mokuzai Gakkaishi 33: 596-604.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93: 10614-10619.

Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E., and Davis, R. W. (1998). Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol 16: 301-306.

Shain, J., and Mackay, J. (1973a). Seasonal fluctuations in respiration of aging xylem in relation to heartwood formation in *Pinus radiata*. Can J Bot 51: 737-741.

Shain, L., and Mackay, J. (1973b). Phenol-oxidizing enzymes in the heartwood of *Pinus radiata*. For Sci 19: 153-155.

Stryer, L. (1988). Biochemistry, Ed 3. WH Freeman, New York, pp 618-623

Tracey, M.V. (1961). Urea and Ureides. *In* K Peach K, M Tracey, eds, Modern Methods of Plant Analysis, Vol 4. Springer Verlag, Heidelberg, pp 119-141

Vogels, G.D., Trijbels, F., and Uffink, A. (1966). Allantoinases from bacterial, plant and animal sources. I. Purification and enzymatic properties. Biochim Biophys Acta 122: 482-496

Walbot, V. (1999). Genes, genomes, genomics. What can plant biologists expect from the 1998 national science foundation plant genome research program? Plant Physiol 119: 1151-1156.

Wardrop, A., and Cronshaw, J. (1962). Formation of phenolic substances in ray paranchyma of angiosperms. Nature 193: 90-92.

Webb, M.A., and Lindell, J.S. (1993). Purification of allantoinase from soybean seeds and production and characterization of anti-allantoinase antibodies. *Plant Physiol.* 103: 1235-41.

Wingler, A., von Schaewen, A., Leegood, R., Lea, P., and Quick, W. (1998). Regulation of leaf senescence by cytokinin, sugars, and light. Plant Physiol 116: 329-335.

CHAPTER 1

Novel gene expression profiles define the metabolic and physiological processes characteristic of wood and its extractive formation in a hardwood tree species, *Robinia pseudoacacia*.

ABSTRACT

Wood is of critical importance to humans as a primary feedstock for biofuel, fiber, solid wood products, and various natural compounds including pharmaceuticals. The trunk wood of most tree species has two distinctly different regions: sapwood and heartwood. In addition to the major constituents, wood contains extraneous chemicals that can be removed by extraction with various solvents. The composition and the content of the extractives vary depending on such factors as, species, growth conditions, and time of year when the tree is cut. Despite the great commercial and keen scientific interest, little is known about the tree-specific biology of the formation of heartwood and its extractives. In order to gain insight on the molecular regulations of heartwood and its extractive formation, I carried out global examination of gene expression profiles across the trunk wood of black locust (*Robinia pseudoacacia* L.) trees. Of the 2,915 expressed sequenced tags (ESTs) that were generated and analyzed in the current study, 55.3 % showed no match to known sequences. Cluster analysis of the ESTs identified a total of 2,278 unigene sets, which were used to construct cDNA microarrays. Microarray hybridization analyses were then performed to survey the changes in gene expression profiles of trunk wood. The gene expression profiles of wood formation differ according to the region of trunk wood sampled, with highly expressed genes defining the metabolic and physiological processes characteristic of each region. For example, the gene

encoding sugar transport had the highest expression in the sapwood, while the structural genes for flavonoid biosynthesis were up-regulated in the sapwood-heartwood transition zone. This analysis also established the expression patterns of 341 previously unknown genes.

Introduction

Wood is a unique renewable material produced by trees using solar energy through a highly ordered developmental process involving cell division/expansion, secondary cell wall synthesis/deposition, lignification, programmed cell death, and heartwood formation (Fukuda, 1996). As a result of radial growth and differentiation, the trunk wood of many tree species has two distinctly different regions: sapwood and heartwood. Sapwood is the outermost portion of the xylem tissue and contains living cells, whereas the heartwood is defined as the "dead" central core of the woody axis and only provides passive support to the tree. Sapwood (young xylem) has three important functions: to conduct sap (water, solutes, and gases) from the roots to all parts of the tree; to provide structural support for the entire tree; and to serve as a reservoir for water, energy, minerals, and solutes. On average, about 10% of the cells in the sapwood are alive (Kozlowski and Pallardy, 1997). The living ray cells in sapwood serve as the source of raw materials for secondary substances. The ray parenchyma may also serve as communication channels radially from the cambium through the sapwood, while axial parenchyma functions largely as a storage tissue.

As sapwood is gradually converted to inactive heartwood, the wood parenchyma cells undergo numerous metabolic changes and produce large quantities of heartwood

extractives such as phenolic compounds, lignin, and aromatic substances that accumulate in the vessels (Magel, 2000). During that process, one annual ring is converted to heartwood each year (Mauseth, 1998). The reserve materials in the parenchyma cells of the sapwood are used for wood formation and the synthesis of heartwood extractives, such as condensed tannins, terpenes, flavonoids, lignans, stilbenes, and tropolones (Burtin et al., 1998; Hillinger et al., 1996a and b; Hillis, 1987; Magel et al., 1994 and 2000). The formation of heartwood is accompanied by a variety of alterations in metabolic conditions such as senescence. Although the events of senescence have been studied at the molecular level during leaf senescence (Miller et al., 1999; Wingler et al., 1998), seed germination (Cercos et al., 1999), and nodule development (Matamoros et al. 1999); the cell maturation and death events occurring during heartwood formation have been difficult to study because of the location and timing of the events. The presence of heartwood is a major determining factor for wood quality and influences the way in which specific woods are utilized. Various wood properties, such as dimensional stability, durability, pulpability, colors and hues, scents and beauty, are affected by extractives. Furthermore, stem wood sequesters large amounts of atmospheric CO2 into a much slower turnover pool and consequently accounts for the largest proportion (20-40%) of total ecosystem aboveground carbon in closed forests (Saxe et al., 1998). Therefore, understanding the regulation of wood formation is of great commercial and keen scientific interest.

In recent years, a genomics approach has been successfully used to examine global gene expression patterns in developing xylem tissues of pine (Allona *et al.*, 1998; Lorenz and Dean, 2002) and poplar (Sterky *et al.*, 1998; Hertzberg *et al.*, 2001).

Although the information derived from those studies undoubtedly provided a powerful means for studying the molecular mechanisms of this important differentiation pathway, it is still insufficient to account for the complete process of wood formation. So far, there has been no report on global examination of gene expression profiles inside trunk wood of mature trees. In order to gain some insight into the transcriptional hierarchy of heartwood and its extractive formation, I examined global gene expression profiles across the stems of 10-year-old *Robinia pseudoacacia* trees by sampling bark, sapwood, and sapwood-to-heartwood transition zone tissues. This report describes the first comprehensive look at global gene expression profiles in trunk wood and provides expression data for many genes of unknown function.

Materials and Methods

RNA isolation and cDNA library construction

Three cDNA libraries of bark/cambial region (BCS), sapwood region (SWS), and transition zone (TZS) of trunk-disk of 10-year-old black locusts (*Robinia pseudoacacia* L.) harvested in early summer (July 27, designated "S") and one cDNA library (TZF) from a black locust harvested at late fall (November 27, designated "F") were constructed using the SMART cDNA library construction kit (λTriplEx2 vector system, Clontech, Palo Alto, CA). Mature trees (20 cm DBH) were fell using a chain saw and made into 25 cm-long logs. The logs were immediately placed on ice and brought back to a wood shop, where thin cookies (~1 cm thick) were made using a table saw. The cookies were immediately submerged in RNA extraction buffer (20 mM EDTA, pH 8.0; 50 mM Tris-HCl, pH 8.0; 0.2 % SDS; 10 mM β-mercaptoethanol). While submerged, the trunk wood

sections (bark/cambial region, sapwood, and transition zone) were carved out by using chisel and hammer, and washed with RNase Away solution (Invitrogen, Carlsbad, CA). The isolated sections (~1 cm^3 cubicles) were frozen in liquid nitrogen and stored at –80°C until needed. For RNA isolation, the frozen samples were first ground in a blender and then further ground to fine powder using mortar and pestle. The ground sample were first passed through the shredder column of DNeasy Maxi kit, and then subjected to total RNA isolation using the RNeasy Maxi kit (Qiagen, Hilden, Germany) and cleaned up by Qiagen RNeasy Mini kit.

Nucleotide sequencing

The cDNAs that were directionally cloned were randomly picked and sequenced to generate ESTs. The sequencing was carried out at the Center for Computational Genomics and Bioinformatics at the University of Minnesota and at the Genomics Technology Supporting Facility at Michigan State University. The sequencing results are posted at the website (http://web.ahc.umn.edu/biodata/blacklocust/).

Sequencing analysis

Raw sequence files were produced from the trace files using the Phred trace-processing program followed by the Phran base-calling program with a quality threshold of 8-10 (Ewing *et al.* 1998). Sequence artifacts were trimmed by the removal of leading and trailing vector sequences in the raw sequence. To obtain the best subsequence where the "N" value is 4% or less of the total number of bases, the number of unknown or "N" bases in a sequence of trimming, leading and trailing high-N sections was determined.

Sequence similarity analyses were completed using a number of database searches. Databases include GenBank, National Center for Biotechnology Information (NCBI) GenPept (Benson *et al.*, 2000), Protein Information Resource (PIR) (Barker *et al.* 2000), Swiss Institute of Bioinformatics SWISS-PROT (Bairoch *et al.*, 2000), TrEMBL (Bairoch *et al.*, 2000) and the National Biomedical Research Foundation NRL3D (Barker *et al.*, 2000). The EST sequences were deposited in the dbEST of the GenBank database.

Contig analysis

In order to generate the unigene sets and contigs of sequencing data, EST data was analyzed using DNA similarity algorithms and the assembly program, Phred/Phrap/Consed (Green and Ewing, 1996). Phrap ("phragment assembly program", or "phil's revised assembly program") was used for assembling shotgun DNA sequence data, constructing a contig sequence as a mosaic of the highest quality parts of reads (rather than a consensus) and providing extensive information about assembly (including quality values for contig sequence) to assist trouble-shooting. Phrap was used in conjunction with the base calls and base quality values produced by the basecaller, Phred; and with the sequence editor/assembly viewer, Consed. Cross-match was based on a "banded" version of SWAT, an efficient implementation of the Smith-Waterman algorithm for comparing any two sets of (long or short) DNA sequences. To calculate estimates for comparison among different libraries, I used a statistical method described by Audic and Claverie (1997). This method was developed to calculate statistical differences from different numbers of different libraries, and now it is being used for "digital gene expression profiles." Based on the method, I calculated significantly

different numbers within the same contig among the libraries, and I applied it to estimate statistical differences within a category as described in Kirst et al., (2002).

PCR amplification of the insert cDNA and microarray printing

λTriplEx2 vector sequences flanking the insert (5'-AAGCAGTGGTATCAACGCAGAGT-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-3') were used to amplify selected EST clones using polymerase chain reaction (PCR). The PCR products were precipitated in ethanol and re-suspended in 3 x SSC (1 x SSC is 0.15 M NaCl and 0.015 M sodium citrate). They were checked for quality by using gel electrophoresis to observe the concentration and multiple bands. PCR products of 2,592 clones were arrayed from 384-well microtiter plates, and DNA was spotted on superaldehyde (Telechem, Sunnydale, CA) glass slides at a high density using an Omnigridder robot (Gene Machines, San Carlos, CA) and 16 ArrayIt chipmaker 2 pins (Telechem). Slides were washed and blocked according to the manufacturer's protocol.

Each glass slide contained two replications of the entire array, each of which consisted of 16 subarrays with 12 rows and 14 columns. Negative control genes, B-cell receptor protein genes, including genes such as Myosin heavy chain gene, Myosin regulation light chain2 and insulin-like growth factor gene, were printed on the top, middle and bottom of each array.

Preparation of labeled cDNA probes

Total RNA (1μg from each sample) was reverse-transcribed and amplified using

the SMAT system (Clontech). To reduce nonspecific PCR amplification, cDNAs were amplified with the fewest cycles. Two micrograms of cDNA were labeled by the incorporation of either Cy5 or Cy3-dCTP (Amersham-Phamacia, Piscataway, NJ) during oligo-dT-primed primer extension in the presence of Klneow DNA polymerase (Promega, Madison, WI) as described by Schaffer *et al.* (1999). The labeled probes were purified using the QiaQuick PCR cleanup kit (Qiagen). The probe samples were denatured by placing them in a 100°C water bath for 3 min, left at room temperature for 30 min, and then used for hybridization.

To minimize the inherent variability of the microarray assay (Lee *et al.*, 2000) and to ensure the reliability of the results, at least two microarray slides (four replicates) were used to analyze the transcript expression of each sample pair. The first slide was probed with cDNAs labeled with Cy-3 and Cy-5 deoxy CTP. To probe the arrays, cDNAs were synthesized and amplified from bark/cambial region, sapwood, or transition zone and labeled by Klenow-mediated incorporation of Cy-3-dCTP or Cy-5-dCTP, respectively. By using independent RNA preparations, the second slide was hybridized by cDNAs reverse labeled with Cy-3 and Cy-5 dCTP from each sample pair to overcome potential artifacts caused by the dye-related differences in labeling efficiency, different laser settings, and nonlinearity of photomultiplier tubes in the scanner. Thus, at least two, and sometimes three or four, independent RNA preparations were made for each biological sample and were used to prepare labeled probes. The hybridization signal from each of the replicate ESTs were averaged and used for analysis.

Hybridization and washing of the DNA microarray

The labeled probes with either Cy3 or Cy5 fluorescent dye were hybridized to a microarray slide in a total volume of 30 μL of hybridization buffer (3.4 x SSC, 0.32% SDS, and 5 μg of yeast tRNA) for 16 hours at 65°C. The slide was then washed at room temperature in 1 x SSC, 0.1% SDS for 10 min, in 1 x SSC for 10 min, and in 0.01 x SSC for 10 min. The slide was centrifuged dry and scanned with a 428 Array scanner (Affymetrix, Palo Alto, CA). Each microarray experiment was repeated twice.

Microarray Data Analysis

The data were analyzed with GenePix Pro3.0 (Axon Instruments Inc., Union City, CA). The scanned data were normalized by using the Global Normalization method (Hihara *et al.*, 2001), in which the image data between Cy3 and Cy5 channels are normalized by adjusting the total signal intensities of two images and the bad spots are removed. The unreliable spots were removed by the following screening. Spots containing clones that had poor amplification or multiple bands, as well as those that were flagged due to a false intensity caused by dust or background on the array, were removed. Spots with <65% of the spot intensity at >1.5-fold that of the background in both channels were ignored (see Stanford Microarray Database Web site, http://genome-www5.stanford.edu/ MicroArray/ SMD/). Clones in one sample that had an average induction greater than 2-fold in another were determined as up-regulated. Comparison of the arrays was achieved using Microsoft Excel and Microsoft Access database. For cluster analysis, Cluster and Treeview software were used (Eisen *et al.*, 1998; available at http://genome-www4.stanford.edu/ MicroArray/SMD/restech.html).

Antisense northern blot analysis

I conducted "Antisense Northern blot analysis" which requires only minute amounts of RNA. Conventional Northern blot analysis protocol requires a large amount of RNA from my inner wood samples, typically transition zone, which are hard to isolate, making those protocols difficult to perform. An alternative was to use antisense RNA (aRNA) amplification method that has been successfully used in other microarray analyses (Wang *et al.*, 2000; Baugh *et al.*, 2001; Dent *et al.*, 2001; for protocol, see Patrick Brown's web site; http://cmgm.stanford.edu/ pbrown/ protocols/ ampprotocol_3.html). Antisense RNA was generated as described previously (Wang *et al.*, 2000). Briefly, aRNA was amplified using Message Amp™ aRNA kit (Ambion, Austin, TX). I began by synthesizing first stand cDNAs from the total RNAs of seedling, bark/cambial region, sapwood, or transition zone, and the first cDNAs were used as templates for the synthesis of second cDNAs. Finally, aRNAs from the second cDNAs were generated by *in vitro* transcription (amplification). About 200ng of aRNAs were separated in a formamide agarose gel, and transferred onto the nylon membrane using the capillary transfer method. The membrane was then hybridized with an isotope-labeled probe. The signal was exposed and detected on an X-ray film.

Results

Trunk Wood cDNA Library Construction and Sequencing

Trunk wood of mature trees can be divided into three main parts: bark/cambial region, sapwood, and heartwood (Figure 1-1). While sapwood contains living ray cells, the cells in heartwood are dead and filled with extractives, which produce intense and

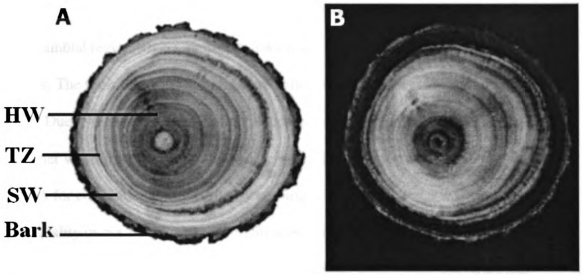


Figure 1-1. Cross-section of a stem from a mature *Robinia pseudoacacia* tree
(A) Cross section under daylight. HW, Heartwood; SW, sapwood; TZ, transition zone. (B) Cross section under UV light. Bright fluorescence in the middle is due to the flavonoids.

bright fluorescence under UV light. In order to analyze the gene expression patterns in different sections of the trunk wood, I constructed four cDNA libraries from the bark/cambial region (BCS), sapwood (SWS), and transition zone of ten-year-old black locusts. The transition zone samples were collected both in the summer and fall (TZS and TZF). Due to the large amounts of polysaccharides and phenolic compounds present in the inner wood tissues, it was not possible to obtain a large enough quantity of pure mRNA for conventional cDNA library construction. Nonetheless, I was able to construct high quality (> 5 x 10^5 pfu) phagemid libraries using the PCR-based cDNA library construction kit (Clontech). The phagemids from each library were converted into plasmids by mass excision in *E. coli.* Over 3,600 individual clones were randomly selected from all four libraries and sequenced using a 5' vector sequencing primer provided in the kit. After trimming vector sequences, clones containing high ambiguous calls (high "N" percent on a DNA sequence) were removed. Finally, a total of 2,915 were chosen for further analyses and sequencing: 895 clones from the BCS library, 999 clones from the SWS library, 880 clones from the TZS library and 141 clones from the TZF library. An average length of 448 bases was obtained and used for contig analysis and database searches.

EST Analysis

The high redundancy of the mRNA in a tissue is approximately reflected in the abundance of its corresponding cDNA in non-normalized libraries. The random sequencing of cDNAs yields information about the ESTs (Adams *et al.*, 1993). Sequence similarities were found by searching various available databases including: GenBank,

National Center for Biotechnology Information (NCBI) GenPept, Protein Information Resource (PIR), Swiss Institute of Bioinformatics SWISS-PROT, TrEMBL and the National Biomedical Research Foundation NRL3D (see the website; http://web.ahc.umn.edu/biodata/blacklocust/). Sequence similarities identified by the BLAST program were considered statistically significant with a Poisson *P* value of $\leq 10^{-5}$. The 1,304 ESTs (44.7%) of the total 2,915 ESTs matched previously sequenced genes. The 909 ESTs of the 1,304 ESTs had significant homology to previously identified genes. The annotations of genes with similarities to an EST were used to assign a putative identification to my EST. The 909 ESTs with similarity to known genes were classified into 13 putative functional categories (Bevan *et al.*, 1998, Covitz *et al.*, 1998), which are listed in Table 1-1. My libraries have a significantly high number of no hit clones, especially in the SWS library (80.6%). As a matter of fact, the portion of no hit clones in other libraries was also high, about 50 %. Allona *et al.* (1998) suggested that the length and quality of cDNA sequences are correlated with the ability to identify similar sequences in public databases. Recent analysis of *Pinus taeda* cDNA clones with longer than 1,000 bases-read revealed that about 95% of the pine genes had homologous sequences in *Arabidopsis* genome (Sederoff *et al.*, 2002). However, the average sequenced length of a SWS clone is not entirely different from the total sequenced length (sequence length distribution data is found at http://web.ahc.umn.edu/biodata/blacklocust/). Many long, high-quality sequences show neither strong nor marginal similarity to sequences in the database, and some no hit clones also show high redundancy (Table 1-2). Therefore, these no-hit clones may represent novel plant genes, reflecting the uniqueness of my samples.

Table 1-1. The functional classification of EST clones

Category	Total	BCS	SWS	TZS	TZF
Cell division and cycle	7	4	0	3	0
Cell wall structure and metabolism	16	5	7	3	1
Chromatin and DNA metabolism	27	8	4	12	3
Cytoskeleton	10	6	3	0	1
Defense	90	46	**5**	35	4
Gene expression and RNA metabolism	76	**38**	20	16	2
Membrane transport	48	33	1	11	3
Miscellaneous	175	73	36	51	15
Primary metabolism	134	46	19	56	13
Protein synthesis and processing	190	84	19	80	7
Secondary and hormone metabolism	55	8	2	**44**	1
Signal transduction	75	26	**11**	32	6
Vesicular trafficking, protein sort, secret	6	2	2	2	0
Unknown, Hypothetical	395	162	65	140	28
Hit clones	1304 (44.7)	541 (60.4)	194 (49.4)	485 (55.1)	84 (59.6)
No hit clones	1611 (55.3)	354 (39.6)	805 (80.6)	395 (44.9)	57 (40.4)
Total sequenced clones	2915	895	999	880	141

The bold number within a row indicates significantly different number of ESTs in the library compared to all the others ($P < 0.01$). For example, the number of ESTs in "Gene expression and RNA metabolism" category was statistically higher in BCS when compared to SWS and TZS libraries. Likewise, the number of ESTs in "Secondary and hormone metabolism" category was significantly higher in TZS than in BCS and SWS.

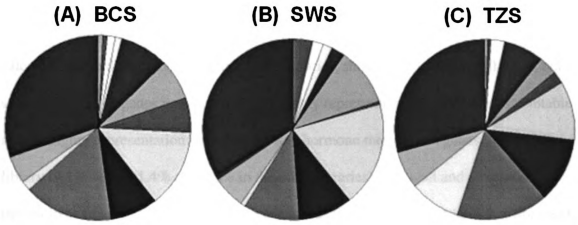


Figure 1-2. Functional classification of the ESTS from three libraries
Due to the high proportion of no-hit clones in the libraries, only those ESTs with significant homology with previously reported sequences were included in this classification.

- Cell division and cycle
- Cell wall structure and metabolism
- Chromatin and DNA metabolism
- Cytoskeleton
- Defense
- Gene expression and RNA metabolism
- Membrane transport
- Miscellaneous
- Primary metabolism
- Protein synthase and processing
- Secondary and hormone metabolism
- Signal transduction
- Vesicular trafficking, protein sort, secret
- Unknown, Hypothetical

The proportions of ESTs in each functional category differed for the EST libraries from the three trunk zones characterized in this study (Figure 1-2.) These data, which were calculated by excluding no-hit clones in all libraries, show a common overall trend, but specific groups of genes were more or less highly represented in specific zones. Notable are the higher representation of secondary and hormone metabolism genes in the TZS library (9.1% versus 1.4% and 0.8% in the other libraries), cell wall and structural metabolism genes in the SWS library (3.6% versus 0.6% and 0.9% in the other libraries), and genes associated with membrane transport in the BCS library (6.1% versus 1.2% and 2.3% in the other libraries).

Contig Analysis

It was attempted to estimate the redundancy of my EST clones on the basis of the contig analysis (Phrap assembly with the 70 % identical value of pairwise sequence). The proportion of singletons was extremely high, representing a calculated level of singletons at 70 % (2,051 out of 2,915 ESTs). Only 864 ESTs were identified in 228 contig sets. As a result of the contig analysis, I obtained a total of 2,278 unigene sets that were submitted to GenBank database (accession numbers BI642054 to BI679372). The most frequently presented gene in my ESTs encodes a hypothetical protein (At2g41250; 1.3 %) followed by an auxin-repressed protein (1.1%) that is expressed in dormant tissues and repressed by auxin treatment (Reddy and Poovaiah, 1990; Stafstrom *et al.*, 1998). It is abundant in all libraries except the sapwood library (Table 1-2). It is notable that the TZF library (2.1%) showed a similar proportion when compared to the TZS library (1.5%). The function of this gene has not yet been identified. The contig analysis identified three

types of metallothionein or metallothionein-like protein genes with different expression patterns in the different zones of the trunk wood. For example, metallothionein (contig 226) is the highest in the TZS library, but metallothionein-like protein (contig 223) is only in the BCS library. Such differential expression of metallothionein genes has been reported in *Arabidopsis* plants (García-Hernández *et al.*, 1998). In addition to providing physical support, trunk wood functions as a conduit for water, nutrient, and photosynthates transport. Aquaporin and phloem-specific protein, which are related to a membrane transport system or phloem, are abundant mainly in the bark/cambial region library. Two types of aquaporin genes that are related to water transport were found in my libraries Aquaporin 1 contig existed in the bark/cambial region library only, while aquaporin 2 contig was present in the transition zone library as well as bark/cambial region library. However, no contig for aquaporin genes were present in the sapwood library. Because the sole abundance of transcript encoding phloem-specific protein exists only in the bark/cambial region library, it is clear that my bark/cambial region sample contained phloem regions. Eight out of the 20 contigs in Table 1-2 were library-specific. In other words, the contigs were present only in one library and absent in the other two libraries. For example, maturase and hypothetical protein (PIR: A05191) are abundant only in the SWS library. Hypothetical protein (At3g03150) and cytochrome b5 DIF-F are highly abundant in the TZS library. Some contig sets contained only no hit clones resulting from the source zone of the samples; for example contig 224 and 216 consist of clones from two transition zone libraries, and clones of contig 213 are in the BCS library. These library specific contigs can be expected because the typical character of each library is dependent upon its location within the trunk wood.

Table 1-2. The redundancy of EST clones based on contig analysis

Annotation	Contig number	Total (%)	BCS (%)	SWS (%)	TZS (%)	TZF (%)
Hypothetical protein (At2g41250)	228	37 (1.3)	27 (3.0)	0	8 (0.9)	2 (1.4)
Auxin-repressed protein	227	33 (1.1)	17 (1.9)	0	13 (1.5)	3 (2.1)
Metallothionein	226	18 (0.6)	2 (0.2)	1 (0.1)	14 (1.6)	1 (0.7)
Maturase	225	13 (0.4)	0	9 (0.9)	4 (0.5)	0
No hit	224	11 (0.4)	0	0	10 (1.1)	1 (0.7)
Metallothionein-like protein	223	11 (0.4)	**11** (1.2)	0	0	0
Hypothetical protein (At3g03150)	222	11 (0.4)	0	0	**11** (1.3)	0
Cytochrome B5 DIF-F	221	10 (0.4)	0	0	9 (1.0)	1 (0.7)
Aquaporin 1	220	9 (0.3)	**9** (1.0)	0	0	0
Aquaporin 2	219	8 (0.3)	4 (0.4)	0	3 (0.3)	1 (0.7)
Hypothetical protein (A05191)	218	8 (0.3)	0	**8** (0.8)	0	0
Phloem-specific protein Vein1	217	8 (0.3)	**8** (0.9)	0	0	0
No hit	216	7 (0.2)	0	0	7 (0.8)	0
Metallothionein Class-II	215	7 (0.2)	2 (0.2)	0	3 (0.3)	2 (1.4)
Ubiquitin	214	7 (0.2)	6 (0.7)	0	1 (0.1)	0
No hit	213	7 (0.2)	7 (0.8)	0	0	0
No hit	212	7 (0.2)	6 (0.7)	0	0	1 (0.7)
Integral membrane transport protein	211	6 (0.2)	2 (0.2)	0	4 (0.5)	0
Hypothetical protein (D75542)	210	6 (0.2)	5 (0.6)	0	1 (0.1)	0
Extensin	83	2	0	0	0	**2** (1.4)

Letters in bold are significantly different from the numbers in the other libraries at $P < 0.05$. For example, Metallothionein-like protein (contig 223) and Aquaporin 1 (contig 220) are significantly more abundant in BCS zone than SWS and TZS. SWS library had significantly higher number of ESTs in the contig 218 (hypothetical protein, PIRA05191) than did BCS and TZS libraries. Contigs 222 (hypothetical protein At3g03150) and 216 (no hit clone) are significantly abundant in TZS compared to BCS and SWS libraries.

Microarray Experiments

In order to produce gene expression profiles related to wood formation and to compare gene expression patterns from different regions of the inner wood from a mature tree, I produced cDNA microarrays carrying ~2,580 unigenes from all four libraries and conducted microarray hybridization experiments. This approach allowed me to examine the expression changes of ~2,580 genes simultaneously and to search the expression patterns of bark/cambial region, sapwood, and transition zone through the use of one-by-one comparisons. I demonstrated that one-by-one experiments can be carried out for specific expressed profiling on continuous samples. My approach involved the comparison of one sample with two other samples and then the generated ratios from the two different experimental sets were plotted. When all of the spots were plotted, I could determine which genes were specifically expressed in Sample A, when compared to Sample B and Sample C. Unlike a reference or loop design, an approach of this type will serve as a one-by-one comparison for a small number of side-by-side samples in transcript profiling studies.

In my microarray experiments, the need to obtain large amounts of RNAs proved to be a challenge. Standard microarray protocols require isolating poly(A) RNA from samples, but my wood sample is too difficult to isolate RNAs. So, the cDNA amplification method was used, and I checked the reproducibility of my experiments. I also confirmed that the amplification method efficiently generates highly reproducible populations of cDNA. This method can be useful in transcript profiling studies with limited amounts of RNA. To test the reproducibility of the two different experiments, the expression ratio derived from one microarray experiment was compared to the expression

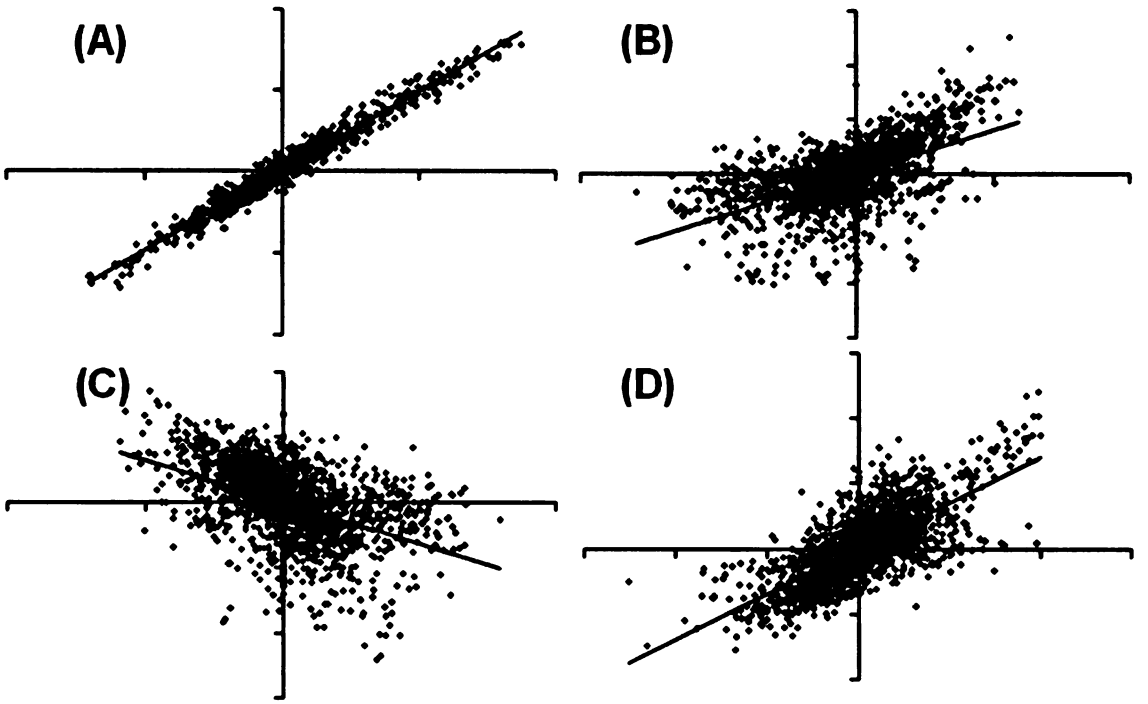


Figure 1-3. Scatter plots of microarray hybridization results
(A) The reproducibility of microarray experiments. Two hybridization experiments of SWS versus BCS were conducted using a dye-swap with Cy5 and Cy3 labeled probes. The natural log values of the Cy5-to-Cy3 ratios were plotted for the two replicates.
(B) The BCS-specific expression pattern. The X-axis is the log-scaled ratio of gene expression for the experiment of BCS over SWS, and Y-axis is the log-scaled ratio of gene expression for the experiment of BCS over TZS.
(C) The SWS-specific expression pattern. X-axis is the log-scaled ratio of SWS over BCS, and Y-axis is the log-scaled ratio of SWS over TZS.
(D) The TZS-specific expression pattern. X-axis is the log-scaled ratio of TZS over BCS, and Y axis is the log-scaled ratio of TZS over SWS. All number values of the log-scaled ratio of gene expression are average values of data points generated from each experiment. Abbreviations: BC, bark/cambial zone; SW, sapwood; TZ, transition zone.

ratio from the other, for all of the normalized clones (Figure 1-3A). The scatter plot shows that the gene expression ratio from the two experiments was remarkably similar (the coefficient of determination $R^2 = 0.96$). Furthermore, the coefficients of determination between other experiments were sufficiently high (> 0.91). These results are similar to those presented in other papers, in which the same or similar amplification protocol was used (Livesey *et al.*, 2000, Hertzberg *et al.*, 2001). Furthermore, using "antisense northern blot" analysis, I confirmed the microarray data as well as the EST results. Thus, the microarray signals from the replicates in my study were highly reproducible and the conclusions derived from this analysis are considered reliable.

Differential Gene Expression across the Stem

In order to investigate distinct differences in gene expression profiles among the three regions of the trunk wood (bark/cambial zone, sapwood, and sapwood-heartwood transition zone) during active tree growth, I compared differentially expressed genes in the three zones. The expression ratios for each zone in comparison to the other two zones were examined using regression analysis. Figure 1-3B shows the scatter plot and regression line of the expression ratios for the bark/cambial region versus the sapwood or transition zone libraries. The trend line of the scatter plot had a slope of 0.401, with $R^2 = 0.2956$. These values indicate a positive correlation, but with little of the variation in the data explained by the regression. On the basis of that graph, I identified the genes that were up or down regulated in BCS tissue using the cutoff values of an expression ratio greater than two-fold or less than 0.5. Of all arrayed cDNAs, 292 clones were

Table 1-3. Summary of up-regulated or down-regulated genes

	Up-regulation			Down-regulation		
Functional category	BC	SW	TZ	BC	SW	TZ
Cell division and cycle	0	0	0	0	0	0
Cell wall structure and metabolism	0	0	0	1	0	0
Chromatin and DNA metabolism	0	0	1	1	0	0
Cytoskeleton	2	0	0	0	0	2
Defense	6	0	2	2	0	6
Gene expression and RNA metabolism	4	0	0	4	0	3
Membrane transport	7	1	0	0	0	5
Miscellaneous	16	0	1	3	0	16
Primary metabolism	8	1	2	1	0	5
Protein synthase and processing	2	0	1	0	0	5
Secondary and hormone metabolism	1	0	8	3	1	0
Signal transduction	1	0	2	1	0	0
Vesicular trafficking, protein sort, secret	1	0	0	0	0	0
Unknown, Hypothetical	22	0	20	7	3	14
No hit	26	1	37	40	0	12
Total	96	3	74	63	4	68

more highly expressed in the bark/cambial region than in sapwood. The expression of 174 clones in the bark/cambial region was higher than those in the transition zone. Ninety-six of the clones were considered as the BCS-specific clones (Tables 1-3 and 1-4). As expected, metallothionein-like protein (contig 223; containing all 11 BCS clones), phloem-specific protein genes (contig 217; containing all 8 BCS clones), and a no-hit clone in contig 213, which is a contig set composed of all 7 BCS clones, were highly expressed in bark/cambial region when compared to other zones. These results corroborate with those of EST redundancy analysis. In addition, genes encoding photosynthesis-related proteins, such as Photosystem II 10K proteins, were highly expressed in the bark/cambial region. In addition, 63 clones were specifically down regulated in bark/cambial region, when compared to sapwood and transition zone (Table 1-3). I compared each functional category to find out how many tissue-specific genes were in each category. Table 1-3 shows the proportion of functional categories of up or down regulated genes in bark/cambial region. Like the EST results, the proportion of the genes categorized as involving membrane transport was high relative to other functional categories in the bark/cambial region. When considering the physiological significance of the region in tree growth and development, the proteins encoded by the genes in each of these categories might be targets of special interest for biotechnological improvement of trees.

Similarly, the expression ratio of the SWS was compared to both the BCS and TZS (Fig 1-3c) and the number of up and down regulated genes was calculated. Unlike the trendline of the bark/cambial region and transition zone scatter plots; the scatter plot of sapwood had a negative slope (0.401), indicating a negative correlation. Many of the

clones that were highly expressed in sapwood vs. bark/cambial region were down regulated in the sapwood vs. transition zone comparison. For example, the expression ratio of the PR-10 gene in the sapwood region vs. bark/cambial region was 6.7, but the ratio in the sapwood region vs. the transition zone was 0.5. Similarly, the expression ratio of the phloem-specific protein gene in the sapwood region versus the bark/cambial region was 0.2. However, in contrast, the ratio in the sapwood region versus transition zone was 4.3, showing that the expression relationships vary for the different proteins in the three regions. In addition, the number of clones specifically up or down regulated in the sapwood region was very small (3 or 4), but when the expression in the sapwood region was compared to bark/cambial region or transition zone, the number of up or down regulated clones increased dramatically (395 up regulated for sapwood vs. bark/cambial region; 177 up regulated for sapwood vs. transition zone, and 291 down regulated for sapwood vs. bark/cambial region; 226 down regulated for sapwood vs. transition zone). This suggests that the sapwood region plays the role of a bridging zone between bark/cambial region and transitional zone. Table 1-5 shows the list of up-regulated genes in the sapwood compared to bark/cambial region and transition zone. In other words, only three ESTs (TZS0226, SWS0332, and SWS0562) had sapwood-specific up-regulation. Interestingly, the transcript of a sugar transport protein, which plays key roles in source-sink relationships (Lalonde *et al.*, 1999), was highly expressed in sapwood. The highly redundant EST (contig 218) in the sapwood library is also highly expressed in the sapwood when compared with the bark/cambial region and transition zone (at the ratio of 2.2 and 1.7, respectively).

Table 1-4. Up-regulated transcripts in the bark and cambial zone

Clone ID	GenBank Acc.	Annotation	BC/SW*	BC/TZ*
Cytoskeleton				
CLS0035	BI677465	Actin depolymerizing factor 5	3.2	3.9
CLS0977	BI678088	(SMC)-like	5.1	3.3
Defense				
CLS0100	BI677437	Superoxide Dismutase (Cu-Zn)	2.4	2.4
CLS0239	BI677578	Metallothionein-like protein	5.4	3.5
CLS0595	BI677862	Superoxide Dismutase (Cu-Zn) sodB	2	2.1
CLS0801	BI677956	Metallothionein	2.8	2.5
CLS0867	BI678223	Glyoxalase	3.2	4.1
CLS0960	BI678073	Proteinase Inhibitor	3.2	2.4
Gene expression and RNA metabolism				
CLS0340	BI677663	Transcriptional regulator, putative	2.8	2.3
CLS0439	BI677739	Zinc Finger Protein ID1	2.2	2.7
CLS0673	BI678286	Transcription Factor like Protein	4.3	4.4
SWS1456	BI679312	Zinc Finger Protein ID1	2.2	2.7
Membrane transport				
CLS0025	BI677455	Lectin precursor, Bark Agglutinin I	7.7	3.8
CLS0051	BI677477	Vacuolar V-H subunit E [Citrus limon]	2.2	2.2
CLS0052	BI677478	Aquaporin	7.5	3.8
CLS0197	BI677550	Lectin	2.1	2.9
CLS0488	BI677780	Tonoplast Intrinsic Protein, delta type	7	5.9
CLS0929	BI678048	Lectin-related polypeptide	4.7	3.3
CLS0950	BI678064	Lectin like protein (hypothetical)	3.3	2.4
Miscellaneous				
CLS0022	BI677452	Phloem-specific protein Vein1	5.3	10.1
CLS0068	BI677490	Chlorophyll a/b binding protein	4.6	2.5
CLS0081	BI677501	MTN5 Gene Precursor	3.5	6.7
CLS0158	BI677521	Phloem-specific protein Vein1	8.8	5.5
CLS0274	BI677608	Trypsin Inhibitor (Serine Proteinase Inhibitor)	3.1	2.9
CLS0376	BI677690	Photosystem II 10K protein	2.6	4.2
CLS0423	BI677726	Photosystem II Protein X precursor	8.1	4.9
CLS0479	BI677773	Phloem-specific protein Vein1	7.1	4.9
CLS0536	BI677820	Core protein	2.8	2.1
CLS0564	BI677840	Photosystem II 10K protein	3	2.1
CLS0603	BI677869	Auxin-repressed protein	3.1	3.1
CLS0805	BI677959	Photosystem I Reaction Centre Subunit VI	5	3.7
CLS0819	BI677971	Trypsin Inhibitor	3.5	2.6
CLS0824	BI677975	Magnesium Chelatase	4.1	3.3
CLS0952	BI678066	Specific Tissue Protein 2	6.3	2.7
CLS1025	BI678111	Ripening related protein	4.2	3

* The ratio was estimated as average value from data points.

Table 1-4. (Cont'd)

Clone ID	GenBank Acc.	Annotation	BC/SW*	BC/TZ*
Primary metabolism				
CLS0017	BI677447	Copper Amine Oxidase precursor	4.4	5.4
CLS0714	BI678307	Lipid Transfer Protein	4.2	3.6
CLS0745	BI677908	Alcohol Dehydrogenase 1	6.2	3.8
CLS0813	BI677966	Alcohol Dehydrogenase 7	3.6	3
CLS0930	BI678049	Acetoacyl-CoA-thiolase	2.3	2.9
CLS0956	BI678070	Blue copper protein	2.5	2.1
CLS1015	BI678039	Lipid Transfer Protein	4.5	2.9
TZS0305	BI642632	Epoxide Hydrolase	4.7	2.3
Protein synthesis and processing				
CLS0811	BI677964	Ribosomal Protein S16 protein	3.7	2.3
CLS0973	BI678084	Peptidylprolyl Isomerase; Cyclophilin (Cyp)	2.4	2
Secondary and hormone metabolism				
CLS0373	BI677687	Monooxygenase	3.5	2.6
Signal transduction				
CLS0377	BI677691	Protein Phosphatase 2C-like	3.2	2.3
Vesicular trafficking, protein sorting, secretion				
CLS0783	BI677940	ER etention receptor Erd2	2.3	2.8
Unknown				
CLS0032	BI677462	Unknown Protein	3.3	3.1
CLS0071	BI677493	Unknown Protein	2.3	2.1
CLS0186	BI677541	Unknown Protein, hypothetical	2.6	2.8
CLS0227	BI677572	Unknown Protein	4	3.3
CLS0273	BI677607	Unknown Protein, hypothetical	3.8	4.8
CLS0297	BI677624	Unknown Protein, hypothetical	2.6	2.5
CLS0324	BI677649	Unknown Protein, hypothetical	2.8	2.5
CLS0333	BI677657	Unknown Protein, hypothetical	2.5	2.2
CLS0342	BI677665	Unknown Protein, hypothetical	2.9	2.8
CLS0362	BI677678	Unknown Protein, hypothetical	3.3	2.2
CLS0406	BI677712	Unknown Protein, hypothetical	3.5	3.4
CLS0493	BI677785	Unknown Protein, hypothetical	3.1	2.9
CLS0537	BI677821	Unknown Protein, hypothetical	2.5	2.3
CLS0630	BI677890	Unknown Protein, hypothetical	10.7	3.1
CLS0740	BI677903	Unknown Protein, hypothetical	4.3	2.9
CLS0776	BI677934	Unknown Protein	3.3	2.7
CLS0800	BI677955	Unknown Protein	4.1	2.9
CLS0854	BI678208	Unknown Protein, hypothetical	2.3	2.1
CLS1014	BI678038	Unknown Protein, hypothetical	2.4	2.7
CLS1037	BI678119	Unknown Protein	3	2.8
CLS1100	BI678177	Unknown Protein, hypothetical	3.3	2.4
CLS1114	BI678186	Unknown Protein, hypothetical	2.5	2.7

* The ratio was estimated as average value from data points.

Table 1-4. (Cont'd)

Clone ID	GenBank Acc.	Annotation	BC/SW*	BC/TZ*
No hit				
CLS0036	BI677466	No hit	3.6	2.4
CLS0055	BI677481	No hit	4.4	2.1
CLS0060	BI677485	No hit (contig 213)	5.8	3.9
CLS0066	BI677489	No hit	6.3	3.2
CLS0201	BI677554	No hit	3.4	2.6
CLS0211	BI677559	No hit (contig 212)	4.3	2.4
CLS0253	BI677591	No hit	5.2	4.3
CLS0349	BI677670	No hit	9.4	12.3
CLS0366	BI677681	No hit	4.9	2.5
CLS0378	BI677692	No hit	3.8	5.5
CLS0426	BI677729	No hit	2.1	2
CLS0461	BI677758	No hit	2.1	4.3
CLS0568	BI677843	No hit	2.8	2.7
CLS0608	BI677873	No hit	2.1	2
CLS0712	BI678306	No hit	2.9	2.3
CLS0739	BI677902	No hit	4	3.3
CLS0827	BI678187	No hit	6.5	4.8
CLS0840	BI678197	No hit	2.8	3.6
CLS0884	BI678235	No hit	8	2.9
CLS0982	BI678092	No hit	2.6	2
CLS1048	BI678129	No hit	2.5	2.5
CLS1095	BI678170	No hit	2.4	2
SWS0575	BI678741	No hit	2.2	3.6
SWS0655	BI678790	No hit	2.6	2.6
TZS0560	BI642188	No hit	3.3	3.2
TZS1309	BI642865	No hit	3.1	2.5

* The ratio was estimated as average value from data points.

Table 1-5. Up-regulated transcripts in the sapwood

Clone ID	GenBank Acc.	Annotation	SW/BC*	SW/TZ*
Membrane transport				
TZS0226	BI642581	Sugar transport protein	2.8	3
	No hit			
SWS0332	BI678553	No hit	6.8	2.3
	Primary metabolism			
SWS0562		Acetyl-CoA Carboxylase Carboxyl Transferase	2.4	3.3

* The ratio was estimated as average value from data points.

There were dramatic differences in gene expression patterns in the transition zone relative to the other two comparisons. As shown in Figure 1-3D, the gene expression ratios in the transition zone versus the other two zones showed a more definite relationship, indicating a higher correlation and lessened variability (slope= 0.6999 and R^2= 0.5123). Genes with low expression ratios in the transition zone verses bark/cambial region were also low at the ratio of transition zone verses sapwood. In addition, genes that were highly expressed in the transition zone compared to bark/cambial region also had a higher expression in the transition zone than in sapwood region. Of 357 genes that had a higher expression in the transition zone than in either the bark/cambial region or sapwood region, 75 genes were specifically up regulated and 68 genes were down regulated in transition zone (Tables 1-3 and 1-6). Tables 1-3 and 1-6 list the proportion of functional categories of up-regulated genes in transition zone. The results of the EST analysis and microarray results show that the proportion of secondary and hormone metabolism is relatively high (10.7%) in the transition zone. In addition, three TZS clones

Table 1-6. Up-regulated transcripts in the transition zone

Clone ID	GenBank Acc.	Annotation	TZ/BC*	TZ/SW*
Chromatin and DNA metabolism				
TZS0924	BI642324	SAP1 Protein	3.1	3.2
Defense				
TZS0124	BI642508	PR-10 Protein	5.3	3.6
TZS0357	BI642069	PR-10 Protein	3.2	2.1
Miscellaneous				
TZS1380	BI642911	NAM-like protein	2.1	2.7
Primary metabolism				
TZS0097	BI642647	Proline Oxidase	2.9	2.4
TZS0519	BI642157	Cytochrome B5 DIF-F	6.3	5.1
Protein synthesis and processing				
TZS0133	BI642514	eIF4F chain p28	4.8	6.8
Secondary and hormone metabolism				
TZS0021	BI642439	Chalcone-Flavone Isomerase	4.4	4.5
TZS0108	BI642499	Naringenin 3-Dioxygenase	6.5	5.5
TZS0312	BI642638	Flavonoid 3',5'-Hydroxylase	4.6	4.8
TZS0424	BI642100	Chalcone Synthase	5	4.2
TZS0751	BI642292	Flavonoid 3'-hydroxylase	3.4	2.6
TZS0854	BI642412	Chalcone Flavone Isomerase	2.6	2.3
TZS0870	BI642387	Dihydroflavonol 4-reductase	7.4	5.7
TZS0962	BI642432	Chalcone Reductase	5.5	4.1
Signal transduction				
TZF0100	BI642916	Ser/Thr-specific protein kinase	2.2	2
TZS0163	BI642537	GTP-binding Protein, ras-like	5.2	4.7
Unknown				
CLS1022	BI678108	Unknown Protein	3.4	3.1
SWS0040	BI678344	Unknown Protein	3.5	5.2
TZF0001	BI642988	Unknown Protein	2.2	2.9
TZF0010	BI642996	Unknown Protein	2.1	3
TZF0020	BI642927	Unknown Protein	2.3	2.8
TZF0137	BI643021	Unknown Protein	2.6	2.3
TZF0160	BI643042	Unknown Protein	2.2	3.1
TZS0046	BI642463	Unknown Protein (contig 222)	4.9	6.2
TZS0295	BI642621	Unknown Protein	3.1	3.8
TZS0445	BI642109	Unknown Protein	6.5	3.9
TZS0547	BI642176	Unknown Protein	3.2	3.1
TZS0826	BI642402	Unknown Protein	2	2.1
TZS0828	BI642404	Unknown Protein	2.7	2
TZS0984	BI642764	Unknown Protein	5.1	5.3
TZS1012	BI642649	Unknown Protein	6.7	5.7
TZS1095	BI642715	Unknown Protein	5.6	6
TZS1161	BI642738	Unknown Protein	2.1	2.3
TZS1232	BI642802	Unknown Protein	3.2	3.2
TZS1287	BI642850	Unknown Protein	2.4	2.6
TZS1375	BI642907	Unknown Protein	2.5	2.6

* The ratio was estimated as average value from data points.

Table 1-6. (Cont'd)

Clone ID	GenBank Acc.	Annotation	TZ/BC*	TZ/SW*
No hit				
SWS0033	BI678339	No hit	2.2	3.2
SWS0663	BI679126	No hit	2.6	2.2
SWS0732	BI678837	No hit	2.2	2.1
SWS0757	BI678855	No hit	2.6	2
SWS0820	BI679148	No hit	3.3	3.3
SWS0855	BI679178	No hit	3.5	2.7
SWS0931	BI678931	No hit	2.7	3.6
SWS0983	BI679030	No hit	3.5	2.2
SWS1086	BI679045	No hit	2.2	2.6
SWS1123	BI679072	No hit	2	4
TZF0086	BI642977	No hit	2.2	2.3
TZF0109	BI643000	No hit	2.1	2.9
TZS0038	BI642455	No hit (contig 224)	4.3	4.6
TZS0145	BI642532	No hit	6.2	5.4
TZS0161	BI642544	No hit	5.8	4.5
TZS0278	BI642608	No hit	3.7	3.3
TZS0362	BI642070	No hit	3.3	3.4
TZS0388	BI642081	No hit	2.1	2.5
TZS0392	BI642084	No hit	6.2	6.6
TZS0437	BI642106	No hit	3.9	2.6
TZS0440	BI642107	No hit	7.2	7.6
TZS0550	BI642179	No hit	6.8	7.7
TZS0617	BI642214	No hit (contig 216)	7.2	11.2
TZS0660	BI642240	No hit	2.2	2.1
TZS0837	BI642379	No hit	4.1	3.9
TZS0903	BI642337	No hit	3.4	2.9
TZS0922	BI642430	No hit	2.1	2.1
TZS0944	BI642356	No hit	6.3	10.4
TZS0953	BI642359	No hit	4.2	3.4
TZS0973	BI642756	No hit	3.6	3
TZS0986	BI642766	No hit	3.7	4.1
TZS1067	BI642697	No hit	2.2	3.3
TZS1112	BI642719	No hit	2.2	2.5
TZS1226	BI642796	No hit	3.8	5.4
TZS1268	BI642831	No hit	2.9	3
TZS1331	BI642873	No hit	3.1	3
TZS1334	BI642879	No hit	2	3.4

* The ratio was estimated as average value from data points.

(one unknown and two no-hit genes) that are highly redundant in the contig sets (contig 222, 224 and 216) were also highly expressed in transition zone. These results, when combined with the EST analysis and microarray data, show that secondary metabolism-related genes are up-regulated in transition zone. No-hit clones were found in a high proportion in this region as well, possibly indicating the relatively unstudied nature of this unique plant tissue zone. Interestingly, the bark/cambial region and the transition zone showed opposite gene-expression patterns, with most of the down-regulated genes in transition zone being up regulated in the bark/cambial region. For example, phloem-specific protein *VEIN*1 was up regulated in bark/cambial region, but it was down-regulated in transition zone. Such contrasting gene expression is not unexpected considering the functional differences of the two regions. Transition zone is the region where many metabolic changes occur in the establishment of the inner wood. In fact, there was a diversity of genes classified into secondary and hormone metabolism as well as a remarkable increase in the proportion of genes related to secondary metabolism in the TZS library. This observation indicates that the transcript expression pattern of transition zone is highly related to secondary metabolism involving the flavonoid biosynthesis pathway.

Identification of the wood formation-associated genes

To visualize the inner-wood gene expression patterns that could potentially identify the wood formation-related genes, I performed hierarchical clustering of the arrayed genes based on microarray results. Data from the 2,580 clones on my cDNA microarray were clustered from the data of two different experiments:

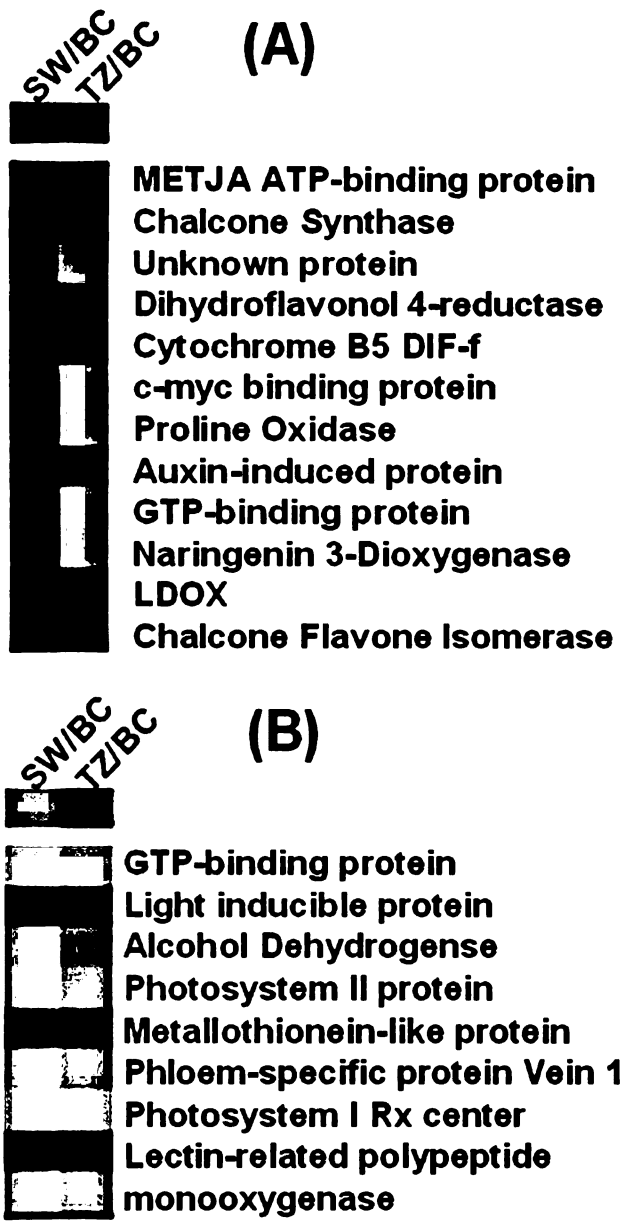


Figure 1-4. Cluster analysis of expression ratios from SWS or TZS vs. BCS
(A) The subcluster of highly expressed genes in inner wood (the sapwood and transition zone) from total hierarchical clustering display.

(B) The subcluster of high expression genes in the bark/cambial zone from total hierarchical clustering display. Only hit EST clones were included for cluster analysis. BC (CZ); bark/cambial zone, SW; sapwood, and TZ; transition zone

SWS versus BCS and TZS versus BCS. Information from clustering not only allows for the identification of related expression patterns of different genes but also shows expression patterns of individual genes over different experiments. Figure 1-4A represents genes that show higher expression in inner wood (i.e., sapwood and transition zone) than in bark/cambial region. Many up-regulated genes in inner wood are not characterized, but there are some interesting genes related to secondary metabolism (chalcone synthase, chalcone flavone isomerase and dihydroflavonol 4-reductase), primary metabolism (proline oxidase, cytochrome B5 DIF-F), and signal transduction (c-myc binding protein and GTP-binding protein ras-like). This observation suggests that these genes may be proximately induced in inner wood and their expression in inner wood may account for the heartwood formation in trunk wood. On the contrary, down-regulated genes are monooxygenase, alcohol dehydrogenase, metallothionein-like protein, phloem-specific protein, and so on (Figure 1-4B).

Confirmation of Microarray Results

Even though the data generated from microarray experiments are reproducible, the data should also be confirmed by northern blot analysis, western blot analysis, or RT-PCR analysis (Seki *et al.*, 2001; Perez-Amador *et al.*, 2001; Yu *et al.*, 2002). In order to compare the expression patterns of each region based on the same reference, I conducted the second microarray experiment using two probes labeled from the RNA populations of the target sample and the seedling control. Due to the nature of trunk wood samples, where only about 10% of the cells are live and the presence of wood extractives makes the isolation of RNA difficult, I was unable to obtain a large amount of mRNA for

conventional northern blot analysis. Accordingly, I performed the approach of "Antisense northern blot analysis" as well as EST results. Expression patterns in the microarray experiment were confirmed by antisense northern blot analysis (Figure 1-5) using Histon H 3.2 as control. Considering the data derived from ESTs, microarray, and "antisense northern blot" analyses; the gene expression profiles reported here are considered reliable.

	(A) Microarray				(B) Antisense Northern Blot
	SD	BC	SW	TZ	SD BC SW TZ
Phloem specific protein	1	2.8	0.1	0.1	
Sugar transport protein	1	0.3	1.2	0.2	
Ser/Thr specific protein kinase	1	5.5	45	63	
WRKY-like protein	1	11	64	56	
Auxin-induced protein (TZS1033)	1	0.6	2.4	4.6	
PR-10	1	0.1	0.7	1.7	
Cytochrome b5 DIF-F	1	1.1	3.5	12	
Flavonoid 3,5-hydroxylase	1	2.8	-	23	
Naringenin 3-dioxygenase	1	8.6	10	50	
No hit (TZS0038, contig224)	1	0.5	9.2	45	
Histon H 3.2	1	0.8	1.1	1.0	

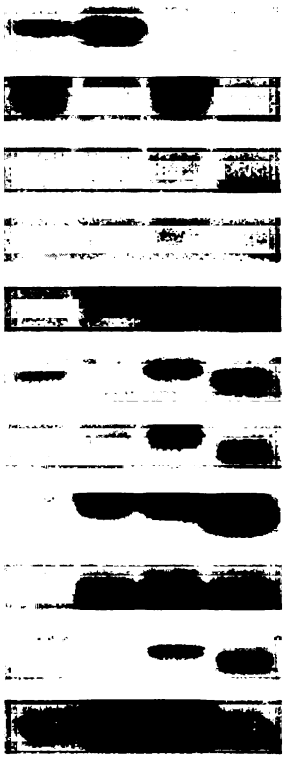


Figure 1-5. Confirmation of microarray Data with antisense northern blot analysis
(A) The expression ratio of the selected genes in microarray analysis with each sample (BC ; bark/cambial zone, SW; sapwood, and TZ; transition zone) versus seedling (SD). Smaller than 1.0 means down-regulation in the target sample compared to the seedling control (sd). Greater than 1.0 indicates up-regulation in the target sample. (B) The expression patterns of the selected genes in seedling, bark/cambial zone, sapwood, and transition zone by antisense northern blot analysis. Each sample lane contained the equal amount of 200 ng of aRNAs.

Discussion

In addition to the major constituents (i.e., cellulose, hemicellulose, and lignin), wood contains many extractives including tannins and other polyphenolics, coloring matter, essential oils, fats, resins, waxes, gum starch, and other secondary metabolites. These extractives make wood a veritable chemical storehouse that provides many organic compounds including biocides for biological control of insects and diseases, adhesives, biofuels, industrial oils, preservatives, pharmacologically active compounds, and rubber. Little is known about the molecular basis for such chemical diversity in heartwood. In order to gain insight on the temporal, spatial, and developmental regulation of the genes involved in the processes of heartwood and its extractive formation, I carried out global examination of gene expression profiles across the stems of mature heartwood-forming trees. It is very difficult to obtain high quality and quantity mRNA from the limited number of live inner wood cells impregnated with extractives. In this study, I characterized the transcriptional profile of black locust wood trunk through the generation of 2,915 ESTs. Assembly of 2,915 ESTs estimated the maximum number of unique genes represented in this set to be 2,278. Because this analysis was performed on 5'end sequences that may arise from multiple non-overlapping segments of the same cDNA, the true number of unique genes may be overestimated. Of the 1,304 matched ESTs that were analyzed, the largest number (30 %) were uncharacterized genes, categorized as hypothetical or unknown proteins. Many of the known transcripts belonged to groups related to housekeeping genes, such as those involved in protein synthesis and processing. The low representation of genes associated with defense in the SWS library and with secondary and hormone metabolism in both the BCS and SWS

libraries are also notable considering the physiological importance of the regions. Thus, the libraries from the different trunk regions have distinctive characteristics based on the position of the sample in the wood. They show a common overall trend of the proportion of genes in the various categories, but with differential expression in some categories that highlights the unique metabolic characteristics of each zone.

The results of my contig analysis show that functional categorization also presents the specified function of each tissue region. A direct proportion comparison of functional categories within each library could not be completed reliably, due to the high percent of no hit clones in the SWS. So, I excluded the portion of no-hit clones and compared the proportions of functional categories in each library. All three libraries contained similar percents of genes in their unknown function groups, as well as their housekeeping function category, which is related to protein synthesis and processing. When compared with other libraries, the proportion of membrane transport- related genes was high in the BCS, but the category of secondary and hormone metabolism membrane was highly scored in the TZS.

It is predicted that the bark/cambial region library would contain many genes categorized into the cytoskeleton, vesicle trafficking, cell division and cycle functioning categories because this region includes actively growing and differentiating cells. However, even though representatives of these classes were found in the library, the proportions were not significantly different from those of other libraries. Additionally, I found that this library has a comparatively large number of genes classified into membrane transport such as: aquaporin, aquaporin-like protein PIP2, plasma membrane intrinsic protein, plasma membrane integral protein ZmPIP2-7, and delta type tonoplast

intrinsic protein. Many transcripts (e.g., the genes encoding phloem-specific proteins) of unknown function or no hit were also highly expressed in this region. When considering the cell growth and division in this region, the proteins encoded by the genes in each of these categories are of interest for future analysis.

The sapwood sample included developing-xylem cells. Like the ESTs derived from the developing-xylem cells of pine and popular, my sapwood library contained a higher concentration of those genes involved in cell wall synthesis, when compared to the other two libraries. However, this library possessed only a few transcripts coding for enzymes involved in the synthesis of lignin. Instead, there were clones corresponding to cell wall structural proteins including: extensin-like proteins, proline-rich proteins, leucine-rich repeat proteins, arabinoxylan arabinofuranohydrolase isoenzyme AXAH-II, and pectinacetylesterases. These results indicate that sapwood gene expression patterns more closely resemble the characteristics of inner wood gene expression than that of developing-xylem. This finding parallels the fact that sapwood is a part of inner wood and has ray cells, which remain alive and maintain their metabolic activity. Based on microarray results few genes are specifically expressed in the sapwood when compared with bark/cambial region and transition zone. One possible explanation for the small number of specifically expressed genes in sapwood is that this region might play bridging roles between bark/cambial region and transition zone.

As in the EST results, genes involved in secondary metabolism were highly expressed in transition zone. Recent research has shown a specific example of this in black walnut (*Juglans nigra*), where flavonoid biosynthesis was up-regulated in the transition zone (Beritognolo et al., 2002). In the July samples of *Juglans nigra*, both

chalcone synthase (CHS) and flavonoid 3'-hydroxylase (F3H) reached their maximum levels of expression in the transition zone, while phenylalanine ammonia lyase (PAL) had increased expression in the outer sapwood. Its activity was much higher in the transition zone than the sapwood of black locust (Magel and Hübner, 1997). In the current study, its expression was increased only in the sapwood (8-fold), but no change in the transition zone. This may reflect the difference in the genotypes used in the studies, sampling time, age, or environmental conditions where the trees were grown. This suggests that the basis for the heartwood chemical diversity among different trees may be at the transcriptional level. My study showed that CHS was up-regulated 5-fold in the transition zone as compared to the bark/cambium region, 4-fold compared to the sapwood, and the level of F3H was increased about 3-fold in the transition zone. These results corroborate with the gene expression patterns observed in black walnut (Beritognolo et al., 2002). The expression of dihydroflavonol 4-reductase (DFR) could not be detected by northern blot, only by RT-PCR in the trunk of walnut, and showed no significant change across the stem sections. However, it was up-regulated 7- and 6-fold in the transition zone of black locust as compared to the bark/cambium region and the sapwood, respectively. This differential expression of the structural genes may explain the difference in the heartwood chemical profiles of the two species. The high expression of cytochrome B5 in the transition zone may be required for flavonoid biosynthesis during heartwood formation. This gene is known to enhance the activity of flavonoid 3',5'-hydroxylase, which catalyzes the 3', 5'-hydroxylation of dihydroflavonols, the precursors of purple anthocyanins (de Vetten, et al., 1999). I found that the expression level of B5 DIF-F (accession number Bl642157) was up-regulated 6- and 5-fold in the transition compared

to the bark/cambium region and the sapwood, respectively. Accordingly, the expression of flavonoid 3',5'-hydroxylase gene was dramatically increased in the transition zone. The high level expression of the genes related to flavonoid biosynthesis in the transition zone fits well with the darkening of the heartwood that is mediated by these genes. The pathogenesis-related class 10 (PR-10), which is a protein related to defense, was also highly expressed in the transition zone, as well as the sapwood region. WRKY, which is a family of plant-specific zinc-finger-type transcriptional factors, was expressed in trunk wood at low levels. Eulgem et al. (1999) reported that the promoter of PR-10 gene is regulated by WRKY proteins as an early defense response in parsley. In addition, Ser/Thr specific protein kinase genes and auxin-induced protein genes were also highly expressed in the transition zone.

The stems of growing trees are characterized by low oxygen and high carbon dioxide (CO_2) concentrations (Carrodus and Triffett, 1975). Under such conditions, Magel (2000) suggested that the products of the accelerated oxidative pentose-phosphate pathway might be used for the synthesis of phenolics. My microarray has two pentose-phosphate pathway genes, transaldolase and fructose bisphosphate aldolase. Both of those genes were up-regulated in the sapwood. The plant hormone ethylene has been suggested as an important regulator of heartwood formation (Nilsson et al., 2002). This view is further supported by the fact that ethylene production was greater in the transition zone than in the outer sapwood (Nelson, 1978) and it stimulates the activity of important enzymes for polyphenol biosynthesis (Roberts and Miller, 1983; Ingermarsson, 1995). Neither of the ethylene receptor gene (accession number Bl677627) and the ethylene-responsive small GTP-binding protein (accession number Bl677741) on my microarray

showed changes in their expression levels across the stem.

One important question in the study of heartwood formation is the source of the carbon-skeletons used for the biosynthesis of heartwood extractives. The evidence gathered so far support the hypothesis that the substrates are derived from imported carbohydrates (Magel and Hübner, 1997; Hauch and Magel, 1998; Magel, 2000). The microarray analysis showed that the gene encoding sucrose transporter (accession number BI642581) was up-regulated 2.8-times in the sapwood, suggesting increased activity of sucrose transporter in the region. Carbohydrates have been shown to be distributed across trunk wood (Magel *et al.*, 1994; Uggla *et al.*, 2001), and there are many reports confirming that sugar transporters play a role in the cell-to-cell and long-distance distribution of sugars throughout the plant (Williams *et al.*, 2000). The high expression of sugar transport protein genes in the sapwood region may indicate that carbohydrates from source tissues are transported to inner wood by the sugar transport proteins. During summer months, starch is accumulated in the sapwood and its accumulation correlates with enhanced sucrose synthase (SuSy) activities in the sapwood (Magel, 2000). Especially, the activity of SuSy increases dramatically at the sapwood-heartwood transition zone, which may lead to enhanced degradation of sucrose in the region where the synthesis and accumulation of phenolic heartwood extractives occur (Magel *et al.*, 1994; Magel and Huber, 1997; Magel, 2000). The activity of SuSy has been proposed as a measure for sink strength in tissues with extensive synthesis of phenolic compounds (Magel, 2000). In this case, my data add additional evidence for the view that heartwood extractives are synthesized at the transition zone using imported carbohydrates, not translocated via the phloem and the wood rays to the heartwood (Steward, 1966).

In summary, I report the first comprehensive study of gene expression profiles deep inside trunk wood of a hardwood tree. My ESTs present unique gene sets that are expressed in uncharted plant tissues. Using DNA microarray analysis, I have identified genes associated with inner wood formation and profiled gene expression patterns in trunk wood. These genes will assist future investigations to unravel the molecular mechanisms regulating the formation of inner wood. Resolving the dilemma of achieving greater environmental protection of forest ecosystems while meeting the increasing demand for forest utilization necessitates gaining a fundamental understanding of the biochemical processes involved in tree growth and development. The findings described here will provide a platform for such attempts.

Literature Cited

Adams, M., Kerlavage, A., Fields, C., and Venter, J. 1993. 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nat Genet 4: 256-267.

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

Audic, S. and Claverie, J-M. 1997. The significance of digital gene expression profiles. Genome Research 7: 986-995.

Bairoch, A. and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28: 45-48.

Barker, W., Garavelli, J., Huang, H., McGarvey, P., Orcutt, B., Srinivasarao, G., Xiao, C., Yeh, L.-S.L., Ledley, R., Janda, J., Pfeiffer, F., Mewes, H.-W., Tsugita, A. and Wu, C. 2000. The Protein Information Resource (PIR). Nucleic Acids Res. 28: 41-44.

Baugh, L., Hill, A., Brown, E., and Hunter, C. 2001. Quantitative analysis of mRNA amplification by in vitro transcription. Nuccleic Acid Res. 29: e29.

Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J., Rapp, B., and Wheeler, D. 2000. GenBank. Nucleic Acids Res. 28: 15-18.

Beritognolo, I., Magel, E., Abdel-Latif, A., Charpentier, J., Jay-Allemand, C., and Breton, C. 2002. Expression of genes encoding chalcone synthase, flavanone 3-hydroxylase, and dihydroflavonol 4-reductase correlates with flavanol accumulation during heartwood formation in *Juglans nigra* L. Tree Physiol 22: 291-300.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S.A., Patel, K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T.M., Terryn, N., Gielen, J., Villarroel, R., and Chalwatzis, N. 1998. Analysis of 1.9 Mb contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391: 485-488.

Burtin, P., Jay-Allemand, C., Charpentier, J., and Janin, G. 1998 Natural wood colouring process in *Juglans* sp. (*J. nigra, J. regia* and hybrid *J. nigra* 23 x *J. regia*) depends on native phenolic compounds accumulated in the transition zone between sapwood and heartwood. Trees 12: 258-264.

Carrodus, B., and Triffett, A. 1975. Analysis of respiratory gases in woody stems by mass spectrometry. New Phytol. 74: 243-246.

Cercos, M., Santamaria, S., and Carbonell, J. 1999. Cloning and characterization of TPE4A, a thiol-protease gene induced during ovary senescence and seed germination in pea. Plant Physiol 119: 1341-1348.

Covitz, P., Smith, L., and Long, S. 1998. Expressed Sequence Tags from a Root-Hair-Enriched *Medicago truncatula* cDNA library. Plant. Physiol. 117: 1325-1332.

Dazzo, F., and Hubbell, H. 1975. Cross-reactive antigens and lectins as determinants of symbiotic specificity in the *Rhizobium*–clover association. Appl. Microbiol. 30: 1017-1033.

de Vetten, N., ter Horst, J., van Schaik, H.-P., de Boer, A., Mol, J., and Koes, R. 1999. A cytochrome b_5 is required for full activity of flavonoid 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flowers. Proc Natl Acad Sci USA 96: 778-783

Dent, G., O'Dell, D, and Eberwine, H. 2001. Gene expression profiling in the amygdala: An approach to examine the moleculat substrates of mammalian behavior. Physiol. & Behavior 73: 841-847.

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

Eulgem, T., Rushton, P., Schmelzer, E., Hahlbrock, K., and Somssich, I. 1999 Early nuclear events in plant defense signaling: rapid gene activation by WRKY transcription factors. EMBO J. 18: 4689-99.

Ewing, B., Hillier, L., Wendl M., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8: 175-85.

Fukuda, H. (1996). Xylogenesis: initiation, progression, and cell death. Ann Rev Plant Physiol Plant Mol Biol 47, 299-325.

García-Hernández, M., Murphy, A., and Taiz, L. 1998. Metallothioneins 1 and 2 have distinct but overlapping expression patterns in Arabidopsis. Plant Physiol. 118: 387-397.

Hauch, S. and Magel, E. 1998. Extractable activities and protein content of sucrose-phosphate synthase, sucrose synthase and neutral invertase in trunk tissues of *Robinia pseudoacacia* L. are related to cambial wood production and heartwood formation. Planta 207: 266-274

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

Hihara, Y., Kamei, A., Kanehisa, M., Kaplan, A., and Ikeuchi, M. 2001. DNA Microarray Analysis of Cyanobacterial Gene Expression during Acclimation to High Light, Plant Cell 13: 793-806

Hillinger, C., Holl, W., and Ziegler, H. 1996a. Lipids and lipolytic enzymes in the trunkwood of *Robinia pseudoacacia* L. during heartwood formation. I. Radial distribution of lipid classes. Trees 10: 366-375.

Hillinger, C., Holl, W., and Ziegler, H. 1996b. Lipids and lipolytic enzymes in the trunkwood of *Robinia pseudoacacia* L. during heartwood formation. II. Radial distribution of lipases and phospholipases. Trees 10: 376-381.

Hillis W. 1987. Heartwood and the exudates, Spinger, Berlin, Heidelberg, New York.

Ingemarsson, B. 1995. Ethylene effects on peroxidases and cell growth patterns in *Picea abies* hypocotyl cuttings. Physiol. Plant. 94: 211-218

Kirst, M., Johnson, A., Retzel, E., van Zyl, L., Craig, D., Li, Z.J., Whetten, R., Baucom, C., Ulrich, E., Hubbard, Kristy, and Sederoff, R. (2002). Quantitative inference in functional genomics of loblolly pine (*Pinus taeda* L.) using ESTs and microarrays. *In* Proc. the 10th IAPTC&B Congress. (Ed) I. Vasil. Kluwer Academic Publishers, Dordrecht, Netherlands.

Kozlowski, T. and Pallardy, S. (1997). Physiology of woody plants. San Diego, Academic Press.

Lalonde, S., Boles, E., Hellmann, H., Barker, L., Patrick, J., Frommer, W., and Ward, J. 1999. The dual function of sugar carriers: transport and sugar sensing. Plant Cell 11: 707-726.

Lee, M., Kuo, F., Whitmore, G., and Sklar, J. 2000. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. Proc. Natl. Acad. Sci. USA 97: 9834–9839.

Livesey, F., Furukawa, T., Steffen, M., Church, G., and Cepko, C. 2000 Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene *Crx*. Current Biology 10:301-310.

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

Magel, E., Jay-Allemand, C., and Ziegler, H. 1994. Formation of heartwood substances in the stemwood of Robinia pseudoacacia L.: II. Distribution of nonstructural carbohydrates and wood extractives across the trunk. Trees 8: 165-171.

Magel, E. and Hübner, B. 1997 Distribution of phenylalanine ammonia lyase and chalcone synthase within trunks of *Robinia pseudoacacia* L. Bot. Acta 110: 314-322

Magel, E. 2000. Biochemistry and physiology of heartwood formation. In Molecular and Cell Biology of Wood Formation. Eds. R. Savidge, J. Barnett and R. Napier. BIOS Scientific Publishers, Oxford, pp363-376.

Matamoros, M., Baird, L., Escuredo, P., Dalton, D., Minchin, F., Iturbe-Ormaetxe, I., Rubio, M., Moran, J., Gordon, A., and Becana, M. 1999. Stress-induced legume root nodule senescence. Physiological, biochemical, and structural alterations. Plant Physiol 121: 97-112.

Mauseth, J. 1998. Botany: an introduction to plant biology. Jones and Bartlett Publishers. Sudbury, Massachusetts.

Miller, J., Arteca, R., and Pell, E. 1999. Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. Plant Physiol 120: 1015-1024.

Nelson, N. 1978. Xylem ethylene, phenol oxidising enzymes and nitrogen and heartwood formation in walnut and cherry. Can. J. Bot. 56: 626-634

Nilsson, M., Wikman, S., and Eklund, L. 2002. Induction of discolored wood in Scots pine (*Pinus sylvestris*). Tree Physiology 22: 331-338

Perez-Amador, M., Lidder, P., Johnson, M., Landgraf, J., Wisman, E., and Green, P. 2001. New molecular phenotypes in the dst mutants of Arabidopsis revealed by DNA microarray analysis. Plant Cell 13: 2703-2717.

Reddy, A., and Poovaiah, B. 1990. Molecular cloning and sequencing of a cDNA for an auxin-repressed mRNA: correlation between fruit growth and repression of the auxin-regulated gene. Plant Mol Biol 14: 127-136.

Roberts, L. and Miller, A. 1983 Is ethylene involved in xylem of differentiation? Plant Physiol. 6: 1-24

Saxe, H., Ellsworth, D. S., and Heath, J. 1998. Tree and forest functioning in an enriched CO_2 atmosphere. New Phytol. 139: 395-436.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E. 2001. Microarray Analysis of Diurnal and Circadian-Regulated Genes in Arabidopsis. Plant Cell 13: 113-123.

Sederoff, R., Kirst, M., Johnson, A., Retzel, E., Whetten, R., Vasques-Kool, J. & O'Malley, D. 2002. Homology of expressed genes in loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. The 10th IAPTC&B Congress, February 23-28, 2002, Orlando, FL.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. 2001. Monitoring the Expression Pattern of 1300 Arabidopsis Genes under Drought and Cold Stresses by Using a Full-Length cDNA Microarray. Plant Cell 13: 61-72.

Stafstrom, J., Ripley, B., Devitt, M., and Drake, B. 1998. Dormancy-associated gene expression in pea axillary buds. Cloning and expression of PsDRM1 and PsDRM2. Planta 205: 547-552.

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

Steward, C. 1966. Excretion and heartwood formation in living trees. *Science* 153: 1068-1074.

Uggla, C., Magel, E., Moritz, T., and Sundberg, B. 2001. Function and dynamics of auxin and carbohydrates during early wood/late wood transition in Scots pine. Plant Physiol 125: 2029-2039

Wang, E., Miller, L., Ohnmacht, G., Liu, E., and Marincola, F. 2000. High-fidelity mRNA amplification for gene profiling. Nat. Biotech. 18: 457-459.

Williams, L., Lemoine, R., and Sauer, N. 2000. Sugar transporters in higher plants – a diversity of roles and complex regulation. Trens in Plant Science 5: 283-290

Wingler, A., von Schaewen, A., Leegood, R., Lea, P., and Quick, W. 1998. Regulation of leaf senescence by cytokinin, sugars, and light. Plant Physiol 116: 329-335.

Yu, Q., He, M., Lee, N., and Liu, E. (2002). Identification of MYC-mediated death response pathways by microarray analysis. J Biol Chem 277: 13059-13066.

CHAPTER 2

Seasonal changes in gene expression at the sapwood–heartwood transition zone of black locust (*Robinia pseudoacacia*) revealed by cDNA microarray analysis

ABSTRACT

Heartwood is a determining factor of wood quality and understanding the biology of heartwood may allow us to control its formation. Heartwood formation is a form of senescence that is accompanied by a variety of metabolic alterations in ray parenchyma cells at the sapwood–heartwood transition zone. Although senescence has been studied at the molecular level with respect to primary growth, the cell maturation and death events occurring during heartwood formation have been difficult to study because of their location and timing. Analysis of global gene expression patterns during the transition from sapwood to heartwood may offer a powerful means of identifying the mechanisms controlling heartwood formation. Previously, I developed cDNA microarrays carrying 2,567 unigenes derived from the bark/cambium region, sapwood and transition zone of a mature black locust tree. Here, I describe the use of these microarrays to characterize seasonal changes in the expression patterns of 1,873 genes from the transition zone of mature black locust trees. When samples collected in summer and fall were compared, 569 genes showed differential expression patterns: 293 genes were up-regulated (> twofold) in summer (July 5) and 276 genes were up-regulated in fall (November 27). More than 50% of the secondary and hormone metabolism-related genes on the microarrays were up-regulated in summer. Twenty-nine out of 55 genes involved in signal transduction were differentially regulated, suggesting that the ray parenchyma cells located in the inner-most part of the trunk wood react to seasonal changes. I established

the expression patterns of 349 novel genes (previously unknown or no-hit), of which 154 were up-regulated in summer and 195 were up-regulated in the fall.

Introduction

Most tree species have a dark-colored zone called heartwood in the central parts of their trunks. Besides the difference in color, heartwood is distinguished from the surrounding pale-colored sapwood by features such as a greater content of extractives, a lower water content and parenchyma cell death. The presence of heartwood is the major factor determining wood quality. Many characteristics of wood, including dimensional stability, durability, pulpability, color, hue, scents and beauty, are affected by the presence of various organic substances classified as extractives. Identifying changes in gene expression during the transition from sapwood to heartwood may help us control heartwood formation.

Heartwood results from cell death and contains no living cells. Studies of heartwood formation have focused on metabolic changes and extractive formation in a narrow zone, called the transition zone (TZ), adjacent to the heartwood. During the transition from sapwood to heartwood, parenchyma cells of the sapwood undergo metabolic changes that result in heartwood formation and increased synthesis of secondary products such as condensed tannins, terpenes, flavonoids, lignans, lipids, stilbenes and tropolones (Hillis, 1987; Magel et al., 1991, 1994; Hillinger et al., 1996a, 1996b; Burtin et al. 1998). There is much evidence that heartwood extractives are synthesized in situ in the sapwood–heartwood transition zone from the breakdown of starch or from soluble sugars (Kumar and Datta, 1987; Nobuchi et al., 1987a, 1987b;

Magel et al., 1994; Hillinger et al., 1996a, 1996b; Magel and Hübner, 1997). Starch hydrolyzed at the transition zone is a primary source of hydroxycinnamic acid and flavonoids (Magel et al., 1994). A period of enhanced metabolic activity has been found at this zone in heartwood-forming species, such as *Acacia* (Baqui and Shah; 1985), black locust (*Robinia pseudoacacia* L.; Holl and Lendzian, 1973; Magel et al., 1991), oak (*Quercus robur* L.; Ebermann and Stich, 1985) and walnut (*Juglans nigra* L.; Nelson et al., 1981). It has also been found that formation and accumulation of heartwood phenolics coincides with the transformation of sapwood to heartwood in black locust (Magel et al., 1994). Hillis and Hasegawa (1963) noted that, 19 days after labeled glucose was administered to the cambial region of *Eucalyptus sieberi* L. A. S. Johnson, it was converted to labeled extractives formed at the heartwood periphery. This conversion was also observed in the transition zone of sugi (*Cryptomeria japonica* D. Don; Higuchi et al., 1969).

Several enzymes show high activity in the transition zone. Elevated activities of phenol oxidizing enzymes have been observed in the transition zone of *Eucalyptus polyanthemos* Schauer (Hillis and Yazaki, 1973). High peroxidase activity was observed in the transition zone of *Eucalyptus elaeophora* F. Muell. (Wardrop and Cronshaw, 1962) and *Fagus sylvatica* L. (Dietrichs, 1964). Baqui et al. (1979) found that succinate dehydrogenase, adenosine triphosphatase and lipase had enhanced activity only in the transition zone of *Melia azedarach* L. Other enzymes reported to be highly active in the transition zone include catechol oxidase, glucose-6-phosphate dehydrogenase, malic dehydrogenase, maltase and amylase (Hillis, 1987). Furthermore, Magel et al. (1991) have shown that two key enzymes for flavonoid biosynthesis (chalcone synthase (CHS)

and phenylalanine ammonia-lyase (PAL)) are highly active in the sapwood–heartwood transition zone of black locust; PAL was active in the youngest wood near the cambium in April and September and active in the transition zone in all seasons, whereas CHS was active only in the intermediate wood. The stimulus for biosynthesis of robinetin, a flavonoid with strong antimicrobial activity, apparently occurs in the transition zone in *Intsia* species (Hillis, 1996).

In temperate zones, heartwood formation occurs in late summer and continues through late fall and the beginning of dormancy, provided temperatures are above 5 °C (Hillis, 1987). Heartwood formation occurs at the time of cambial dormancy in radiata pine (*Pinus radiata* D. Don) (Shain and Mackay, 1973), walnut and cherry (Nelson, 1978). Studies of the cytology and coloration of extractives suggest that heartwood formation commences in midsummer and continues in the fall and winter seasons in sugi (Nobuchi et al., 1987a) and black locust (Nobuchi et al., 1987b). Beritognolo et al. (2002) studied seasonal changes in the expression of several flavonoid biosynthetic genes during heartwood formation in *Juglans nigra* and demonstrated that flavonol accumulation in the sapwood–heartwood transition zone was correlated with transcript levels of key flavonoid structural genes.

Black locust is one of the most extensively studied hardwood species with regard to heartwood formation (Magel et al., 1991, 1994; Hillinger et al., 1996a, 1996b; Hauch and Magel, 1998). Its sapwood is yellowish and narrow, whereas the heartwood ranges from green to greenish yellow, dark yellow or golden brown. I chose to work with black locust because it is a heartwood-forming species with clearly distinguished sapwood and heartwood zones. In addition, black locust has a relatively small genome (637 Mbp),

making it a suitable species for genomic experiments. I previously isolated mRNA from the inner wood of mature black locust trees (> 10 years old) and analyzed a large number of expressed sequence tags (2,915 ESTs) (Yang et al., 2003). I then constructed cDNA microarrays carrying 2,567 unigenes identified from the EST analysis (a list of the unigenes is available at http://forestry.msu.edu/biotech/Projects.htm). To gain an increased understanding of the genetic regulation of heartwood formation, I report here a series of microarray hybridization experiments performed to examine changes in gene expression in the sapwood–heartwood transition zone of black locust during active growth (July) and during the transition from active growth to dormancy (November).

Materials and Methods

Plant material

In summer (July 5) and fall (November 27), two mature 10-year-old (20-cm diameter at breast height) black locust trees growing at the Tree Research Center on the campus of Michigan State University were felled and cut into 25-cm-long logs. The logs were immediately placed on ice and transported to a wood shop where ~1-cm-thick disks were cut with a table saw. The disks were immediately submerged in RNA extraction buffer (50 mM Tris-HCl buffer, pH8.0, containing 20 mM EDTA, 0.2% SDS and 10 mM 2-mercaptoethanol). The sapwood–heartwood transition zone, located 2–4 cm inside the outer edge of each disk, was carved out with a chisel and hammer that had been prewashed with RNase Away solution (Invitrogen, Carlsbad, CA). The isolated sections (~1 cm 3 cubes) of the transition zone sampled in summer (TZS) and fall (TZF) were frozen in liquid nitrogen and stored at –80 °C until use.

Microarray preparation

The construction of microarrays carrying 2,567 unigenes that were expressed in the bark/cambium, sapwood and transition zone of a mature black locust has been described previously (Yang et al., 2003). Briefly, the PCR-amplified products of the selected unigenes were precipitated with ethanol, re-suspended in 3 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M sodium citrate) and checked for quality by gel electrophoresis. The PCR products were arrayed from 384-well microplates and spotted on superaldehyde glass slides (Telechem, Sunnydale, CA) at a high density with an Omnigridder robot (Gene Machines, San Carlos, CA) and a 2-pin 16 ArrayIt chipmaker (Telechem). Each slide contained duplicate spots for each gene. Slides were rinsed twice in 0.2% SDS for 2 min, and washed in distilled water. To denature the DNA, the slides were transferred to boiling water for 2 min. After denaturing the DNA, the slides were dipped in blocking solution (1.5 g NaBH4 in 450 ml of phosphate buffered saline to which 133 ml of 100% ethanol was added) for 5 min to reduce free aldehydes, and then washed with 0.2% SDS solution and distilled water.

Hybridization and washing of the DNA microarray

For total RNA isolation, frozen TZS and TZF samples were ground in a blender and reground to fine powder with a mortar and pestle. The ground samples were then passed through the shredder column of a DNeasy Maxi kit subjected to total RNA isolation using the RNeasy Maxi kit (Qiagen, Hilden, Germany), and cleaned up with a Qiagen RNeasy Mini kit. The cDNAs were synthesized in reverse transcription reactions using 1 μg of total RNA and then PCR-amplified using the SMART system (ClonTech,

Palo Alto, CA) and the fewest possible cycles (15–18 cycles) to minimize generation of non-specific PCR products. To generate probes for array hybridization, 2 μg of the PCR-amplified cDNAs were labeled by incorporation of either Cy5- or Cy3-dCTP (Amersham-Pharmacia, Piscataway, NJ) during oligo-dT-primed primer extension in the presence of Klenow DNA polymerase (Promega, Madison, WI), as described by Schaffer et al. (2001). The labeling reactions were purified with the QiaQuick PCR cleanup kit (Qiagen). Probe samples were denatured at 100 °C for 3 min, left at room temperature for 30 min, and then used for hybridization. Probes labeled with either Cy3 or Cy5 fluorescent dye were hybridized to a microarray slide in a total volume of 30 μl of hybridization buffer (3.4 × SSC, 0.32% SDS, and 5 μg of yeast tRNA) for 16 h at 65 °C. Three replicate slides were hybridized for each sample pair, including one dye swap experiment in which the labeling of the sample in each pair was reversed on the second slide by using independent RNA preparations. The third slide was a repeat of the first hybridization to check the reliability of the results. Each slide was then washed at room temperature in 1 × SSC, 0.1% SDS for 10 min, in 1 × SSC for 10 min, and in 0.01 × SSC for 10 min. The slides were centrifuge-dried and scanned with a 428 Array scanner (Affymetrix, Palo Alto, CA).

Microarray data analysis

Microarray data were analyzed with GenePix Pro 3.0 (Axon Instruments, Union City, CA). The scanned data were normalized by the Global Normalization method (Hihara et al., 2001), which normalizes the image data between Cy3 and Cy5 channels by adjusting the total signal intensities of two images and removing unreliable spots. The

unreliable spots were discarded based on the following screening. Spots containing clones that had poor amplification, or multiple bands, as well as those that were flagged because of a false intensity caused by dust or background on the array, were removed. Spots with < 65% of the spot intensity at > 1.5-fold that of the background in both channels were ignored (see Stanford Microarray Database Web site, http://genome-www5.stanford.edu/Micro-Array/SMD/). Clones in one sample that had an average induction greater than twofold in another were determined as up-regulated. Data management and analyses were carried out with Microsoft Excel and Microsoft Access database. After normalization, I calculated the means and coefficients of variation for the observed signal intensities in each channel and the ratio of signals from three replicates. All data were then converted to base 2 logarithm. Pairwise comparisons were made by simple linear correlation analysis to evaluate variations within and between microarrays and within and between slides (Fernandes et al., 2002, Swidzinski et al., 2002). To determine the strength of the relationship between replicates, the Pearson correlation coefficient (r) was calculated.

Antisense Northern blot analysis

Antisense RNA (aRNA) was generated as described by Wang et al. (2000). Briefly, aRNA was amplified with a Message Amp aRNA kit (Ambion, Austin, TX). I began by synthesizing first stand cDNAs from the total RNA of seedling, bark/cambial region, sapwood, or transition zone, and the first cDNAs were used as templates for the synthesis of second cDNAs. Finally, aRNAs from the second cDNAs were generated by in vitro transcription (amplification). About 200 ng of aRNA was amplified, separated in

a formamide agarose gel, and transferred onto a nylon membrane by the capillary transfer method. The membrane was then hybridized with a labeled probe. The signal was exposed and detected on X-ray film. Histon H3.2 gene was used as a control, which was similarly expressed in both TZS and TZF based on my microarray results.

Results

Microarray experiments

Correlations between replicate microarrays were high (Table 2-1), confirming that PCR amplification generated highly reproducible populations of cDNAs. Even the lowest r value obtained, 0.86, was well within the range of significance ($P < 0.05$). Hybridization in dye reversal experiments was more variable between replicate microarrays than between replicate elements within the same slide (cf. Fernandes et al., 2002). In my analysis, the highest correlation coefficient ($r = 0.96$) was between duplicate spots on the same slide. The r-value of two experiments in which the same dye was used to label each sample was higher than the r-value of the two-dye reversal experiments. Furthermore, two experiments using the same dye to label each sample were less variable than two-dye reversal experiments. These data imply that replication reliability can be affected by dye reversion and that hybridizations should be replicated at least three times, with at least one replicate being a dye swap experiment. Statistically, to detect up-regulated genes (> twofold increase) on the microarrays, the log2 (TZS/TZF) should fall at least two standard errors away from zero (no change in log2 expression) when considering the three replicate arrays (Swidzinski et al., 2002). Thus, the log2 standard deviation (SD) for each gene, averaged across three replicate arrays, must be less than or equal to 0.5.

Table 2-1. Summary of correlation coefficients of ratios from TZS versus TZF
Numbers 1 and 2 in parentheses represent duplicate 1 and duplicate 2 of the same slide. Abbreviations: TZS or TS, transition zone harvested on July 5; TZF or TF, transition zone harvested on November 27; Cy3, Cy3-deoxyCTP; and Cy5, Cy5-deoxyCTP.

	Replication 1		Replication 2		Replication 3	
	TS-Cy3 *VS.* TF-Cy5(1)	TS-Cy3 *VS.* TF-Cy5(2)	TS-Cy5 *VS.* TF-Cy3(1)	TS-Cy5 *VS.* TF-Cy3(2)	TS-Cy3 *VS.* TF-Cy5(1)	TS-Cy3 *VS.* TF-Cy5(2)
TZS-Cy3 *VS.* TZF-Cy5 (1)	-	0.942	0.91	0.89	0.92	0.92
TZS-Cy3 *VS.* TZF-Cy5 (2)		-	0.92	0.91	0.94	0.93
TZS-Cy5 *VS.* TZF-Cy3 (1)			-	0.93	0.91	0.93
TZS-Cy5 *VS.* TZF-Cy3 (2)				-	0.86	0.91
TZS-Cy3 *VS.* TZF-Cy5 (1)					-	0.96

In 1,873 genes out of 2,580 considered in the analysis, 90% of the means had a SD < 0.5. These 1,873 genes were selected for further analyses.

Differential gene expression between TZS and TZF

Analysis of the microarray data revealed significant changes in transcript levels of the selected 1,873 genes. Figure 2-1 shows that the number of genes was normally distributed in log2 (TZS/ TZF). Only the genes for which the expression varied more than twofold when averaged across six data points (two duplicates per slide and three replications) for each treatment were considered to represent significant changes in gene expression (Girke et al., 2000, Reymond et al., 2000). Genes for which the log2 (TZS/TZF) was more than one were considered as highly expressed genes in TZS. Conversely, genes in which the log2 (TZS/TZF) was less than –1 were considered as highly expressed genes in TZF or as being down-regulated in TZS. About 30% of the 1,873 genes selected displayed significant expression changes. A total of 293 genes were up-regulated in TZS and 276 genes were highly expressed in TZF when compared with TZS (Table 2-2). About 570 genes were differentially expressed between summer and fall.

To assess seasonal changes (from summer to fall) in gene expression patterns of the transition zone, I classified the up-regulated genes in TZS or TZF into 14 functional categories (Table 2-2). The percentage of up-regulated genes to total genes in each sample was similar: 15.6% for TZS (293 genes out of 1,873 total genes), and 14.7% for TZF (276 genes out of 1,873 total genes). The percentages of up-regulated genes to total genes for several categories were similar in the TZS and TZF groups; however, the

percentages for some categories varied between the two groups. For example, the percentage of the genes in the 'unknown' or 'no-hit' category was higher in TZF than in TZS. Conversely, the percentage of genes classified in the 'cell division and cycle', 'defense' and 'secondary and hormone metabolism' categories was higher in TZS than in TZF. Although these findings provide insight into the characteristics of each sample group, I caution that the numbers of genes included in the 'cell division and cycle' and 'vesicular trafficking, protein sort and secretion' categories were only six and five, respectively. The high proportion of secondary and hormone metabolism-related genes among the TZS up-regulated genes indicates distinctive features of the TZS sample. Additionally, the portion of up-regulated genes related to 'signal transduction' in each sample (5.4% in TZS, 4.7% in TZF) was higher than the percentage values calculated in the Total (2.9%). Furthermore, 53% of the genes within the 'signal transduction' category were differentially expressed.

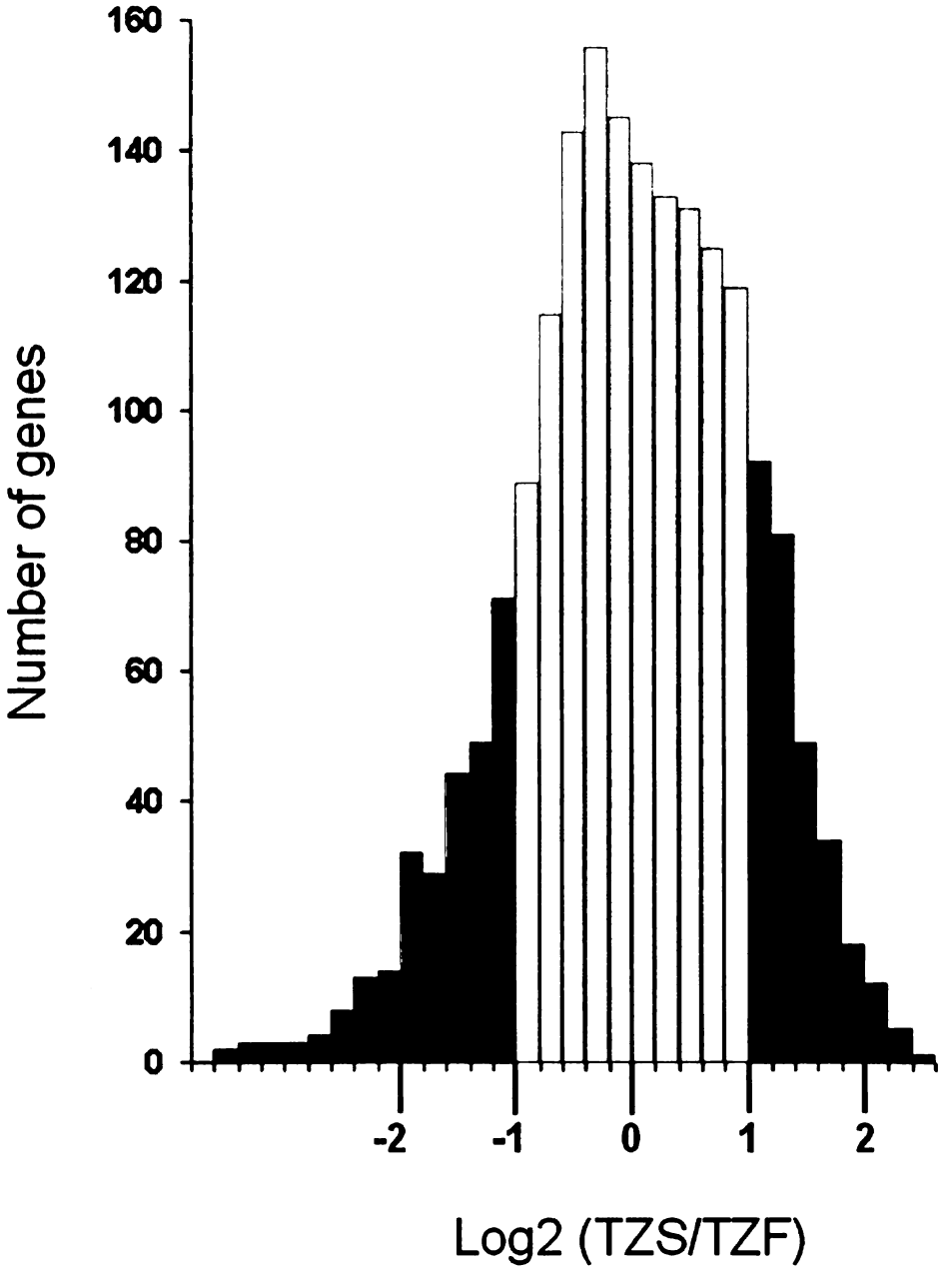


Figure 2-1. Distribution of mean ratios of expression from TZS and TZF samples Total RNAs from the transition zone sampled on July 5 (TZS) and on November 27 (TZF) were labeled with fluorescence dyes (Cy3 or Cy5) and hybridized with the microarrays. The ratio for each clone is the mean of three replicates.

Table 2-2. Functional classification of up-regulated clones in transition zones
Numbers are of selected clones for microarray analysis after normalization. Numbers in parentheses indicate the percentage of each number per total number.

Category	Total	TZS	TZF
Cell division and cycle	6 (0.3)	4 (1.4)	0 (0.0)
Cell wall structure and metabolism	13 (0.7)	2 (0.7)	1 (0.4)
Chromatin and DNA metabolism	26 (1.4)	7 (2.4)	3 (1.1)
Cytoskeleton	6 (0.3)	0 (0.0)	1 (0.4)
Defense	43 (2.3)	13 (4.4)	7 (2.5)
Gene expression and RNA metabolism	55 (2.9)	7 (2.4)	8 (2.9)
Membrane transport	19 (1.0)	4 (1.4)	2 (0.7)
Miscellaneous	85 (4.5)	22 (7.5)	15 (5.4)
Primary metabolism	107 (5.7)	31 (10.6)	22 (8.0)
Protein synthase and processing	74 (4.0)	15 (5.1)	8 (2.9)
Secondary and hormone metabolism	31 (1.7)	16 (5.4)	1 (0.4)
Signal transduction	55 (2.9)	16 (5.4)	13 (4.7)
Vesicular trafficking, protein sort, Secretion	5 (0.3)	2 (0.7)	0 (0.0)
Unknown, No hit	1348 (72)	154 (33)	195 (71)
Total clones	1873	293	276

Genes up-regulated in summer (TZS)

A total of 293 genes were up-regulated in TZS (139 known and 154 unknown/no hit genes) (Table 2-3). Three out of four cell division and cycle-related genes that were up-regulated in TZS were SKP1 (suppressor of kinetochore protein) genes. Two cell wall biosynthesis-related genes were highly expressed in TZS. Stress-inducible genes such as pathogenesis-related class 10 gene (PR-10), and genes coding for aluminum-induced protein, metallothionein and dehydration induced protein, were highly expressed in TZS. It is notable that the gene coding for aquaporin-like protein PIP2 was up-regulated in TZS, as were genes coding for dehydration-induced protein. Two genes coding for auxin-repressed protein and one auxin-induced protein were up-regulated in TZS, suggesting that the expression of these genes could be differentially regulated in the transition zone. The gene for sucrose synthase, which is known to be involved in the production of substrates for heartwood formation in the transition zone (Shah et al., 1981, Nobuchi et al., 1982), and the gene for sugar transport protein were 3.7- and 3.9-fold up-regulated in TZS, respectively. Other TZS up-regulated genes in the 'primary metabolism' category included those encoding cytochrome B5 DIF-F (2.8-fold), phenylalanine ammonia lyase (PAL, 3.3-fold), adenosine kinase-like protein (4.4-fold) and glycine hydroxymethyltransferase (5.1-fold). Sixteen of the 31 genes in the 'secondary and hormone metabolism' category were up-regulated in TZS. Nine genes were involved in the flavonoid pathway. Two lignin biosynthesis-related genes, genes encoding Caffeoyl-CoA o-methyltransferase and Caffeic acid o-methyl-transferase, were up-regulated in TZS. Two genes encoding glutathione S-transferase were more highly expressed in TZS than in TZF. Two genes encoding monooxygenase and one gene for fumarylacetoacetase

Table 2-3. Categorized genes whose expression was up-regulated in summer (TZS)

CloneID	GenBank Acc.	Annotation	TZS/TZF
Cell division and cycle			
CLS0043	BI677470	Skp1	2.3
CLS0092	BI677511	Calcineurin B Subunit	2.2
CLS0877	BI678229	SKP1	3.0
TZS0380	BI642078	SKP1 Homolog	2.8
Cell wall structure and metabolism			
CLS0298	BI677625	Pectinesterase-like protein	2.0
CLS0529	BI677813	Polygalacturonase-like protein	2.3
Chromatin and DNA metabolism			
CLS0083	BI677503	HMG1/2-like Protein (SB11 Protein)	2.4
CLS0418	BI677721	Histone H4	3.5
CLS0521	BI677806	Histone H2B1	2.3
CLS0549	BI677829	DNA Topoisomerase VI subunit B-like protein	2.2
CLS0968	BI678080	DNA-Binding Protein, hypothetical	2.2
TZS0239	BI642590	Histone H2A	3.1
TZS0517	BI642156	DNA Polymerase Accessory Protein	2.2
Defense			
CLS0021	BI677451	Dehydration-induced protein	2.7
CLS0176	BI677536	L-ascorbate peroxidas, cytosolic	5.2
CLS0243	BI677582	Superoxide Dismutase (Cu-Zn)	2.2
CLS0450	BI677748	Metallothionein	2.2
CLS0513	BI677800	Disease Resistance Protein-like	2.3
CLS0609	BI677874	Peroxidase pxdD precursor	2.5
TZS0036	BI642453	Aluminum-induced Protein	2.9
TZS0124	BI642508	PR-10 Protein	2.0
TZS0304	BI642629	Metallothionein	2.1
TZS0357	BI642069	PR-10 Protein	4.2
TZS0535	BI642166	Dehydration-induced Protein	2.4
TZS0539	BI642170	DEHYDRATION STRESS-INDUCED PROTEIN.	5.0
TZS0643	BI642228	Cystein Proteinase Inhibitor	2.3
Gene expression and RNA metabolism			
CLS0034	BI677464	Glycine-rich RNA-binding protein	2.4
CLS0074	BI677494	Transcription factor	2.2
CLS0394	BI677705	Zinc Finger protein DOF	2.7
CLS0497	BI677788	C-myc binding protein MM-1-like protein	4.4
CLS0515	BI677802	RNA polymerase II subunit-like protein	2.6
CLS0842	BI678199	RNA polymerase II subunit	2.0
SWS0968	BI678966	Poly(A)-Binding Protein	2.1
Membrane transport			
CLS0285	BI677616	Plasma Membrane Intrinsic Protein	3.4
CLS1026	BI678112	Secretory Carrier-Associated Membrane Protein	2.3
TZS0226	BI642581	Sugar Transport Protein	3.9
TZS0255	BI642596	Aquaporin-like protein PIP2.	3.2

Table 2-3 (Cont'd)

CloneID	GenBank Acc.	Annotation	TZS/TZF
Miscellaneous			
CLS0081	BI677501	MTN5 Gene Precursor	3.3
CLS0162	BI677524	Germin-like Protein	2.7
CLS0274	BI677608	Trypsin Inhibitor (Serine Proteinase Inhibitor)	2.3
CLS0464	BI677761	auxin-repressed protein	2.8
CLS0536	BI677820	Core protein	2.9
CLS0669	BI678285	PI starvation-induced protein	2.1
CLS0778	BI677935	DnaJ Protein homolog	2.6
CLS0786	BI677943	Ribulose-bisphosphate carboxylase	2.2
CLS1025	BI678111	Ripening related protein	2.2
TZS0314	BI642642	Pollen Specific Protein	2.1
TZS0319	BI642056	Oxygen-evolving Enhancer Protein	3.0
TZS0363	BI642071	Tumor-related Protein	2.8
TZS0422	BI642099	Integrase	2.1
TZS0473	BI642130	auxin-repressed protein	2.7
TZS0542	BI642171	Farnesylated protein GMFP5	2.6
TZS0662	BI642242	Germin-Like Protein	3.2
TZS0717	BI642278	MAGO NASHI Protein	2.4
TZS0815	BI642398	DnaJ Protein homolog	2.2
TZS0928	BI642348	Farnesylated protein	2.1
TZS0964	BI642747	erythrocyte-binding protein	2.4
TZS1033	BI642665	Auxin-induced protein	2.1
TZS1380	BI642911	NAM (no apical meristem)-like protein	2.9
Primary metabolism			
CLS0031	BI677461	NADH Dehydrogenase (hypothetical)	2.6
CLS0279	BI677611	GDP-FUCOSE SYNTHETASE	2.0
CLS0350	BI677671	Enolase	2.1
CLS0478	BI677772	Adenosine kinase-like protein	4.4
CLS0484	BI677776	PHOSPHATIDYLINOSITOL 3-KINASE	2.2
CLS0633	BI677893	GDP-D-Mannose-4,6-Dehydratase	3.4
CLS0813	BI677966	Alcohol Dehydrogenase 7	2.4
CLS0834	BI678193	S-Adenosyl-L-Methionine Synthetase	3.5
CLS0861	BI678214	Inorganic pyrophosphatase-like protein	2.1
CLS0913	BI678259	Soluble Inorganic Pyrophosphatase	2.2
CLS0930	BI678049	Acetoacyl-CoA-thiolase	2.6
CLS0935	BI678052	Allyl Alcohol Dehydrogenase-like protein	2.8
CLS0956	BI678070	Blue copper protein	2.9
CLS0992	BI678099	Sucrose Synthase	3.7
CLS1064	BI678145	Thioredoxin	4.3
CLS1078	BI678150	Glycine hydroxymethyltransferase	5.1
SWS0965	BI678963	dTDP-D-Glucose 4,6-Dehydratase	2.5
TZS0048	BI642465	Cytochrome B5 Reductase	2.5
TZS0054	BI642472	Xyloglucan endo-1,4-beta-D-glucanase	3.8
TZS0094	BI642644	ATPase (H+transporting), ATP synthase	2.6
TZS0171	BI642542	Alpha-Amylase/Subtisin Inhibitor	2.2

Table 2-3 (Cont'd)

CloneID	GenBank Acc.	Annotation	TZS/TZF
Primary metabolism			
TZS0193	BI642554	Citrate Synthase	2.6
TZS0511	BI642151	UDP Glucose 4-Epimerase	2.1
TZS0519	BI642157	Cytochrome B5 DIF-F	2.8
TZS0533	BI642165	Acyl-CoA-Binding Protein	2.1
TZS0657	BI642237	Glyceraldehide 3-Phosphate Dehydrogenase	2.3
TZS0703	BI642268	Phosphoenolpyruvate Carboxylkinase	3.2
TZS0919	BI642344	BISPHOSPHATE NUCLEOTIDASE	2.7
TZS0971	BI642754	DNA-dependent ATPase	2.1
TZS1265	BI642829	Glutamine Synthetase PR-2	2.2
TZS1352	BI642887	Phenylalanine Ammonia Lyase	3.3
Protein synthesis and processing			
CLS0093	BI677512	Eukaryotic translation initiation factor eIF4E	3.7
CLS0216	BI677564	Polyubiquitin	2.7
CLS0272	BI677606	Legumain precursor (vicilin peptidohydrolase)	2.5
CLS0625	BI677885	Chaperonin	2.4
CLS0769	BI677928	Polyubiquitin	2.3
CLS0886	BI678236	Ferredoxin [2Fe-2S] II	2.4
CLS0916	BI678261	Ubiquitin extension protein	3.1
CLS1002	BI678027	Proteasome beta, 20S subunit PBG1	2.3
CLS1009	BI678033	Papain-like Cystein Proteinase Isoform I	2.3
TZF127	BI643016	subtilisin-like proteinase (EC 3.4.21.-)	2.7
TZS0133	BI642514	eIF4F chain p28	2.4
TZS0364	BI642072	Ubiquitin-like Protein	2.2
TZS0530	BI642162	Translation Elongation Factor 1-beta	2.6
TZS0739	BI642284	Immunophilin	2.3
TZS1359	BI642899	Formyltransferase purU	2.0
Secondary and hormone metabolism			
CLS0302	BI677629	Leucoanthocyanidin Dioxygenase-like Protein	2.4
TZS0021	BI642439	Chalcone-Flavone Isomerase	2.7
TZS0063	BI642476	Isoflavone Reductase	3.0
TZS0108	BI642499	Flavanone 3-hydroxylase	3.3
TZS0121	BI642505	Monooxygenase	2.6
TZS0216	BI642565	Monooxygenase	2.8
TZS0312	BI642638	Flavonoid 3',5'-Hydroxylase	3.1
TZS0424	BI642100	Chalcone Synthase	2.8
TZS0608	BI642208	Caffeoyl-CoA O-methyltransferase	3.0
TZS0672	BI642248	Fumarylacetoacetase	2.3
TZS0721	BI642279	Glutathione S-Transferase	2.7
TZS0751	BI642292	Flavonoid 3'-hydroxylase	2.4
TZS0799	BI642363	Glutathione S-Transferase	3.5
TZS0870	BI642387	Dihydroflavonol 4-reductase	3.5
TZS0962	BI642432	Chalcone Reductase	2.8
TZS1298	BI642857	caffeic acid O-methyltransferase-like protein	3.9

Table 2-3 (Cont'd)

CloneID	GenBank Acc.	Annotation	TZS/TZF
Signal transduction			
CLS0213	BI677561	Casein kinase II alpha subunit	2.1
CLS0214	BI677562	CSF-1 PROTEIN	2.1
CLS0252	BI677590	ADP-ribosylation factor	2.3
CLS0300	BI677627	Ethylene receptor homolog	2.0
CLS0377	BI677691	Protein Phosphatase 2C-like	4.1
CLS0503	BI677793	14-3-3 protein homolog Vfa-1433b	2.3
CLS0899	BI678246	Light-inducible protein ATLS1	2.6
TZS0111	BI642495	Serine/Threonine Protein Kinase	2.7
TZS0163	BI642537	GTP-binding Protein, ras-like	3.9
TZS0224	BI642579	Casein Kinase	2.0
TZS0258	BI642600	ADP-Ribosylation Factor	2.2
TZS0311	BI642643	Nucleoside Diphosphate Kinase I	4.0
TZS0353	BI642065	Phosphoglycerate Kinase 1	3.9
TZS0548	BI642177	GTP-Binding Protein	3.1
TZS0745	BI642288	ADP-Ribosylation Factor	2.6
TZS1034	BI642667	Signal peptidase 18 kDa subunit	2.3
Vesicular trafficking, protein sorting, secretion			
CLS0775	BI677933	Dynamin-like protein 4 (GTP-binding proteins)	2.4
TZS1114	BI642720	GTPase Activating Protein-like (GAP)	2.5

were also among genes up-regulated in TZS. Signal transduction-related genes were differentially expressed in the transition zone depending on the sampling time. For example, two genes for GTP-binding protein were up-regulated in TZS, but five genes for GTP-binding protein were up-regulated in TZF or down-regulated in TZS.

Genes up-regulated in fall (TZF)

Seventy-one percent of the 276 TZF up-regulated genes are either of unknown function or no-hit compared with 53% of the TZS up-regulated genes, indicating that more novel genes are up-regulated in the fall than in the summer (Table 2-4). The gene encoding DNA mismatch repair protein (accession no. BI679071) had 6.7-fold higher expression in TZF compared with TZS. Two genes encoding cysteine protease inhibitor (BI678216 and BI642350) were up-regulated in TZF, whereas another gene (BI642228) was highly expressed in TZS. Eight genes categorized in 'gene expression and RNA metabolism' were up-regulated in TZF compared with only six in TZS. A gene encoding reverse transcriptase, in particular, was highly up-regulated (4.9-fold) compared with TZS. Considering the TZF sampling time (November 27), these data suggest that gene expression activity in the transition zone continues during the dormancy period. Furthermore, many of the primary metabolism-related genes were also up-regulated in TZF. For example, expression of the gene for acetyl-CoA carboxylase carboxyl transferase (BI678729) was 16.5-fold higher in TZF than in TZS. Other up-regulated genes in the primary metabolism category included those encoding ATP synthase (7.2-fold), cytochrome C oxidase (4.2-fold), epoxide hydrolase (5.9-fold), ethanolaminephosphotransferase (4.8-fold), and starch phosphorylase (3.2-fold).

Table 2-4. List of up-regulated genes in transition zones harvested in fall (TZF)

Clone ID	GenBank Acc.	Annotation	TZF/TZS
Cell wall structure and metabolism			
SWS1028	BI678995	Beta-glucosidase precursor	2.6
Chromatin and DNA metabolism			
CLS0968	BI678080	DNA-Binding Protein, hypothetical	2.0
SWS1024	BI678992	histone H1.41	5.4
SWS1122	BI679071	DNA Mismatch Repair Protein	6.7
Cytoskeleton			
TZF011	BI642920	cytoskeletal protein homolog	3.6
Defense			
SWS0593	BI679110	NADPH-Cytochrome P450 Reductase	2.1
CLS0974	BI678085	ABC Transporter-like protein	2.3
SWS1009	BI678982	Elicitor-responsive gene	2.4
TZS0931	BI642350	Cystein Proteinase Inhibitor	2.4
CLS0863	BI678216	Cystatin (cysteine proteinase inhibitor)	3.4
CLS0240	BI677579	wound-inducible basic protein	3.6
TZS0462	BI642120	Oxygenase, pathogen-induced	3.8
Gene expression and RNA metabolism			
SWS1456	BI679312	Zinc Finger Protein ID1	2.3
SWS0664	BI679127	RNA Polymerase	2.5
SWS0786	BI678876	RNA Helicase ATP-Dependent	2.6
CLS1011	BI678035	RING-H2 Zinc finger protein	2.7
SWS0307	BI678530	RNA polymerase beta' subunit	2.8
CLS0258	BI677595	heat shock transcription factor-like protein	3.5
TZS0302	BI642630	Glycine-rich RNA Binding Protein	3.7
SWS0180	BI678403	Reverse Transcriptase	4.9
Membrane transport			
CLS0025	BI677455	Lectin precursor, Bark Agglutinin I	2.5
SWS1005	BI678978	Tonoplast Intrinsic Protein alpha	3.5
Miscellaneous			
CLS0517	BI677803	Tripeptidyl-Peptidase I	2.0
TZF094	BI642983	nonstructural protein 1	2.1
TZS0351	BI642064	ATAF2 Protein	2.2
SWS1013	BI678984	S1 self-incompatibility	2.3
TZS0103	BI642503	NOI Protein	2.3
TZF013	BI642922	POLYPROTEIN	2.6
SWS0481	BI678670	AINTEGUMENTA-LIKE PROTEIN	2.6
CLS0369	BI677684	Synaptobrevin homolog F10N7.40	2.7
TZS0291	BI642616	RP42	2.9
SWS0193	BI678438	Altered response to gravity	3.5
TZS1138	BI642726	Steroid-Binding (progesterone-binding) Protein	3.9
SWS0990	BI679034	Protein Translocation Complex sec61 Gamma Chain	3.9
SWS0615	BI678759	SUCROSE CARRIER SCA1.	4.7
CLS0975	BI678086	Mitochondrial processing peptidase	4.9
SWS0064	BI678369	Maturase-like Protein	7.2

Table 2-4 (Cont'd)

Clone ID	GenBank Acc.	Annotation	TZF/TZS
Primary metabolism			
SWS0982	BI678976	Lipid Transfer Protein	2.0
CLS0089	BI677508	dTDP-D-Glucose 4,6-Dehydratase	2.0
CLS0888	BI678221	Methyltransferase	2.3
SWS1286	BI679214	NADH Dehydrogenase (ubiquinone)	2.3
TZS1144	BI642729	Outer membrane lipoprotein-like	2.3
CLS1041	BI678124	Granule-bound glycogen (starch) synthase	2.3
CLS1062	BI678144	Methyltransferase	2.4
TZS0305	BI642632	Epoxide Hydrolase	2.5
SWS1067	BI679019	Aminotransferase-like protein	2.5
SWS0616	BI679124	NADH-Plastoquinone Oxidoreductase	2.6
TZS0502	BI642145	UDP Glucose 4-Epimerase	2.6
TZS0287	BI642622	Lipid Transfer Protein	2.6
SWS0564	BI678731	ATPase, Cadmium-transporting	2.8
TZF050	BI642948	GDSL-motif lipase/acylhydrolase	2.8
CLS0760	BI677920	Thioredoxin-like proteins	2.9
TZF150	BI643032	cytochrome b5 - common tobacco	3.0
TZS1266	BI642830	Starch Phosphorylase	3.2
TZS0958	BI642329	Cytochrome c oxidase subunit 6b-1	4.2
SWS0401	BI678642	Ethanolaminephosphotransferase	4.8
TZS0266	BI642604	Epoxide Hydrolase	5.9
SWS1451	BI679307	ATP Synthase beta chain	7.2
SWS0562	BI678729	Acetyl-CoA Carboxylase Carboxyl Transferase	16.5
Protein synthesis and processing			
SWS0868	BI679188	Ubiquitin-Conjugating Enzyme	2.3
TZS0297	BI642624	Aminoacylase 1	2.4
SWS0901	BI678907	Histidine-tRNA ligase	2.4
CLS0417	BI677720	Ribosomal Protein S12, Chloroplast 30S	3.1
SWS0983	BI679028	Heat Shock Protein	3.7
CLS0264	BI677601	ClpP-like Protease	3.7
SWS0197	BI678442	ClpP-like Protease	4.4
TZS1286	BI642848	Heat Shock Protein 70	4.4
Secondary and hormone metabolism			
CLS0373	BI677687	Monooxygenase	2.6
CLS0631	BI677891	Auxin-binding protein T85 precursor	2.0
TZS1040	BI642673	Protein Phosphatase	2.1
CLS0984	BI678094	GTP-binding protein sra1	2.2
CLS0433	BI677735	Calcium-dependent protein kinase	2.2
SWS1034	BI679000	Light-inducible protein ATLS1	2.2
CLS0941	BI678057	GTP-binding protein GTP13	2.4
SWS0952	BI678950	Protein Kinase PVPK-1	2.8
CLS0441	BI677741	small GTP-binding protein	3.0
SWS0491	BI678678	WD-Repeat Protein-like	3.7
TZS0467	BI642124	GTP-binding Protein	4.0
TZS0026	BI642444	GTP-Binding Protein SUP1	4.7
CLS0869	BI678222	Calmodulin-related protein	6.2
SWS0339	BI678559	Receptor Kinase	9.4

Many protein synthesis and process-related genes were also up-regulated in TZF, including genes encoding two ClpP-like proteases (4.7- and 3.7-fold), two heat shock proteins (3.7- and 4.4-fold), and ribosomal protein S12 (3.1-fold). Most importantly, a similar number (13) of signal transduction genes were up-regulated in TZF compared with 16 genes in TZS. Expression of a gene encoding for receptor kinase (BI678559) was increased 9.4-fold in TZF compared with TZS. Other up-regulated signal transduction genes include those encoding for calmodulin-related protein (6.2-fold), four GTP-binding proteins (2.2- to 4.7-fold) and WD-repeat protein-like (3.7-fold). However, only one gene, encoding for monooxygenase (BI677687), in the 'secondary and hormone metabolism' category was up-regulated in TZF. It had been thought that biosynthesis of heartwood extractives (i.e., mostly secondary metabolites) was most active in fall, rather than in summer when a total of 16 genes in the category were up-regulated. Overall, the TZF gene expression data strongly suggest that many biological activities (e.g., gene expression, protein synthesis and signal transduction) occur after trees enter dormancy.

Confirmation of microarray results

To further support my microarray results, and to confirm the seasonal expression of the secondary metabolism related genes, antisense Northern blot was carried out on 10 selected genes (the genes encoding PAL, C4H, CHS, CHR, CHI, F3H, F3'H, F3',5'H, DFR and IFR) that are involved in the phenylpropanoid metabolic pathway. When the expression of these genes was compared at three seasonal points (spring, sampled on March 5; summer, sampled on July 5; fall, sampled on November 27), the expression of all selected genes was highest in summer (Figure 2-2), which corroborates the microarray

results. In addition, the flavonoid metabolism-related genes were differentially expressed according to the sampling season with highest expression levels in the summer samples and similar expression levels in the spring and fall samples. These results suggest that the flavonoid pathway is activated and secondary metabolites are produced during the summer.

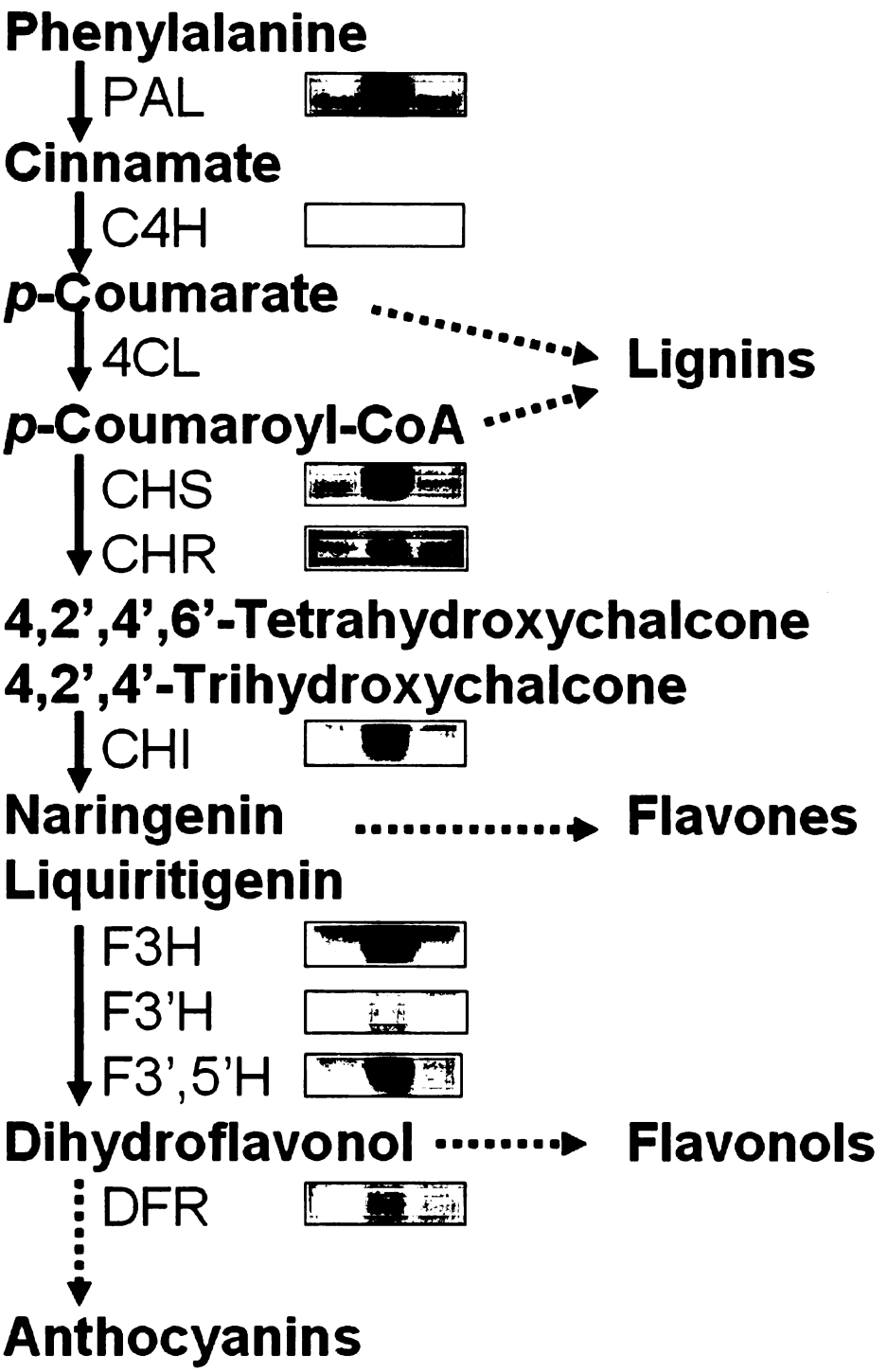


Figure 2-2. Metabolic pathway leading to phenylpropanoids

Discussion

Transverse trunk sections of mature trees usually contain two zones: a pale-colored outer zone, the sapwood, and a dark-colored inner zone, the heartwood. The pale-colored sapwood contains storage material (starch and lipids), whereas the dark color of the heartwood reflects the presence of various organic substances, the so-called extractives. The nature of the extractives determines the specific heartwood color and durability. In black locust, which is known for its resistance to wood decay (Scheffer et al., 1944), most of the extractives are phenolics. Although the biochemistry of heartwood has been studied, gene expression studies of heartwood formation are limited (Magel, 2000; Beritognolo et al., 2002; Yang et al., 2003). My work provides the first comprehensive analysis of the global gene expression changes occurring in the transition zone of a mature tree from active growth (July 5) to the start of dormancy (November 27).

During the transition from sapwood to heartwood, the ray parenchyma cells in the sapwood–heartwood transition zone undergo a form of programmed cell death (Magel, 2000). The reserve materials in these cells are converted to secondary metabolites, which may serve as defense chemicals. This complex and slow event of heartwood formation is presumably orchestrated by enzymes involved in the breakdown of storage materials, secondary metabolite biosynthesis, and senescence. The genes encoding such enzymes are thought to be induced in the transition zone at the time of heartwood formation. In the temperate zones, heartwood formation occurs in late summer to late fall and at the beginning of dormancy provided temperatures are above 5 °C (Hillis, 1987). Earlier studies have indicated that heartwood formation occurs at the time of cambial dormancy

in pine (Shain and Mackay, 1973), walnut and cherry (Nelson 1978). Studies of the cytology and coloration of the extractives suggest that heartwood formation commences in midsummer and continues in the fall and winter seasons in sugi (Nobuchi et al. 1987a) and black locust (Nobuchi et al. 1987b). Based on my survey of global gene expression changes from active growth (July 5) to the start of dormancy (November 27), I found that more than 50% of the secondary and hormone metabolism-related genes on the microarray were expressed at a twofold or greater level in July (TZS) compared with November (TZF). In addition, 31 genes in the 'primary metabolism' category were up-regulated in TZS, but only 22 genes were up-regulated in TZF. The expression levels of 53% (29 out of 55) of the signal transduction-related genes were differentially expressed in the transition zone, suggesting that the cells located in the innermost part of the trunk are affected by environmental conditions, such as seasonal change.

Higuchi (1997) proposed that the physiological processes leading to the transition from sapwood to heartwood occur in three main steps: (1) alterations in the metabolism of living parenchyma cells activated by hormonal (e.g., ethylene) and physical factors, such as reduction of water content; (2) increased metabolic activity and synthesis of heartwood extractives; and (3) cellular death, liberation of extractives stored in the vacuole, oxidation and polymerization of extractives by chemical or enzymatic processes. Among the TZS up-regulated genes, there are many representative genes that support this proposed scheme of heartwood formation. For example, three *SKP1* (suppressor of kinetochore protein) genes, a component of ubiquitin ligase complex, were up-regulated in TZS. The function of SKP1 protein is unclear, but plant SKP1 proteins involved in proteolysis play a role in cell differentiation, programmed cell death and defense

mechanisms. The *SKP1* gene is related to the regulation of proteolysis in different stages of the cell cycle (Bai et al., 1996; Deshaies, 1999). The mutant of *Arabidopsis* SKP1, *ask1*-1, shows reduced auxin responses, abnormal flowers and reduced lateral roots (Gray et al., 1999; Zhao et al., 2001). In tobacco, the *NbSKP1* gene plays an important role in the *N* gene-mediated resistance response to tobacco mosaic virus (Liu et al., 2002). The high level of expression of *SKP1* and *SKP1* homolog in TZS suggests that the transition zone undergoes programmed cell death during summer. However, because most of the SKP1 proteins studied are involved in post-translational regulation of gene expression, a proteomics-based investigation should provide more insights into the role of black locust SKP1 protein during heartwood formation. The pathogenesis-related class 10 (PR-10), which is a protein related to defense, was also highly expressed in TZS. The *PR-10* genes are induced, not only by pathogen infections, but also by a large variety of environmental stresses (Dubos and Plomion, 2001; McGee et al., 2001; Borsics and Lados, 2002). Metallothioneins (MTs) are ubiquitous low molecular weight cysteine-rich proteins and polypeptides of extremely high metal and sulfur content, and are required for heavy metal tolerance in animals and fungi. In plants, the transcripts of metallothionein genes accumulate during leaf senescence as well as heavy metal detoxification (Buchanan-Wollaston, 1997). There are three genes that are induced under dehydration conditions. One of them is homologous to *ERD15*, a dehydration-induced gene in *Arabidopsis* (Kiyosue et al., 1994). High peroxidase activity has been observed in the transition zone of various wood species: *Juglans* (Nelson, 1977), *Picea* (Estebauer et al., 1978), *Quercus* (Ebermann and Stich, 1982 and 1985), *Paulownia* (Ota et al., 1991) and *Larix* (Korori et al., 1998). Dehon et al. (2001 and 2002) noted that peroxidases in walnut trunk were

involved in heartwood formation and its brown coloration. The up-regulation of an ethylene receptor homologous gene (accession no. BI677627) suggests that ethylene signaling is involved in heartwood formation. Nelson et al. (1981) reported that ethylene production is related to heartwood formation in black walnut. In addition, ethylene is an important factor controlling heartwood formation in pine (Nilsson et al., 2002). The high level of expression of these genes related to environmental stress, defense or senescence indicates that, during the summer, the transition zone is subjected to various stressful conditions such as dehydration and senescence, and that the ray parenchyma cells in the transition zone undergo heartwood formation through the expression of these genes.

Heartwood formation requires degradation of carbohydrate storage materials at the transition zone to generate energy and substrates necessary for biosynthesis of heartwood extractives. Biochemical studies have shown that storage material is consumed during the sapwood–heartwood transition (Shah et al., 1981; Nobuchi et al., 1982, 1987a, 1987b; Nair, 1988; Kumar and Datta, 1989). Carbohydrates stored at the transition zone may be used by the cells to produce phenolics (Magel et al., 1991, 1994). Magel (2000) proposed that the sucrose synthase pathway is associated with the biosynthesis of phenolic extractives during heartwood formation and, therefore, the gene coding for sucrose synthase can be considered a marker gene of heartwood formation. Because the bulk of phenolic compound synthesis is dependent on imported carbon, the sugar transport system should be involved in heartwood formation. The genes encoding carbohydrate transporter proteins and sucrose degradation enzymes were highly expressed during the period of heartwood formation (Magel, 2000). I found that these genes were up-regulated during summer rather than in late fall, suggesting that

carbohydrate transport and break down occur during summer in the trunk wood of black locust.

Biosynthesis of secondary metabolites is a major process that leads to accumulation of heartwood extractives. Cytochrome B5 *DIF-F* gene was highly expressed in the TZS sample. This gene product stimulates the activity of specific Cyt P450s such as a flavonoid 3',5'-hydroxylase (F3,5'H) in petunia (de Vetten et al., 1999). Many secondary metabolites are toxic to the cells that produce them and are, therefore, sequestered in the vacuole once they are made. Programmed cell death of the ray parenchyma cells at the transition zone may involve the release of secondary metabolites stored in cell vacuoles. Glutathione S-transferase (GST) has been shown to play a role in vacuolar sequestration of flavonoids (Mueller et al., 2000), and GSTs may also serve as binding proteins for anthocyanin to sequestrate flavonoids into vacuoles. Two of the three genes encoding GST on the microarrays were up-regulated in TZS.

In black locust, the activities of two key enzymes of the phenylpropanoid biosynthetic pathway (PAL and CHS) increase in the transition zone compared with the outer sapwood (Magel et al., 1991; Magel and Hübner, 1997). These genes are more highly expressed in the transition zone than in the bark/cambial zone and sapwood region (Yang et al., 2003). Beritognolo et al. (2002) studied seasonal changes in the expression of several flavonoid biosynthetic genes during heartwood formation in *Juglans nigra* and found that flavonol accumulation at the sapwood–heartwood transition zone was correlated with transcript levels of key flavonoid biosynthetic genes such as those encoding CHS, DFR, and F3H. I compared the expression of these genes among three seasons (spring, summer and fall). Expression of the secondary metabolism-related genes

was highest in summer, suggesting that the biosynthesis of phenylpropanoids and other heartwood extractives in the ray parenchyma cells at the transition zone is most active during summer. Biosynthesis of flavonoid and anthocyanin has been extensively studied (Forkmann and Martens, 2001), but the biosynthetic pathways for many natural products in heart wood remain to be elucidated. For example, even though more chemical analyses have been carried out on the heartwood flavonoids of Leguminosae, to which black locust belongs, than on heartwood flavonoids of any other plant family (Rowe, 1989), the specific flavonoid biosynthetic pathway of this family has not been established and no explanation has been found for the absence of 5-hydroxylation in this family. Three major flavonoids in black locust heartwood are dihydrorobinetin (3,7,3',4',5'-pentahydroxyflavanonol), robinetin (3,7,3',4',5'-pentahydroxyflavone) and leucorobinetidin (tetrahydroxy-flavan-3,4-diol) (J.A. Duke, http://www.ars-grin.gov/duke/).

When studying heartwood formation, it is crucial to select the periods characterized by high activities in extractive accumulation and heartwood expansion. In *R. pseudoacacia*, these processes appear to be seasonally regulated and linked to tree phenology. Accumulation of extractives and radial expansion of heartwood follow the period of cambial activity and are more marked in late summer and autumn than in spring. A year-round sampling of black locust trees has just been completed to monitor the temporal dynamics of heartwood formation. Analyses of extractives and transcription profiles are currently being carried out on these samples.

Microarray analysis was performed to examine changes in global gene expression patterns in the sapwood–heartwood transition zone of mature black locust trees between

the termination of active growth and the onset of dormancy. I identified candidate genes that are likely to be involved in heartwood formation, established expression patterns of many unknown genes, and provided insights into seasonal changes in the molecular events leading to heartwood expansion and extractive formation. Although these findings need to be substantiated at the protein or phenotypic level before biologically meaningful conclusions can be drawn, they indicate the potential for bio-technological improvement of forest trees for production of value-added forest products (e.g., decay resistance, novel pharmaceuticals) and disease and insect resistance.

Literature Cited

Alizadeh, A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran and X. Yu. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403: 503-511.

Bai, C., P. Sen, K. Hofmann, L. Ma, M. Goebl, J.W. Harper and S.J. Ellwdge.1996. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell 86: 263–274.

Baqui, S. and J. Shah. 1985. Histoenzymatic studies in wood of *Acacia auriculiformis* Cunn. during heartwood formation. Holzforshchung 39: 311-320.

Baqui, S., J. Shah, R. Pandalai and I. Kothari. 1979. Histochemical changes during transition from sapwood to heartwood in *Melia azedarach*. Indian J. Exp. Biol. 17: 1032-1037.

Beritognolo, I., E. Magel, A. Abdel-Latif, J.P. Charpentier, C. Jay-Allemand and C. Breton. 2002. Expression of genes encoding chalcone synthase, flavanone 3-hydroxylase, and dihydroflavonol 4-reductase correlates with flavanol accumulation during heartwood formation in *Juglans nigra* L. Tree Physiol. 22: 291-300.

Borsics, T. and M. Lados. 2002. Dodder infection induces the expression of a pathogenesis-related gene of the family PR-10 in alfalfa. J. Exp. Bot. 53: 1831-1842.

Buchanan-Wollaston, V. 1997. The molecular biology of leaf senescence. J. Exp. Bot. 48: 181-199.

Burtin, P., C. Jay-Allemand, J.P. Charpentier and G. Janin. 1998. Natural wood colouring process in *Juglan* sp. (*J. nigra*, *J. regia* and hybrid *J. nigra* x *J. regia*) depends on native phenolic compounds accumulated in the transition zone between sapwood and heartwood. Trees 12: 258-264.

de Vetten, N., J. ter Horst, H.-P. van Schaik, A. de Boer, J. Mol and R. Koes. 1999. A cytochrome b_5 is required for full activity of flavonoid 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flowers. Proc. Natl. Acad. Sci. USA 96: 778-783.

Dehon, L., L. Mondolot, M. Durand, C. Chaliès, C. Andary and J.J. Macheix. 2001. Differential compartmentation of o-diphenols and peroxidase activity in the internal sapwood of walnut tree (*Juglans nigra* L.). Plant Physiol. Biochem. 39: 339–344.

Dehon, L., J. Macheix and M. Durand. 2002. Involvement of peroxidases in the formation of the brown coloration of heartwood in *Juglans nigra*. J. Exp. Bot. 53: 303-311.

Deshaies, R.J. 1999. SCF and cullin/ring H2-based ubiquitin ligases. Annu. Rev. Cell Biol. 15: 435–467

Dietrichs, H. 1964. Das Verhalten von Kohlenhydraten bei der. Holzforschung 18: 14-24.

Dubos, C. and C. Plomion. 2001. Drought differentially affects expression of a PR-10 protein, in needles of maritime pine (*Pinus pinaster* Ait.) seedlings. J. Exp. Bot. 52: 1143-1154.

Ebermann, R. and K. Stich. 1982. Peroxidase and amylase isoenzymes in the sapwood and heartwood of trees. Phytochemistry 21: 2401–2402.

Ebermann, R. and K. Stich. 1985. Distribution and seasonal variations of wood peroxydase activity in oak (*Quercus robur*). Wood Fiber Sci. 17: 391–396.

Eisen, M.B. and P.O. Brown. 1999. DNA arrays for analysis of gene expression. Methods Enzymol. 303: 179-205.

Estebauer, H., D. Grill and M. Zotter. 1978. Peroxidase in needles of *Picea abies* (L.) Karst. Biochemica Physiologie et Pflanzenphysiologie 172: 155–159.

Fernandes, J., V. Brendel, X. Gai, S. Lal, V. Chandler, R. Elumalai, D. Galbraith, E. Pierson and V. Walbot. 2002. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. Plant Physiol. 128: 896-910.

Forkmann, G. and S. Martens. 2001. Metabolic engineering and applications of flavonoids. Current Opinion Biotechnol. 12: 155-160.

Girke, T., J. Todd, S. Ruuska, J. White, C. Benning and J. Ohlrogge. 2000. Microarray analysis of developing *Arabidopsis* Seeds. Plant Physiol. 124: 1570-1581.

Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M. Loh, J.R. Downing, and M.A. Caligiuri. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286: 531-537.

Gray, W.M., J.C. del Pozo, L. Walker, L. Hobbie, E. Risseeuw, T. Banks, W.L. Crosby, M. Yang, H. Ma and M. Estelle. 1999. Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. Genes Dev. 13: 1678–1691.

Hauch, S. and E. Magel. 1998. Extractable activities and protein content of sucrose-phosphate synthase, sucrose synthase and neutral invertase in trunk tissues of *Robinia pseudoacacia* L. are related to cambial wood production and heartwood formation. Planta 207: 266-274.

Higuchi, T. 1997. Biochemistry and Molecular Biology. Springer Verlag: Berlin.

Higuchi, T.,Y. Onda and Y. Fujimoto. 1969. Biochemical aspects of heartwood formation with special reference to the site of biogenesis of heartwood compounds. Wood Res. 48: 15-30.

Hillinger, C., W. Höll and H. Ziegler. 1996a. Lipids and lipolytic enzymes in the trunkwood of *Robinia pseudoacacia* L. during heartwood formation. I. Radial distribution of lipid classes. Trees 10: 366-375.

Hillinger, C., W. Höll and H. Ziegler. 1996b. Lipids and lipolytic enzymes in the trunkwood of *Robinia pseudoacacia* L. during heartwood formation. II. Radial distribution of lipases and phospholipases. Trees 10: 376-381.

Hillis, W.E. 1987. Heartwood and tree exudates. Berlin, New York, Springer-Verlag.

Hillis, W.E. 1996. Formation of robinetin crystals in vessels of *Intsia* species. IAWA J. 17: 405-419.

Hillis, W. E. and M. Hasegawa. 1963. The formation of polyphenols in trees. I. Administration of 14-C-glucose and subsequent distribution of radioactivity. Phytochem. 2: 195-199.

Hillis, W. E. and Y. Yazaki. 1973. Wood polyphenols of *Eucalyptus polyanthemos*. Phytochem. 12: 2969-2977.

Höll, W. and K. Lendzian. 1973. Respiration in the sapwood and heartwood of *Robinia pseudoacacia*. Can J. Bot. 52: 727-734.

Kiyosue, T., K. Yamaguchi-Shinozaki and K. Shinozaki. 1994. ERD15, a cDNA for a dehydration-induced gene from *Arabidopsis thaliana*. Plant Physiol. 106: 1707.

Korori S.A.A., H. Pichorner and R. Ebermann. 1998. Seasonal alterations of peroxidase and catalase isoenzymes in branches and seeds on three different species of *Larix*. Plant Peroxidase Newsletter 3: 12–17.

Kummer, A. and S. Datta. 1989. Histochemical and histoenzymological changes during heartwood formation in a timber tree *Shorea robusta*. Proc. Indian Sci. (Plant Sci.) 99: 21-27.

Liu, Y., M. Schiff, G. Serino, X. Deng and S. Dinesh-Kumar. 2002. Role of SCF ubiquitin-ligase and the COP9 signalosome in the *N* gene–mediated resistance response to tobacco mosaic virus. Plant Cell 14: 1483-1496.

Magel, E. 2000. Biochemistry and physiology of heartwood formation. In Molecular and Cell Biology of Wood Formation. Eds. R. Savidge, J. Barnett and R. Napier. BIOS Scientific Publishers, Oxford, pp363-376.

Magel, E. and B. Hübner. 1997. Distribution of phenylalanine ammonia lyase and chalcone synthase within trunks of *Robinia pseudoacacia* L. Bot. Acta. 110: 314-322.

Magel, E., A. Drouet, A. Claudot and H. Ziegler. 1991. Formation of heartwood substances in the stemwood of *Robinia pseudoacacia* L. I. Distribution of phenylalanine ammonium-lyase and chalcone synthase across the trunk. Trees 5: 203-207.

Magel, E., C. Jay-Allemand and H. Ziegler. 1994. Formation of heartwood substances in the stemwood of *Robinia pseudoacacia* L.: II. Distribution of nonstructural carbohydrates and wood extractives across the trunk. Trees 8: 165-171.

McGee, J., J. Hamer and T. Hodges. 2001. Characterization of a PR-10 pathogenesis-related gene family induced in rice during infection with Magnaporthe grisea. Mol. Plant Microbe Interact. 14: 877-886.

Mueller, L., C.D. Goodman, R.A. Silady and V. Walbot. 2000. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration is a flavonoid-binding protein. Plant Physiol. 123: 1561-1570.

Nair, M. 1988. Wood anatomy and heartwood formation in Neem (*Azadirachata indica* A. Juss.). Bot. J. Linn. Soc. 97: 79-90.

Nelson, N.D., W.J. Rietveld and J.G. Isebrands. 1981. Xylem ethylene production in five black walnut families in the early stages of heartwood formation. For. Sci. 27: 537-543.

Nelson, N.D. 1977. Xylem ethylene, phenol-oxidizing enzymes, and nitrogen and heartwood formation in walnut and cherry. Can. J. Bot. 56: 626–634.

Nelson, N.D. 1978. Xylem ethylene, phenol-oxidising enzymes and nitrogen and heartwood formation in walnut and cherry. Can. J. Bot. 56: 626-634.

Nelson, N.D., W.J. Rietveld and J. Isebrands. 1981. Xylem ethylene production in five black walnut families in the early stages of heartwood formation. For. Sci. 27: 537-543.

Nilsson, M., S. Wikman and L. Eklund. 2002. Induction of discolored wood in Scots pine (*Pinus sylvestris*) Tree Physiol. 22: 331-338.

Nobuchi, T., K. Kuroda, R. Iwata and H. Harada. 1982. Cytological study of the seasonal features of Sugi (*Cryptomeria japonica* D. Don). Mokuzai Gakkaishi 28: 669-976.

Nobuchi, T., K. Takai and H. Harada. 1987a. Distribution of heartwood phenols in the trunk of Sugi (*Cryptomeria jaonica* D. Don) and partial characterization of heartwood formation. Mokuzai Gakkaishi 33: 88-96.

Nobuchi, T., N. Tokuchi and H. Harada. 1987b. Variability of heartwood formation and cytological features in broadleaved trees. Mokuzai Gakkaishi 33: 596-604.

Perou, C.M., T. Sorlie, M.B. Eisen, M. Van de Rijn, S. Jeffrey, C.A. Rees, J.R. Pollock, D.T. Ross, H. Johnsen and L.A. Akslen. 2000. Molecular portraits of human breast tumors. Nature 406: 747-752.

Pinkel, D., R. Segraves, S. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W.-L. Kuo, C. Chen and Z. Zhai. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. 20: 207-211.

Reymond, P., H. Weber, M. Damond and E. Farmer. 2000. Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. Plant Cell 12: 707-720.

Rowe, J. 1989. Natural products of wood plants I. Berlin, New York, Springer-Verlag. Schaffer, R., J. Landgraf, M. Accerbi, V. Simon, M. Larson and E. Wisman, E. 2001. Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. Plant Cell 13: 113-123.

Scheffer, T., H. Lachmund and H. Hopp. 1944. Relation between hot-water extractives and decay resistance of black locust wood. J. Agri. Res. 68: 415-426.

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

Shah, J., S. Baqui, R. Pandalai and K. Patel. 1981. Histochemical changes in *Acacia nilotica* during transition from sapwood to heartwood. IAWA Bulletin 2: 31-36.

Shain, J. and J. Mackay. 1973. Seasonal fluctuations in respiration of aging xylem in relation to heartwood formation in *Pinus radiata*. Can J. Bot. 51: 737-741.

Swidzinski, J.A., L.J. Sweetlove and C.J. Leaver. 2002. A custom microarray analysis of gene expression during programmed cell death in *Arabidopsis thaliana*. Plant J. 30: 431-446.

Wang, E., L.D. Miller, G.A. Ohnmacht, E.T. Liu and F.M. Marincola. 2000. High-fidelity mRNA amplification for gene profiling. Nat. Biotech. 18: 457-459.

Wardrop, A. and J. Cronshaw. 1962. Formation of phenolic substances in ray paranchyma of angiosperms. Nature 193: 90-92.

Yang, J., S. Park, D.P. Kamdem, D.E. Keathley, E. Retzel, C. Paule, V. Kapur and K.-H. Han. 2003. Novel insight into trunk-wood gene expression profiles in a hardwood tree species, *Robinia pseudoacacia*. Plant Mol. Biol. 52: 935-956.

Zhao, D., Q. Yu, M. Chen and H. Ma. 2001. The *ASK1* gene regulates B function gene expression in cooperation with *UFO* and *LEAFY* in *Arabidopsis*. Development 128: 2735–2746.

CHAPTER 3

Functional characterization of allantoinase genes from Arabidopsis and a nonureide-type legume black locust

ABSTRACT

The availability of nitrogen is a limiting factor for plant growth in most soils. Allantoin and its degradation derivatives are a group of soil heterocyclic nitrogen compounds that play an essential role in the assimilation, metabolism, transport, and storage of nitrogen in plants. Allantoinase is a key enzyme for biogenesis and degradation of these ureide compounds. Here, I describe the isolation of two functional allantoinase genes, *AtALN* (Arabidopsis allantoinase) and *RpALN (Robinia pseudoacacia* allantoinase), from Arabidopsis and black locust *(Robinia pseudoacacia*). The proteins encoded by those genes were predicted to have a signal peptide for the secretory pathway, which is consistent with earlier biochemical work that localized allantoinase activity to microbodies and endoplasmic reticulum (Hanks et al., 1981). Their functions were confirmed by genetic complementation of a yeast mutant *(dal*1) deficient in allantoin hydrolysis. The absence of nitrogen in the medium increased the expression of the genes. In Arabidopsis, the addition of allantoin to the medium as a sole source of nitrogen resulted in the up-regulation of the *AtALN* gene. The black locust gene *(RpALN*) was differentially regulated in cotyledons, axis, and hypocotyls during seed germination and seedling growth, but was not expressed in root tissues. In the trunk wood of a mature black locust tree, the *RpALN* gene was highly expressed in the bark/cambial region, but had no detectable expression in the sapwood or sapwood-heartwood transition zone. In

addition, the gene expression in the bark/cambial region was up-regulated in spring and fall when compared with summer, suggesting its involvement in nitrogen mobilization.

Introduction

Nitrogen is a key component of plant metabolism and its availability is often a limiting factor for plant growth in most soils. Heterocyclic nitrogen compounds (i.e. purines, pyrimidines, and their degradation products) represent major sources of soil organic nitrogen (Schulten and Schnitzer, 1998). Among these, allantoin (ALN) and its degradation product allantoic acid (ALA) are nitrogen-rich organic compounds with a C:N ratio of 1:1, and they play an essential role in the assimilation, metabolism, transport, and storage of nitrogen in plants (Schubert and Boland, 1990). In addition, they serve as effective carriers of the biologically fixed nitrogen in ureide-type legumes, and provide nitrogen storage with minimal expense of reduced carbon. For example, these compounds constitute 70% to 80% (w/v) of the organic nitrogen in the xylem sap of nodulated soybean *(Glycine max*; McClure and Israel, 1979). In a tree legume (black locust, *Robinia pseudoacacia*), about 3% (w/w) of xylem nitrogen was ureide (Atkins et al., 1991). Furthermore, ALN and ALA are central metabolites in the seasonal nitrogen cycle of perennial species such as maple *(Acer* spp.) and comfrey *(Symphytum officinale*; for review, see Schubert and Boland, 1990), indicating their key role in nitrogen storage and cycling. In legumes, the organic nitrogen fixed by soil bacteria (e.g. *Rhizobium* sp.) can be transported to the other parts of the plant as amides (Asn and Gln; amide-type legume) or as ureides (ALN and ALA; ureide-type legume).

The major route for ALN biogenesis is the purine oxidation pathway (also called

ureide pathway and ALN degradation pathway; Figure 3-1). The first step in this pathway is the degradation of nucleic acid purine moieties (adenine and guanine) to uric acid. After two consecutive oxidation reactions by urate oxidase and hydroxy-isourate hydrolase, the uric acid is converted to ALN (Raychaudhuri and Tipton, 2002). Allantoinase (ALN amidohydrolase, EC 3.5.2.5) catalyzes the hydrolysis of ALN to form allantoic acid, which is a key reaction for biogenesis and the degradation of ureides (Vogels et al., 1966; Noguchi et al., 1986). The resulting ALA is then further metabolized to ammonium, urea, and glyoxylate (Muñoz et al., 2001). Allantoate degradation can be catalyzed by allantoate amidohydrolase (EC 3.5.3.9) or allantoate amidinohydrolase (EC 3.5.3.4) to produce ureidoglycolate, which is metabolized to glyoxylate by ureidoglycolate urea-lyase (EC 4.3.2.3) or ureidoglycolate amidohydrolase (EC 3.5.3.19). Allantoate amidohydrolases and ureidoglycolate amidohydrolases generate ammonium, whereas allantoate amidinohydrolase and ureidoglycolate urea-lyase release urea. This pathway serves different roles and is evolutionarily distinct in plants, animals, and microorganisms. It is primarily used for salvage or excretion of nitrogen from purines in animals (Campbell and Bishop, 1970; Stryer, 1988). On the other hand, microorganisms use it to extract nitrogen from a variety of sources in the external environment (Cooper, 1980). Although its main function in animals is nitrogen excretion, many leguminous plant species use this pathway to store and recycle nitrogen. The pathway for ureide degradation in nonlegume plants has not been well documented. Recently, Desimone et al. (2002) demonstrated that Arabidopsis, a nonlegume model species, could take up and use ALN as a sole nitrogen source. This requires enzymatic degradation of the mobilized ALN inside the cells. The first step of the catabolism is catalyzed by allantoinase.

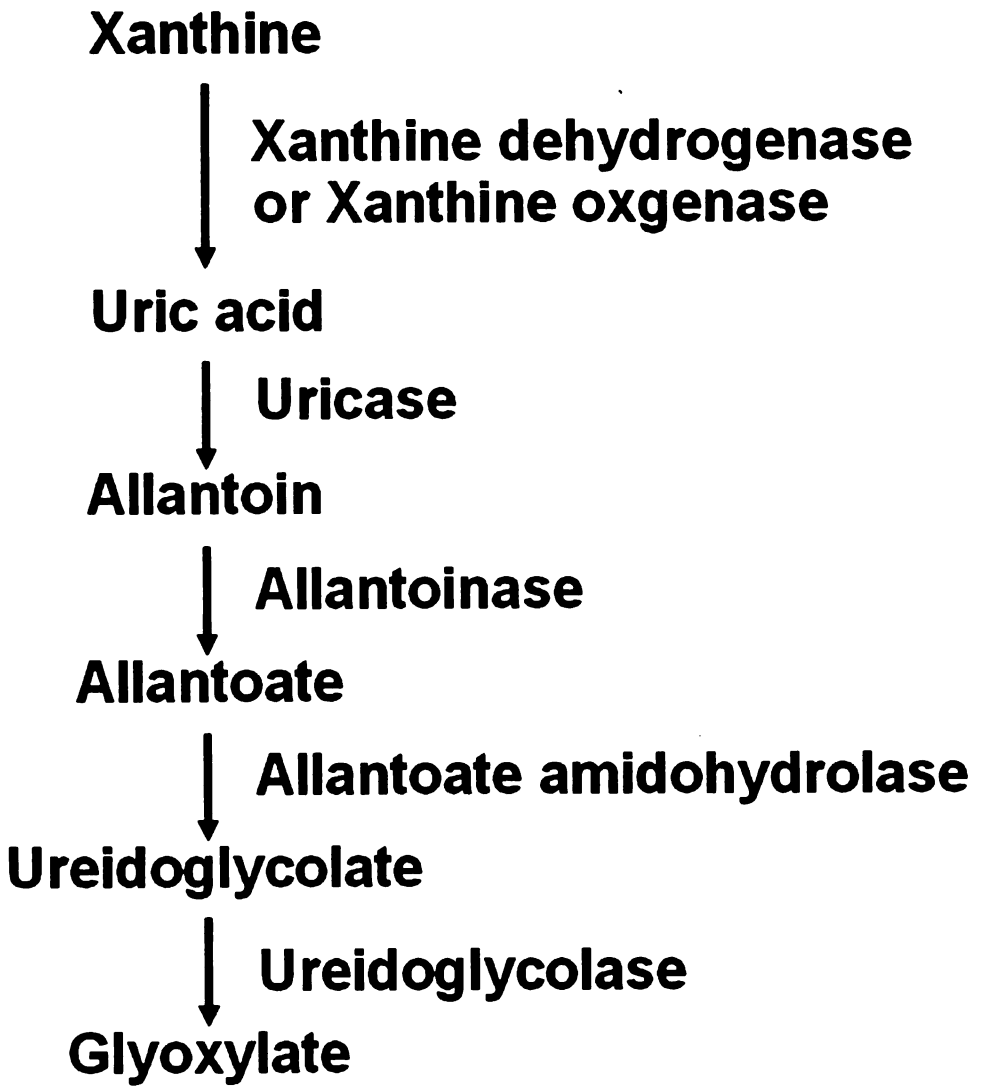


Figure 3-1. Catabolism of ureide in plants

Allantoinase is present in a wide variety of organisms, including animals, bacteria, fungi, and plants (for review, see Schubert and Boland, 1990). It is important for ureide biogenesis and degradation. Despite the long history of allantoinase study in plants (mainly legumes), my understanding of the physiological significance of the enzyme in plant growth and development is still limited. Molecular analysis of allantoinase genes is the first step in my approach to elucidate the relevance of ALN for plant nutrition. The genes encoding allantoinase have been cloned from yeast *(Saccharomyces cerevisiae*; Buckholz and Cooper, 1991) and bullfrog *(Rana catesbeiana*; Hayashi et al., 1994), but no plant allantoinase genes have been characterized yet. In this report, I isolated and characterized functional plant allantoinase genes from a tree legume (black locust, *Robinia pseudoacacia*) and Arabidopsis (nonlegume model species). The black locust allantoinase gene was discovered as a part of my trunk wood expressed sequence tag (EST) project, and the Arabidopsis gene was identified by a homologous search with the black locust gene. Their functions were confirmed by genetic complementation of a yeast mutant defective in allantoinase function. Differential expression of the genes was investigated with regard to the stage of development, seasonal variation, presence or absence of nitrogen, and different tissues.

Materials and Methods

Plant Materials and Culture Conditions

Arabidopsis (Colombia ecotypes) and black locust *(Robinia pseudoacacia*) seedlings were grown in petri dishes under light conditions (16 h of light per day)at 25°C. For black locust seed germination and seedling growth, seeds were heated in boiling

water for 1 min, soaked in sterile water for 1 h, and transferred into glass jars containing 0.7% (w/v) agar (without nitrogen) or 0.7% (w/v) agar plus 10 mm ammonium nitrate (with nitrogen). For Arabidopsis plants, seeds were subjected to cold treatments for 3 d in sterile water, and then were transferred onto Murashige and Skoog (1962) medium with 2% (w/v) Glc and 10 mm ammonium nitrate or 10 mm ALN (ALN medium) as a nitrogen source. All procedures were performed under sterile conditions.

Cloning, Multiple Alignment, and Phylogenetic Analysis

I used the cDNA library prepared from the 4- to 6-d-old seedlings of black locust to obtain the full-length cDNA clone of black locust allantoinase. About 100,000 plaques were screened with the partial cDNA clone (GenBank accession no. BI643000) and were labeled with [α-^{32}P]dCTP. Five positive phages were randomly selected and inserts were recovered as pBluescript SK phagemids according to the manufacture's instruction (Stratagene, Cedar Creek, TX). All of them were sequenced and shown to be identical clones. I used the sequenced black locust full-length cDNA clone to search for other plant homologous clones on the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and the TIGR (http://www.tigr.org) sequence databases. The full-length cDNA clone of Arabidopsis allantoinase gene (TAIR clone no. 3989737) was obtained from TAIR. The multiple alignment of amino acid sequences was carried out by CLUSTALW (Thompson et al., 1994) using default parameters, and the phylogenetic tree was created by DRAWTREE (PHYLIP unrooted phylogenetic tree, Felsenstein, 1989).

Identification of ataln T-DNA Insertional Mutant

Seven T-DNA insertion mutants (SALK-000325, SALK-000327, SALK-000282, SALK-013427, SALK-013985, SALK-013986, and SALK-142607) of the AtALN gene generated by the Salk Institute Genomic Analysis Laboratory (http://signal.salk.edu/) were obtained from the Arabidopsis Biological Resource Center (Ohio State University, Columbus). The seeds were planted on kanamycin plates and the resulting kanamycin-resistant plants were transferred to soil and seeds were harvested separately from individual plants. Subsequently, the seeds were screened on Murashige and Skoog media supplied with 10 mm ALN as a sole nitrogen source. After an additional phenotypic selection, I randomly chose one mutant line *(ataln m2-1)* that did not grow well on ALN media for further analyses. To confirm the mutant line as a homozygote, PCR was performed with the genomic DNA of *ataln m2-1* mutants using gene-specific oligo-nucleotides (forward1, 5'-AGAGATATGGAGAGAACTTTGCTT-3'; forward2, 5'-GGAAGGAGACATTGATATGCTGAG-3'; reverse, 5'-AAGTTAAGTAGTTGCAAGTTGCAG-3') or one gene-specific primer and one left-border-specific primer. As expected, the progenies of the homozygous mutant line showed the mutant phenotype and kanamycin resistance with no segregation. The mutant line did not have any noticeable morphological alterations or growth defects under the normal condition.

RNA Isolation

Using the Trizol reagent (Invitrogen, Carlsbad, CA), total RNAs were isolated from seedlings of Arabidopsis and black locust. RNA extraction from inner wood was

performed as previously described (Yang et al., 2003). Briefly, mature trees (20 cm diameter at breast height) were felled using a chain saw and were made into 25-cm-long logs. The logs were immediately placed on ice and brought back to a wood shop, where thin cookies (approximately 1 cm thick) were made using a table saw. The cookies were immediately submerged in RNA extraction buffer (20 mM EDTA, pH 8.0, 50 mM Tris-HCl, pH 8.0, 0.2% [w/v] SDS, and 10 mM β-mercaptoethanol). While submerged, the trunk wood sections (bark/cambial region, sapwood, and transition zone) were carved out by using a chisel and hammer, and were washed with RNase Away solution (Invitrogen). The isolated sections (approximately 1-cm 3 cubicles) were frozen in liquid nitrogen and stored at -80°C until needed. For RNA isolation, the frozen samples were first ground in a blender and then further ground to a fine powder using a mortar and pestle. The grounded sample was first passed through the shredder column of DNeasy Maxi kit (Qiagen, Hilden, Germany), and was then subjected to total RNA isolation using the RNeasy Maxi kit (Qiagen) and cleaned up by RNeasy Mini kit (Qiagen).

Northern Blots

Total RNA (5 or 10 μg) was separated on a 1.2% (w/v) agarose gel containing 1X MOPS buffer with 2.2 M formaldehyde at 60 V for 3.5 h. The RNA was then washed twice in water and transferred by capillary action to a nylon membrane (Hybond N+; Amersham Pharmacia Biotech, Piscataway, NJ). Membranes were prehybridized for 2 h at 42°C in 50% (v/v) formamide, 6X SSC, 5X Denhardt's solution, 0.5% (w/v) SDS, and 20 μg mL^{-1} sonicated salmon sperm DNA. Hybridizations were performed overnight at 65°C after adding the randomly primed ^{32} P-labeled cDNA probe in the prehybridization

buffer. Membranes were washed at 65°C for 10 min each with washing solution I (2.0X SSC and 0.1% [w/v] SDS), washing solution II (0.5X SSC and 0.1% [w/v] SDS), or washing solution III (0.1X SSC and 0.1% [w/v] SDS). Random-labeled probes were made with a kit (Rediprime II; Amersham Pharmacia Biotech) using an internal fragment from the full-length cDNA of *AtALN* and *RpALN*1.

RT-PCR

One microgram of total RNA, isolated from wild-type and mutant plants, was reverse transcribed using a random primer in a 20 μL reverse transcription reaction as recommended by the manufacturer (SuperScript II reverse transcriptase; Invitrogen). One microliter of the first-strand cDNA reaction was used as a template for PCR amplification with specific oligo-nucleotide primers for *AtALN* (forward, 5'-AGAGATATGGAGAGAACTTTGCTT-3'; reverse, 5'-AAGTTAAGTAGTTGCAAGTTGCAG-3'). As an internal control, plant 18s primers (Ambion, Austin, TX) were used. The PCR products were loaded on ethidium bromide-stained agarose gels.

Antisense Northern Blot

Antisense northern-blot analysis was performed as previously described (Yang et al., 2003). For the generation of antisense RNAs, total RNAs (2 μg) were amplified using Message AmpTM aRNA kit (Ambion). I began by synthesizing first stand cDNAs from the total RNAs of seedling, bark/cambial region, sapwood, or transition zone, and the first cDNAs were used as templates for the synthesis of second cDNAs. Finally, aRNAs

from the second cDNAs were generated by in vitro transcription (amplification). About 200 ng of aRNAs was separated in a formamide agarose gel and transferred onto the nylon membrane using the capillary transfer method. The membrane was then hybridized with an isotope-labeled probe. The signal was exposed and detected on an X-ray film. Histon 3.2, which showed constitutive expression in black locust tissues (Yang et al., 2003), was used as a control.

Genetic Complementation of Yeast (Saccharomyces cerevisiae) Mutant with RpALN and AtALN Genes

To conduct the functional complementation test of RpALN and AtALN in yeast, I found the yeast allantoinase (DAL1) gene deletion mutant on the *Saccharomyces* Genome Deletion Project web page (http://wwwsequence.stanford.edu/ group/ yeast_deletion_project/ deletions3.html), and bought the homozygous diploid strain of *dal1* deletion mutant (item no. 4035962) from American Type Culture Collection (Manassas, VA). The homozygous diploid of the mutant is in the background of BY4743 (BY4741: MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0/BY4742: MAT_his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0). The deficiency of the mutant, using ALN as a sole nitrogen source, was determined on solid plates containing different concentrations of ALN as sole nitrogen sources. Optimal selective conditions were found at 2 g L^{-1}. The *dal1* mutant was transformed with an expression vector pYES2 (Invitrogen) harboring a full-length cDNA of RpALN or AtALN gene. PCR products generated from the full-length cDNA of *RpALN* or *AtALN* gene were digested by *Ba*mHI and *Xb*aI and ligated into the *Ba*mHI-*Xb*aI site of plasmid pYES2. At the first, transformants were selected on solid plates that

were composed of a yeast nitrogen base, 2% (w/v) dextrose, 0.5% (w/v) ammonium sulfate, and a synthetic complete supplement mixture minus uracil (Qbiogene, Carlsbad, CA). RpALN or AtALN transformants were then confirmed by PCR analysis using gene-specific primers. After uracil selection, each transformant was reselected on a solid plate containing yeast nitrogen base, 2% (w/v) Gal, 1% (w/v) raffinose, and 0.2% (w/v) ALN (Sigma, St. Louis) as a nitrogen source, and 1/100-fold reduced synthetic complete supplement mixture minus uracil to supply Leu and His. The colonies capable of growing were reselected under the same conditions. As a negative control, the mutant was transformed with the empty vector. After a single colony from each transformed mutant was chosen, growth complementation tests were performed using ALN as a sole nitrogen source (Desimone et al., 2002).

Protein Extraction and Enzyme Assay

Crude extracts for ALN assays of RpALN and AtALN were prepared according to the method described by Bell and Webb (1995) with slight modification. Plant samples (0.5–1.0 g) were ground with 4X volumes of 20 mM Tris-HCl and 500 mM NaCl, pH 7.5, 100 μM dithiothreitol, 10 mM EDTA, 0.1% (w/v) polyvinyl-polypyrrolidone, and a mixture of protease inhibitors, including phenylmethylsulfonyl fluoride (2 mM), aprotini (10 μg mL^{-1}), leupeptin (10 μg mL^{-1}), and pepstatin (10 μg mL^{-1}) in a chilled mortar, with a pestle. Extracts were centrifuged at 30,000*g* for 30 min at 4°C and were filtered through two layers of Miracloth (Calbiochem, San Diego), and the supernatant was collected. Then, the supernatant was filtered through 0.2-μm filters (Millipore, Bedford, MA) to remove particulate matter. Extracts were kept on ice or -20°C until they were used. The

assay for allantoinase activity was based on the procedure reported by Schubert (1981) and Romanov et al. (1999). The enzyme extract was incubated for 20 min at room temperature with 10 mM ALN in 50 mM Tris-HCl buffer. The pH was set to 7.5. Reaction solution (0.5 mL) was then treated with 0.125 mL of 0.2 N HCl, and was heated in a boiling water bath for 4 min to degrade ALA to glyoxylate and urea. After cooling for 30 sec, 0.4 mL of concentrated HCl and 0.125 mL of potassium ferricyanide solution (16 mg mL^{-1}) were added and incubated at 37°C for 30 min. The total reaction solution was immediately applied to measure the A520 arising from the final product of this reaction, glyoxylate diphenylformazan. All assays were replicated at least twice, and controls were included to account for endogenous ALA in the enzyme extract and non-enzymatic degradation of ALN.

Results

Allantoinase Genes from Black Locust and Arabidopsis

In my previous EST analysis from the trunk wood of black locust (Yang et al., 2003), I found an EST that was homologous to other reported allantoinase genes such as those found in *Escherichia coli* (Kim et al., 2000), yeast (Buckholz and Cooper, 1991), and bullfrog (Hayashi et al., 1994). To obtain a full-length cDNA clone, I used 4- to 6-d-old seedlings to construct a cDNA library, and then screened them using the EST as a probe. The full-length cDNA, named *RpALN*, was comprised of 1,820 bp with an open reading frame of 1,536 bp, which encodes a 56,375-D protein with 512 amino acid residues. The protein was predicted by TargetP V1.0

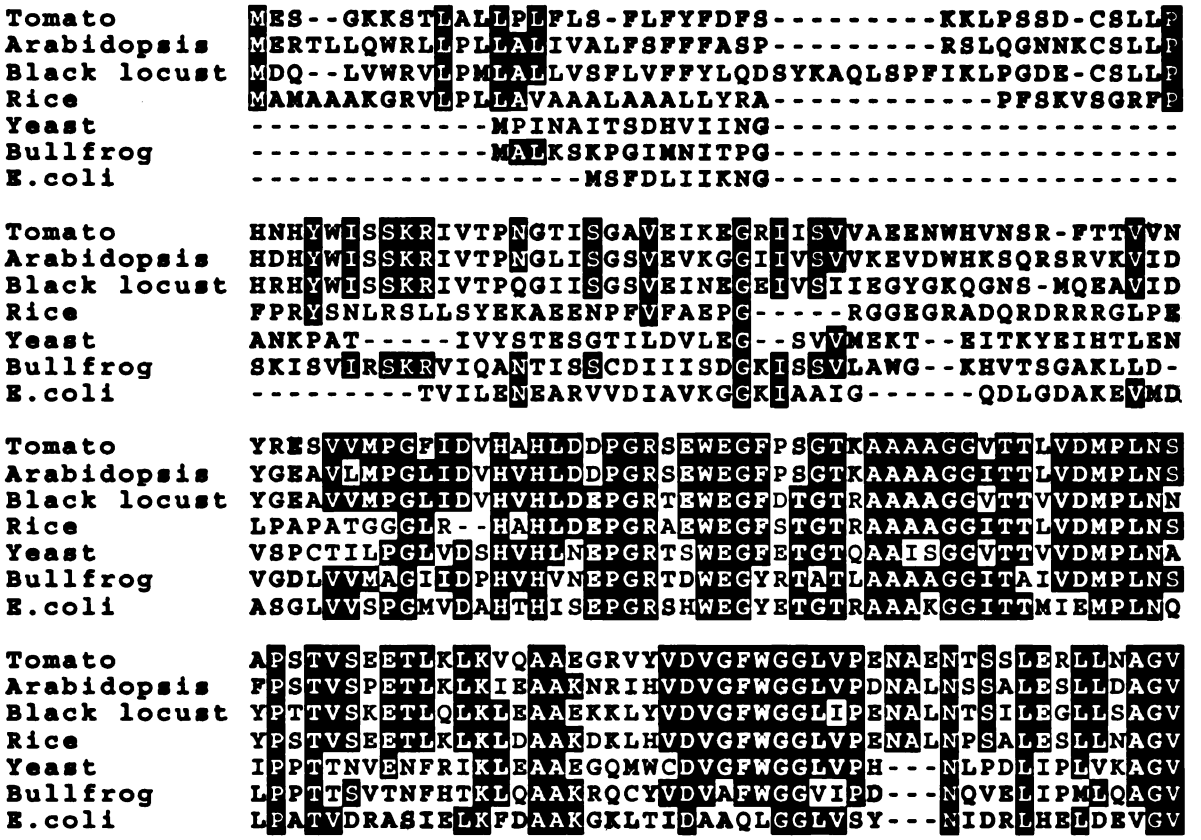


Figure 3-2. Alignment of allantoinase amino acid sequences

Black locust putative allantoinase, Arabidopsis putative allantoinase and tomato putative allantoinase were compared with published allantoinases from *E. coli*, yeast and bullfrog. Completely conserved amino acids in all species are marked in black; identical amino acids in dark gray; similar amino acids in light gray; different amino acids in white. The conserved residues responsible for metal biding are indicated with asterisks.

(http://www.cbs.dtu.dk/services/TargetP/; Nielsen et al., 1997; Emanuelsson et al., 2000) to contain a signal peptide for the secretory pathway and a cleavage site at the 30th amino acid. The *RpALN* sequence information was used to carry out BLASTN searches against public databases (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov and The Institute for Genomic Research [TIGR], http://www.tigr.org). I found an Arabidopsis homologous gene at the locus (accession no. AT4G04955). The BLASTX analysis with the RpALN amino acid sequence further confirmed its homology with an E value of 0.0 (BLASTX score = 670). I obtained a full-length cDNA clone (The Arabidopsis Resource [TAIR] clone no. 3989737) for this gene from TAIR (http://www.Arabidopsis.org). The size of the inserted gene was 1,552 bp with an open reading frame of 1,518 bp, encoding a protein with 506 amino acid residues and a calculated molecular mass of 55,438 D. Again, this protein was predicted to have a signal peptide for the secretory pathway and a cleavage site at the 26^{th} amino acid. The Arabidopsis genome has only one copy of this gene. The deduced amino acid sequences of the RpALN and AtALN share highly conserved residues that are responsible for metal binding that were found in other reported ALN proteins (Figure 3-2).

Function of RpALN and AtALN

After cloning the *RpALN* and *AtALN*, I confirmed their functions via genetic complementation of a yeast mutant *(dal*1) deficient in ALN hydrolysis. This *dal1* mutant grows slowly with ALN as a sole nitrogen source because no enzymatic degradation of ALN to urea can occur (Winkler et al., 1988). However, both transformants carrying *RpALN* or *AtALN* cDNA gene grew well (Figure 3-3), suggesting the enzymes encoded

by the plant genes could catalyze the hydrolysis of ALN to allantoic acid. Recently, Desimone et al. (2002) have shown that an Arabidopsis ureide permease (AtUPS1) could genetically complement an ALN uptake-deficient yeast mutant.

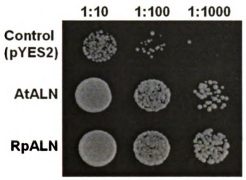


Figure 3-3. Functional complementation of a yeast allantoinase mutant
The dal1 mutant was transformed with the empty vector (pYES2), with pYES2 carrying *AtALN*, or with pYES2 carrying *RpALN*. Three cell dilutions (1:10, 1:100, and 1:1,000) were plated on yeast minimal medium containing 0.1% allantoin as sole nitrogen sources and incubated for 5 days at room temperature.

Arabidopsis Mutant ataln

Consisting of 15 exons and 14 introns, the *AtALN* gene is located on chromosome 4 of the Arabidopsis genome (Figure 3-4A). To examine the phenotypic consequence of the *AtALN* gene mutation, I screened T-DNA insertion mutants received from the Salk Institute Genomic Analysis Laboratory (http://signal.salk.edu/). First, single-copy T-DNA insertion was confirmed by 3:1 (resistant:sensitive) segregation in kanamycin

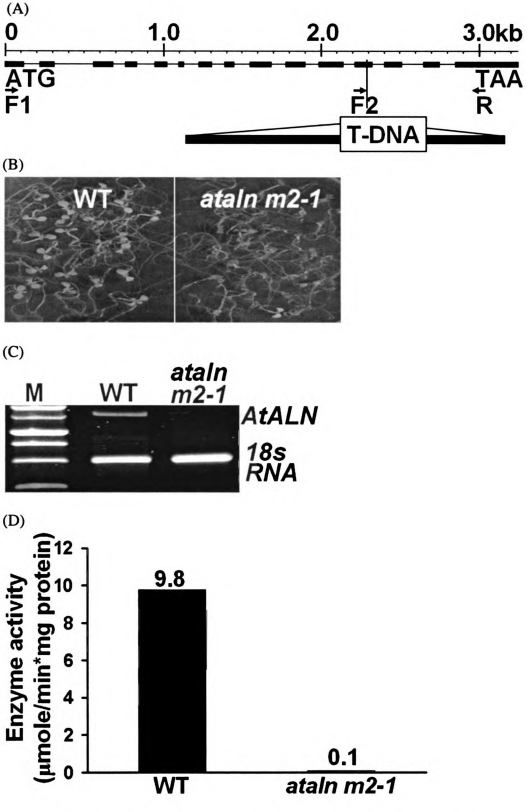


Figure 3-4. Identification of T-DNA insertional mutants of Arabidopsis allantoinase

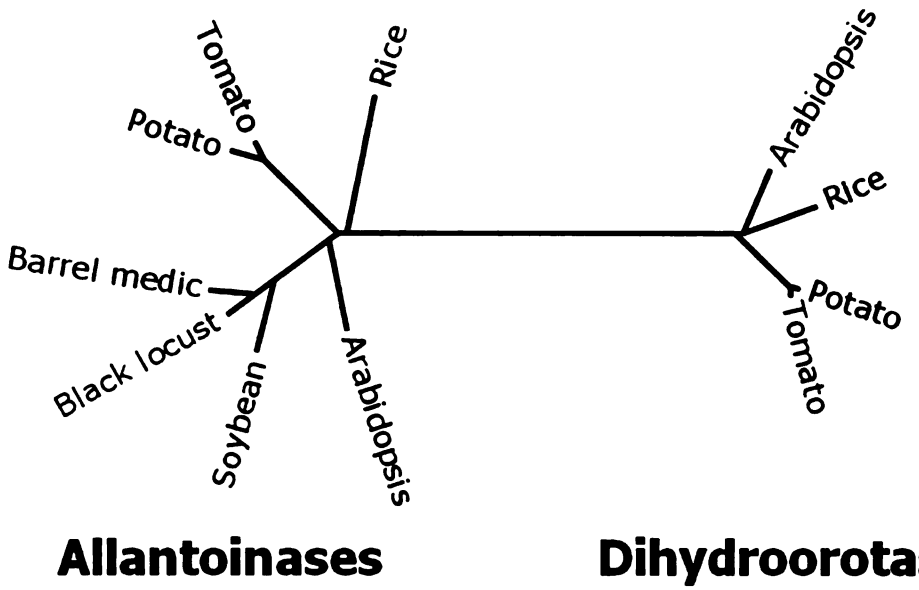
selection. From the SALK-000327 line, I identified one mutant line *(ataln m2*-1) that did not grow well on Murashige and Skoog medium (Murashige and Skoog, 1962), which was composed of 10 mm ALN, instead of ammonium nitrate (Figure 3-4 B). I confirmed the T-DNA insertion in the Arabidopsis genome using PCR analysis with the gene-specific primer sets, which failed to detect stable *AtALN* transcripts in the mutant plants when compared with wild-type plants (Figure 3-4 A and C). As expected, the progenies of the homozygous mutant line showed the mutant phenotype and kanamycin resistance with no segregation. No allantoinase activity was detected in the mutant plants. However, wild-type seedlings had an allantoinase activity of about 10 μmol min^{-1} mg^{-1} protein when cultured on ALN media (Figure 3-4D). This level was higher than the activity of wild-type seedlings (1 μmol min^{-1} mg^{-1} protein) and the black locust seedlings (5 μmol min^{-1} mg^{-1} protein) cultured on Murashige and Skoog media, suggesting that the *AtALN* gene may be regulated by exogenous nitrogen conditions. In the presence of 10 mm ALN, the mutant seedlings were smaller when compared with the wild-type seedlings. Furthermore, the leaf of the mutant was colored pale green, and the mutant cotyledons showed very little red. Notably, wild-type seedlings and mutant seedlings that were subjected to poor nitrogen conditions (less than 10 μM ammonium nitrate) showed similar phenotypes. Regardless of the nitrogen levels in the media, the *ataln* mutant growth patterns did not differ from those of the wild-type seedlings previously reported by Desimone et al. (2002). However, when the mutant seedlings were transferred from the ALN media onto Murashige and Skoog media containing 10 mm ammonium nitrate, the seedlings recovered, leaf color returned to dark green, and the plants grew normally and produced flowers.

Phylogenetic Analysis

This Arabidopsis ALN (AtALN) has been annotated as a putative dihydroorotase (DHO) or DHO-like protein. Because ALN and DHO, along with hydantoinase and dihydropyrimidinase, belong to the same superfamily of cyclic amidohydrolases (EC 3.5.2.-), this annotation is not surprising. Furthermore, many plant ALN gene sequences were not correctly annotated (mainly as "unknown" or "expressed protein"). Amino acid sequence alignment of RpALN identifies seven plant ALNs (Arabidopsis [*Arabidopsis thaliana*], barrel medic [*Medicago truncatula*], potato [*Solanum tuberosum*], rice [*Oryza sativa*], soybean [*Glycine max*], and tomato [*Lycopersicon esculentum*]) with high levels of amino acid sequence similarity (Figure 3-5). To examine the structural relationship of ALNs and DHOs, I constructed an unrooted phylogenetic tree with the seven plant ALNs and four plant DHOs (Figure 3-5). First, the analysis clearly showed that the two enzymes were unmistakably different at the amino acid level. The two enzyme groups (ALNs and DHOs) were distinctly separated in the phylogenetic tree, ALNs having very low similarity to DHOs (less than 12% similarity at the amino acid level). Within the same enzyme group, it was obvious that the proteins from the same family had the most structural similarity to each other. For instance, black locust ALN was most similar to that of soybean (GmALN; 82% similar) and barrel medic (MtALN; 86% similar), just as potato ALN was almost identical to the tomato ALN (95% similar). Furthermore, potato DHO (StDHO) was the most similar to tomato DHO (LeDHO; 98% similar).

In this report, I confirmed the function of putative ALNs from two species (black locust and Arabidopsis) and I suggest that the six plant sequences, which are currently annotated as "unknown" proteins, are ALNs based on sequence homology and

(A)



(B)

		Allantoinase (ALN)						Dihydroorotase (DHO)			
		BL	BM	TM	PT	AT	RC	TM	PT	AT	RC
ALN	Soybean (SB)	82	76	59	56	62	49	10	10	4	9
	Black locust (BL)	100	86	66	64	67	55	10	10	6	9
	Barrel medic (BM)		100	66	63	67	57	11	11	7	8
	Tomato (TM)			100	95	67	55	11	11	9	8
	Potato (PT)				100	67	52	11	11	7	10
	Arabidopsis (AT)					100	57	10	10	6	8
	Rice (RC)						100	8	11	7	6
DHO	Tomato (TM)							100	98	77	77
	Potato (PT)								100	77	77
	Arabidopsis (AT)									100	76

Figure 3-5. Dendrogram and similarity of allantoinases and dihydroorotases

phylogenetic relationship. Allantoinase belongs to a superfamily of metal-containing amidohydrolases and is predicted to have a common three-dimensional fold (Holm and Sander, 1997). Especially, the conserved residues (Figure 3-2, asterisks) responsible for metal biding were found in all reported ALNs. Recently, the recombinant *E. coli* ALN was shown to have the highest activity levels when observed in cell extracts from cultures supplemented with zinc and cobalt (Mulrooney and Hausinger, 2003).

RpALN Gene Expression

During seed germination and seedling growth, nitrogen metabolism is a critical factor for the survival of the tree in a natural environment. To determine the *RpALN* expression pattern during early development of black locust, northern-blot analysis was carried out using total RNAs isolated from cotyledons, axis, hypocotyls, and roots in different seed germination stages. In black locust, an axis emerges from the seed coat at 1 d post-imbibition. The axis continues to grow and shows the line of root and shoot junction at 4 DAG. At 6 DAG, cotyledons open, become wrinkled, and true leaves sprout. In the axis and hypocotyls, the *RpALN* gene expression level peaked at 2 to 4 DAG, and then dropped at 6 DAG. On the other hand, its expression in cotyledon was the highest at 6 to 9 DAG (Figure 3-6A) and reduced to the basal level at 19 DAG. No significant signal was detected in root tissues (4–6 DAG). The differential expression of the gene during the seed germination and seedling growth was also observed at the level of enzyme activity. The allantoinase activity was the highest in axis/hypocotyledon at 4 DAG and in cotyledon at 6 DAG (Figure 3-6B).

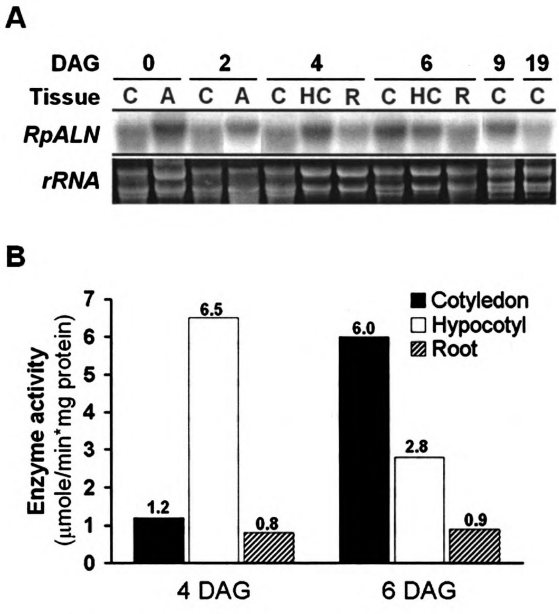


Figure 3-6. Differential expression of black locust allantoinase

(A) Northern blot analyses of *RpALN* in black locust seedlings. *RpALN* mRNA is differentially expressed in seedling tissues. According to the number of days after germination, total RNA (10 μg) from the indicated tissues was electrophoresed and blotted. Then, the RNA was probed with [α^{32}P]dCTP-labeled RpALN cDNA. Each lower panel presents the corresponding agarose gel stained with ethidium bromide, showing ribosomal RNAs.

(B) Enzyme assay of allantoinase of seedling tissues. Total proteins were extracted from three tissues of 4 day-old or 6 day-old seedlings: cotyledon (black bar), hypocotyl (white bar) and root (hatched bar). Results represent means of two independent experiments.

Trunk wood serves as a conduit for long-distance transport of nutrient and represents one of the largest sink tissues in trees. The expression pattern of *RpALN* gene was examined in the bark/cambial zone, sapwood, and sapwood-heartwood transition zones of mature black locust trees (Figure 3-7). The bark/cambial zone is the outer most region, the sapwood is the bright-colored middle region, and the heartwood is the dark, inner region. The transition zone, the interphase between sapwood and heartwood, is where the carbohydrates are converted to heartwood extractives. Because most of the cells in trunk wood are dead, the quantity and quality of mRNA that can be extracted from such tissue is limited. To overcome this problem, I used antisense northern-blot analysis, which included an in vitro transcription process to amplify mRNAs (Yang et al., 2003). My results showed that the bark/cambial region had the highest expression level (Figure 3-7A). I then examined the seasonal regulation of *RpALN* gene expression using bark/cambial zone tissues collected in spring, summer, and fall. In temperate deciduous trees, leaf nitrogen is translocated to perennial tissues (e.g. trunk wood) during the fall and the stored nitrogen is remobilized for spring shoot growth (Taylor and May, 1967; Ryan and Bormann, 1982). I found that the expression level of *RpALN* gene was higher in the spring and fall samples compared with that of the summer sample (Figure 3-7B), suggesting that the *RpALN* gene may be involved in nitrogen resorption and storage during the fall and resupply during the spring growth.

Effects of Nitrogen on RpALN and AtALN Gene Expression

To test whether nitrogen affects allantoinase gene expression, I carried out northern-blot analysis using black locust and Arabidopsis seedlings that were cultured on

Murashige and Skoog medium supplemented with or without nitrogen. The absence of nitrogen in the medium appears to induce the expression of RpALN gene (Figure 3-8A) during seedling germination and growth (2 and 4 DAG). In Arabidopsis seedlings, the absence of nitrogen or the presence of ALN as a sole nitrogen source dramatically increased *AtALN* gene expression (Figure 3-8B).

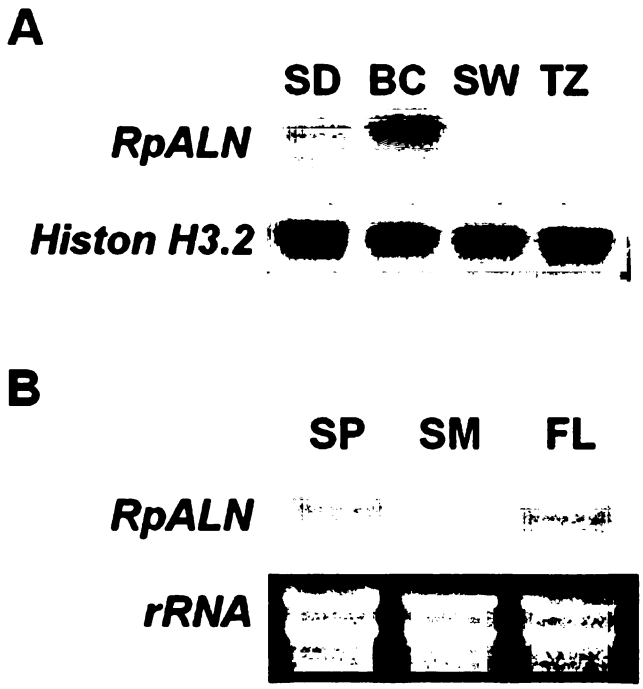


Figure 3-7. RpALN gene expression in trunk wood of black locust
(A) Antisense northern blot analysis of *RpALN* mRNA in seedling (SD), bark/cambial zone (BC), sapwood (SW), and transition zone (TZ) by antisense northern blot analysis. Histon H 3.2 transcript levels were indicated as a control. Each sample lane contained an equal amount of 200 ng of aRNAs.
(B) Northern blot analysis of RpALN mRNA using three different seasons of trunk wood. *RpALN* mRNA is differentially expressed in bark/cambial zones according to spring (SP), summer (SM) and fall (FL). Total RNA (10 μg) from the indicated season samples was separated and blotted, and the mRNA was then detected with [α^{32}P]dCTP-labeled RpALN cDNA. Lower panel presents the corresponding agarose gel stained with ethidium bromide, showing ribosomal RNAs.

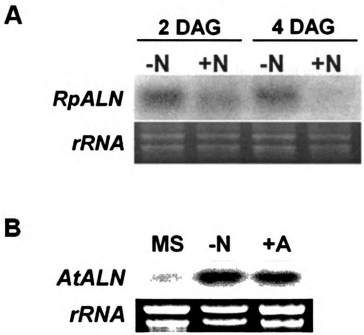


Figure 3-8. Nitrogen influence on the expression of allantoinase genes
(A) Expression of RpALN mRNA under nitrogen conditions. Total RNAs were isolated from 2 day-old or 4 day-old seedlings grown under nitrogen absent condition (-N) or supplied with 10 mM ammonium nitrate (+N).
(B) Northern blot analysis of AtALN mRNA in different nitrogen conditions. Total RNAs were isolated from 8 day-old seedlings cultured on MS medium supplemented with nitrogen (MS), without nitrogen (-N) or with 10 mM allantoin instead of the nitrogen sources (+A). Total RNA (10 μg) from the indicated season samples was separated and blotted, and the mRNA was then detected with [α^{32}P]dCTP-labeled RpALN or AtALN cDNA. Each lower panel presents the corresponding agarose gel stained with ethidium bromide, showing ribosomal RNAs.

Discussion

Biochemical analyses have shown that ureides such as ALN and ALA are present in higher plants, including ureide-type legume species such as soybean (Tracey, 1961; Atkins et al., 1991; Cheng et al., 2000). Desimone et al. (2002) demonstrated that even non-legume Arabidopsis can take up and use ALN as a nitrogen source. Here, I cloned and characterized two full-length cDNA clones (*AtALN* and *RpALN*) encoding allantoinase, further confirming that the ALN degradation pathway exists in Arabidopsis as well as in amide-type legumes (e.g. black locust). To my best knowledge, this is the first report on the cloning of functional allantoinase genes in plants. Using these two ALN gene sequences in a BLAST search against public databases, I identified additional ALN genes from various plant species including barrel medic, potato, soybean, potato, rice, and tomato. Multiple sequence alignment showed that all of the ALNs share high sequence similarity with known ALNs from other non-plant organisms (bullfrog, *E. coli*, and yeast), and identified characteristic conserved residues for metal binding. On the other hand, their sequence similarity with plant DHOs was very low ($\leq$ 11%). Plant ALN genes appear to be expressed abundantly in various tissues, suggesting that the ureide pathway is universal among plant species. In fact, Desimone et al. (2002) demonstrated that Arabidopsis could uptake and use ALN as a sole nitrogen source. The BLAST search against EST databases and the rice genome sequence database on the TIGR identified 15 ESTs from developing stems, developing leaves, drought-treated plants of barrel medic, 12 ESTs from dormant tuber, mature tuber, leaves/petioles of potato, and 31 ESTs from most tomato organs. Other plant species for which ALN genes were identified include rice, cotton (*Gossypium hirsutum*), soybean, and barley (*Hordeum vulgare*).

Until now, plant allantoinase genes have been mis-annotated as putative, unknown, or DHO. Particularly, the current Arabidopsis genome annotation calls *AtALN* a putative DHO or DHO-like protein. At least three pieces of evidence confirm that the *RpALN* and *AtALN* genes encode allantoinase. First, both of the genes were able to genetically complement a yeast mutant defective in ALN hydrolysis. Second, an Arabidopsis mutant *(ataln m2-1)* with T-DNA insertion in the 12th exon of *AtALN* was not able to use ALN as sole source of nitrogen, whereas wild-type plants grew on ALN medium. Third, a phylogenetic tree constructed with seven plant ALNs and four plant DHOs showed that the two enzyme groups were unmistakably different at the amino acid level.

Seed germination and seedling growth are the developmental stages that can affect the survival of the plant in its natural ecosystem. Developmental and metabolic programs are accomplished during the early stage of plant life (Howell, 1998; Holdsworth et al., 1999; Eastmond et al., 2000). Different metabolic pathways are carried out in each tissue of the seedling during its growth and development. For example, a nutrient source tissue (e.g. cotyledon) performs metabolisms that generate nutrients transported to sink tissues, whereas a sink source tissue (e.g. axis) conducts metabolisms involved in cell growth. Thus, an axis cell likely demands more nitrogen and carbon than does a cotyledon cell. It is necessary even for the same metabolism to be differentially controlled depending on the tissues involved. Enzyme activities required for the production of ureides were observed in young soybean seedlings (Polayes and Schubert, 1984). My results showed that black locust seedlings differentially express allantoinase genes depending on tissue type and developmental stages. Although these findings

suggest that allantoinase may be differentially regulated according to the nitrogen demand, it is not known whether the allantoinase degradation pathway plays any nutritional role in seedling development and growth.

Many plant purine biosynthetic enzymes appear to be localized in organelles (Smith and Atkins, 2002). Previous reports indicate that allantoinase is localized in microbodies of plants and animals (for review, see Schubert and Borland, 1990). Hanks et al. (1981) reported that allantoinase was localized in the endoplasmic reticulum (ER). It has been suggested that this ER localization may play a role in the export of ALA from the cell. However, the ER localization of allantoinase has not been independently confirmed. In this regard, my finding that both of the two plant allantoinases have putative transit sequences is interesting. However, further protein localization studies, using green fluorescent protein (GFP) fusion proteins in transgenic plants, can provide direct evidence for the intracellular localization of allantoinase. The cloning of the two functional ALN genes can facilitate such studies.

Signals derived from nitrogen treatment or nitrogen shortage are certainly involved in triggering widespread changes in gene expression and metabolic pathways. Cooper et al. (1990) have shown that in yeast, the ureide pathway is transcriptionally up-regulated under the nitrogen-deficient conditions. It appears that allantoinase genes are down-regulated in nitrogen-rich conditions, but up-regulated in nitrogen-limiting conditions. For example, the yeast allantoinase gene *(dal*1) was up-regulated by nitrogen depletion as well as amino acid starvation (Gasch, et al., 2000; also web supplement at http://genome-www.stanford.edu/yeast stress/index.shtml). In Arabidopsis, cDNA microarray data available from the Stanford Microarray Database (http://genome-

www5.stanford.edu/MicroArray/SMD/) showed that the ALN gene is down-regulated by nitrogen treatment, potassium nitrate versus no treatment (experiment ID nos. 3787 and 3789), the potassium chloride versus ammonium chloride (experiment ID nos. 12308 and 12309), and the potassium chloride versus potassium nitrate (experiment ID nos. 10849 and 10851). In addition, previous studies showed that nitrate treatment decreased the nodule mass and nitrogen-fixing activity, which subsequently resulted in the reduction of ureide content in the xylem sap in soybean (Israel and McClure, 1980; McNeil and LaRue, 1984) and cowpea (Pate et al., 1980) plants. My data showing nitrogen-induced down-regulation of ALN gene expression confirm these earlier physiological and biochemical observations.

In addition to the developmental regulation of ALN, a differential gene expression among different tissue types and seasonal variation might offer some insight on nitrogen metabolism in perennial woody plant species. In the current study, I found that black locust ALN genes had the highest expression levels in bark/cambium tissue compared with sapwood, sapwood-heartwood transition zone, or seedlings. In the fall, nitrogen in the leaves is mobilized and transported to storage tissues such as bark and xylem ray cells (Kang and Titus, 1980; Chapin and Kedrowswki, 1983). The stored nitrogen is then remobilized to support the spring growth (Taylor and May, 1967; Ryan and Bormann, 1982). It is notable that the *RpALN* gene expression was up-regulated in spring and fall compared with summer, again supporting the previous physiological observations. This seasonal differential expression of *RpALN* gene strongly suggests that *RpALN* may be involved in the seasonal cycle of nitrogen resorption, storage, and recycling in temperate deciduous trees. Furthermore, the cloning of two plant ALN genes will facilitate further

studies on the molecular and biochemical basis for nitrogen uptake, metabolism, and recycling in plants.

Literature Cited

Atkins CA, Storer PJ, Young EB (1991) Translocation of nitrogen and expression of nodule-specific uricase (nodulin-35) in *Robinia pseudoacacia*. Physiol Plant 83: 483-491.

Bell JA, Webb MA (1995) Immunoaffinity purification and comparison of allantoinases from soybean root nodules and cotyledons. Plant Physiol 107: 435-441.

Buckholz RG, Cooper TG (1991) The allantoinase (DAL1) gene of *Saccharomyces cerevisiae*. Yeast 7: 913-923.

Campbell JW, Bishop SH (1970) Nitrogen metabolism in molluscs. *In* JW Campbell, ed, Comparative Biochemistry of Nitrogen Metabolism, Vol 1. Academic Press, London, UK, pp 103-206.

Chapin FS, Kedrowski RA (1983) Seasonal changes in nitrogen and phosphorus fractions and autumn retranslocation in evergreen and deciduous taiga trees. Ecology 64: 376-391.

Cheng XG, Nomura M, Takane K, Kouchi H, Tajima S (2000) Expression of two uricase (Nodulin-35) genes in a non-ureide type legume, *Medicago sativa*. Plant Cell Physiol 41: 104-109.

Cooper TG (1980) Selective gene expression and intracellular compartment: two means of regulating nitrogen metabolism in yeast. Trands Biochem Sci 5: 332-334.

Cooper TG, Ferguson D, Rai R, Bysani N (1990) The GLN3 gene product is required for transcriptional activation of allantoin system gene expression in *Saccharomyces cerevisiae*. J Bacteriol 172: 1014-1018.

Desimone M, Catoni E, Ludewig U, Hilpert M, Schneider A, Kunze R, Tegeder M, Frommer WB, Schumacher K (2002) A novel superfamily of transporters for allantoin and other oxo derivatives of nitrogen heterocyclic compounds in Arabidopsis. Plant Cell 14: 847-856.

Eastmond PJ, Germain V, Lange PR, Bryce JH, Smith SM, Graham IA (2000) Post-germinative growth and lipid catabolism in oilseeds lacking the glyoxylate cycle. Proc Natl Acad Sci USA 10: 5669-5674.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1005-1016.

Felsenstein J. (1989) PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11: 4241-4257.

Hanks JF, Tolbert NE, Schubert KR (1981) Localization of enzymes of ureide biosynthesis in peroxisomes and microsomes of nodules. Plant Physiol 68: 65-69.

Hayashi S, Jain S, Chu R, Alvares K, Xu B, Erfurth F, Usuda N, Rao MS, Reddy SK, Noguchi T, Reddy JK, Yeldandi AY (1994) Amphibian allantoinase. Molecular cloning, tissue distribution, and functional expression. J Biol Chem 269: 12269-12276.

Holdsworth M, Kurup S, McKibbin R (1999) Molecular and genetic mechanisms regulating the transition from embryo development to germination. Trends Pharmacol Sci 4: 275-280.

Holm L, Sander C (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins 28: 72-82.

Howell SH (1998) Seedling development. *In* Molecular Genetics of Plant Development. Cambridge University Press, Cambridge, UK, pp 83-102.

Israel DW, McClure PR (1980) Nitrogen translocation in the xylem of soybeans. *In* FT Corbin, ed, World Soybean Research Conference II: Proceedings, Westview Press, Boulder, Colorado, pp 111-127.

Kang SM, Tirus JS (1980) Qualitative and quantitative changes in nitrogenous compounds in senescing leaf and bark tissues of apple. Physiol Plant 50: 285-290.

Kim GJ, Lee DE, Kim HS (2000) Functional expression and characterization of the two cyclic amidohydrolase enzymes, allantoinase and a novel phenylhydantoinase, from *Escherichia coli*. J Bacteriol 182: 7021-7028.

McClure PR, Israel DW (1979) Transport of nitrogen in the xylem of soybean plants. Plant Physiol 64: 411-416.

McNeic DL, LaRue TA (1984) Effect of nitrogen source on ureides in soybean. Plant Physiol 66: 720-725.

Mulrooney SB, Hausinger RP (2003) Metal ion dependence of recombinant *Escherichia coli* allantoinase. J Bacteriol 185: 126-34.

Muñoz A, Piedras P, Aguilar M, Pineda M (2001) Urea is a product of ureidoglycolate degradation in chickpea. Purification and characterization of the ureidoglycolate urea-lyase. Plant Physiology 125: 828-834.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 473-479.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Eng. 10: 1-6.

Noguchi T, Fujiwara S, Hayashi S (1986) Evolution of allantoinase and allantoicase involved in urate degradation in liver peroxisomes. A rapid purification of amphibian allantoinase and allantoicase complex, its subunit locations of the 2 enzymes, and its comparison with fish allantoinase and allantoicase. J Biol Chem 261: 4221-4223.

Pate JS, Atkins CA, White ST, Rainbird RM, Woo KC (1980) Nitrogen nutrition and xylem transport of nitrogen in ureide-producing grain legumes. Plant Physiol 65: 961-965.

Polayes DA, Schubert KR (1984) Purine synthesis and catabolism in soybean seedlings. The biogenesis of ureides. Plant Physiol 75: 1104-1110.

Raychaudhuri A, Tipton PT (2002) Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism. Plant Physiol 130: 2061-2068.

Romanov V, Merski MT, Hausinger RP (1999) Assays for allantoinase. Analytical Biochemistry 268: 49-53.

Ryan DF, Bormann FH (1982) Nutrient resorption in northern hardwood forests. BioScience 32: 29-32.

Schubert KR, Boland MJ (1990) The ureides. *In* BJ Miflin, PJ Lea, eds, The Biochemistry of Plants, Vol. 16. Academic Press, San Diego, pp 197-283.

Schulten HR, Schnitzer M (1998) The chemistry of soil organic nitrogen: A review. Biol Fertil Soils 26: 1-15.

Shubert KR (1981) Enzymes of purine biosynthesis and catabolism in *Glycine max*. Plant Physiol 68: 1115-1122.

Smith PMC, Atkins CA (2002) Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. Plant Physiol 128: 793-802

Stryer L (1988) Biochemistry, Ed 3. WH Freeman, New York, pp 618-623

Taylor BK, May LH (1967) Nitrogen metabolism, translocation and recycling in apple trees. Hort. Rev. 4: 204-246

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

Tracey MV (1961) Urea and Ureides. *In* K Peach K, M Tracey, eds, Modern Methods of Plant Analysis, Vol 4. Springer Verlag, Heidelberg, pp 119-141

Vogels GD, Trijbels F, Uffink A, (1966) Allantoinases from bacterial, plant and animal sources. I. Purification and enzymatic properties. Biochim Biophys Acta 122: 482-496

Winkler RG, Blevins DG, Polacco JC, Randall DD (1988) Ureide catabolism in nitrogen-fixing legumes. Trends Biochem Sci 13: 97-100

Yang J, Park S, Kamdem DP, Keathley DE, Retzel E, Paule C, Kapur V, Han K-H (2003) Novel gene expression profiles define the metabolic and physiological processes characteristic of wood and its extractive formation in a hardwood tree species, *Robinia pseudoacacia*. Plant Mol Biol 52: 935-956.

CONCLUSIONS

1) In this study, ESTs present unique gene sets that are expressed in uncharted deep inside trunk wood.

2) EST analysis shows that the libraies from the different trunk regions have distinctive characteristics.

3) The gene expression profiles are different based on the regions of trunk wood, and highly expressed genes represnt the character of each region.

4) Depending on season, the gene expression patterns are different.

5) The genes related to secondary metabolism are highly expressed in the summer season when compared with other seasons.

6) Allantoinase gene is not specific in ureid-type legume,but general in plant kingdom.

7) The black locust allantoinase gene is differentially regulated in seedling accoding to the nitrogen demand.

8) The allantoinase gene is involved in nitrogen mobilization in black locust.

9) Negatively, nitrogen regulates the expression of allantoinase genes.