This is to certify that the
dissertation entitled

Comparative and Functional Genomics Approaches to Understand Environmental Adaptation in Woody Perennial, *Populus*

presented by

Sunchung Park

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Plant Breeding and Genetics

Major Professor's Signature

12/2/04

Date

COMPARATIVE AND FUNCTIONAL GENOMICS APPROACHES TO UNDERSTAND ENVIRONMENTAL ADAPTATION IN WOODY PERENNIAL, *POPULUS*

By

Sunchung Park

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

COMPARATIVE AND FUNCTIONAL GENOMICS APPROACHES TO UNDERSTAND ENVIRONMENTAL ADAPTATION IN WOODY PERENNIAL, *POPULUS*

By

Sunchung Park

In order to survive various environmental changes over their long life span, woody perennials have evolved several different physiological properties compared with annual herbaceous plants such as vegetative dormancy, seasonal nutrient reallocation, and extensive frost hardiness. Despite these distinguishable properties, most of our knowledge on the molecular basis underlying these processes comes from extrapolation based on findings from annual plant studies. Since poplar ESTs and genomic resources as tree model plant have now been available, I have taken comparative and functional genomic approaches to understand the genetic regulation of these properties. First I have identified a set of 25,282 unigenes from 105,831 poplar ESTs obtained from public databases. I then estimated poplar multigene families. Several multigene families had a higher copy number in poplar than in *Arabidopsis,* implying potential lineage-specific proliferation of poplar protein families. The protein families with such expansion may reflect the adaptation to extensive environmental stresses in perennial poplar. In addition, comparison of poplar unigenes with the *Arabidopsis* transcriptomes revealed that genes involved in transcriptional regulation are the most divergent while metabolism-related genes are most conserved. Second, to understand genetic regulation underlying seasonal growth, I carried out a series of global transcriptional analysis using stem samples from poplar trees grown in field and under controlled environment conditions. The results

showed that extensive metabolic switch and alteration in cellular functions occurred during transition from rest to quiescent. The genes involved in the pentose phosphate pathway were elevated and defense-related genes were overrepresented during early winter whereas the genes involved in fermentation and fatty acid β-oxidation were upregulated and signaling-related genes were overrepresented during late winter. Furthermore, clustering of these genes into eight groups has revealed that plants regulate seasonal growth by integrating environmental factors with developmental stages. For example, short daylength could modulate the cold signal to achieve maximum cold hardiness. This mechanism may help plants to control cold hardiness more closely timed with seasonality than unpredictable temperature alone does. Finally, in this study, an auxin-repressed gene from black locust (*Robinia pseudoacacia*) was investigated. The gene is posttranscriptionally regulated and negatively associated with shoot elongation.

ACKNOWLEDGEMENTS

This study was supported by Department of Forestry, Plant Breeding&Genetics, Genome Technology Support Facility, and Tree Research Centr in Michigan State University. I thank all the staff for their kind assistance.

I want to thank several people who have contributed to this work in one way or another. First of all, I would like to thank to my supervisor, Dr. Kyung-Hwan Han for guiding and advising me during the course of the study. I also thank my guaidance committee, Dr. Steve van Nocker, Dr. Rebecca Grumet and and Dr. Dean Dellapenna for their helpful comments, encouragement, and positive criticism. Dr. Daniel E. Keathley is acknowledged for his careful work in reviewing the manuscript and coauthoring one paper. As well, Merilyn Ruthig are acknowledged for fruitful editing manuscripts and her friendship. Especially, I want to thank Dr. Joanne H.Whallon for giving me the opportunity to serve as teaching assistant and her support. I thank my colleagues at the Lab for sharing every day life.

Above all, I wish to express my deepest gratitude to my wife, Sookyung for rewarding collaboration, invaluable support and coauthoring one paper. Certainly, I would like to thank my family for their interest and support to my work and life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Trees represent the majority of terrestrial biomass production, providing a range of ecological services, such as carbon sequestration, bioremediation, nutrient cycling and biofilteration. Trees are also economically important by producing a variety of wood-based products including timber, pulp and paper. In addition, trends in developed countries to reduce their use of non-renewable resources and to conserve the environment are likely to increase the use of products generated from renewable resources such as wood (Bhalerao et al., 2003; Wullschleger et al., 2002)

Despite the economical and ecological importance of trees, our understanding at the molecular level of tree development and adaptation to environments is very limited, compared to that in herbaceous plants. Most of what we know about the regulation of these processes comes from research on *Arabidopsis thaliana*, a small herbaceous annual plant. Until recently complex biological questions of these types could not be reasonably addressed in woody plants. However, the development of *Populus* genomic resources (e.g., its complete genome sequences and a large number of ESTs from various tissues) and the recent advances in genomic and bioinfomatic tools allow us to take integrated functional and comparative genomics approaches to address these questions (Brunner et al., 2004).

Populus as a model woody plant

Trees are unique among plants since they can achieve extreme longevity and great size. Bristlecone pine (*Pinus longaeva*), for example, can live to be more than 4700 years old, and coast redwood (*Sequoia sempervirens*) and giant sequoia (*Sequoiadendron*

giganteum can exceed 100 m in height and 25 m in circumference, respectively. Because of their great potential size and longevity, trees differ developmentally in so many respects from herbaceous plants, including patterns of apical dominance, perennial growth habit, secondary xylem (wood) formation, mature and juvenile phase change, dormancy cycles, seasonal nutrient reallocation and reproduction. Consequently, tree development is unlikely to be understood based solely on information from herbaceous systems.

Arabidopsis thaliana has been widely accepted as a model plant in plant biology since it has a small genome, rapid life cycle and various value in experimental studies at the physiological and biochemical levels (Goodman et al., 1995). However, it has often been questioned whether arabidopsis can serve as the model for tree biology. For example, secondary xylem formation in Arabidopsis could be used to study wood formation in tree but it is essential that results be placed in the context provided by parallel studies with trees (Chaffey et al., 2002; Oh et al., 2003). Another such example could be the annual growth cycle, which is essential for perennial growth of woody plants. The complex regulation patterns of the annual growth cycle involve a complex of interactions between environmental signals (including day length and temperature) and plant signal transduction pathways. It is hard to imagine that *Arabidopsis* could be adequate for the study of such processes (Taylor, 2002). Thus, to understand further unique development and adaptation processes that occur in trees, a 'model' tree is needed.

Populus has been accepted as a 'model' woody plant to complement the genetic resources being developed in *Arabidopsis*. The genus *Populus* (poplars, cottonwoods and

aspens) belongs to the Salicaceae family and is comprised of about 30 species, all found in the Northern hemisphere and exhibiting some of the fastest growth rates observed in temperate trees. Poplar provides an attractive model system for several reasons. First, and most important, are the recent completion of the *Populus* genome sequence and a large number of expressed sequence tags (ESTs) have been produced and analyzed from various tissues of *Populus* species (Hertzberg et al., 2001; Kohler et al., 2003; Sterky et al., 1998). Other reasons include the relatively small genome size (ca. 550 Mbp, only 4x larger than Arabidopsis but 40x smaller than pine), the large number of molecular genetic maps and the relative ease of genetic transformation. *Populus* can also be propagated vegetatively, making mapping populations immortal and facilitating the production of large amounts of clonal materials for experimentation. In addition, the genus *Populus* has genuine commercial value as a tree for timber, plywood, pulp and paper (Dinus et al., 2001)

Functional genomics approach

Because of their longevity, trees possess various means of adapting to environmental changes. While fundamental cellular and physiological aspects of gene function in plants can be far more precisely analyzed in *Arabidopsis*, rice and other annual model plants, a key challenge for poplar research is how to conduct affordable experiments that enable gene-level inferences to be made about these whole organism-level attributes. Until recently complex biological question of these types could not be reasonably addressed in woody plants. Today, genomics is, however, providing tools that allow examination of whole-plant responses at single-gene resolution, thus providing a bridge between

molecular biology and whole-plant physiology. This will provide an integrated view of tree responses at the finest scale (Matsuyama et al., 2002; Seki et al., 2001). As a result, genomics may provide a mechanism for translating the understanding of cellular processes to system-level understanding of tree growth.

On the other hand, the study of woody plants presents many challenges to the use of the conventional genetic methods such as the self-crossing, inbreeding lines and recessive mutant screening due to the long generation time and long time-to-flower, even though physiological and biochemical experiments are more readily conducted. As a result, there are no well-characterized mutant collections of forest trees. Therefore, reverse genetics, in which a gene sequence is first obtained and then its function learned via directed alteration in transgenic, is the route taken in most transgenic experiments with trees. However, the efficiency of reverse genetic method is highly depending on the potential of the selected candidate genes, and the candidate genes should be carefully selected for functional evaluation. Functional genomic approaches could facilitate the selection process of the candidate genes by providing the single gene resolution in global scale.

Comparative Genomic approach

The real power of genomics lies in the fact that all known organisms share the same genetic code. Sequences for many genes have been largely conserved through evolutionary time, so that information gained about the functioning of genes in one organism can be applied to distantly related organisms. For example, the functions of more than half the genes in *Arabidopsis* were inferred from experimental data gathered in

animal, yeast and microbial model species (Somerville and Somerville 1999). Furthermore, comparative analyses of the sequences of expressed genes have provided a powerful means for gene discovery (e.g., the identification of species-specific genes and multigene family) as well as the study of the molecular basis of plant diversity (Aravind et al., 2000; Lespinet et al., 2002; Riechmann et al., 2000).

Recently, the complete genome sequence of *Arabidopsis thaliana* and rice (*Oryza sativa* L.) has been available (Eckardt 2000) and efforts have been initiated to sequence the genomes of other plants, including *Medicago truncatula,* a model legume (Frugoli and Harris 2001). In addition, large numbers of ESTs are being generated from a wide variety of plant species and are publicly available (http://www.tigr.org/tdb/tgi/-plant.shtml).

Since most of genomic resouces are generated from herbaceous plants, the physiological properties unique to perennial woody plants could be addressed through comparative genomics. It can also be expected that, through comparative genomics, many of the rapid advances in understanding structural and regulatory genes in other model organisms can be directly translated to poplar and other tree species. And the reverse will also be true; research with *Populus* will aid in understanding the genetic regulation on metabolic processes in other plants.

Annual growth cycle in woody perennials

Trees growing in temperate climates are subjected to seasonal environmental variations such as chilling/frost and drought. Thus, temperate woody perennials have evolved annual growth cycles that involve alternation between active growth and vegetative

dormancy (Arora et al 2003). The annual growth cycle of woody perennials is closely timed with local seasonal changes, which could be enabled by interaction between endogenous (e.g., adaptable genetic factors) and environmental factors (e.g., photoperiod and temperature) (Perry, 1971). The genetic adaptation to local climatic conditions is a dominant determining factor for geographical distribution, growth capacity, and survival of tree species (Samish, 1954; Weiser, 1970). Thus, understanding of genetic regulation on annual growth cycle is not only of scientific interest, but also of economic and environmental significance.

The annual growth cycle has two major traits, dormancy and cold hardiness, which are genetically and temporally correlated with each other (Chen et al., 2002). Dormancy, in general, is defined as temporary suspension of visible growth of any plant structure containing a meristem (Lang, 1987). Transition from active to dormancy state occurs gradually (Perry, 1971), and termination of meristematic cell division is one of the first visible indicators of the establishment of a dormant state. Physiologically, dormancy can be distinguished into three states: paradormancy, ecodormancy and endodormancy (Lang, 1987). The paradormancy describes dormancy when the initial reaction leading to growth control involves a specific signal originating in or initially perceived in a different structure from the one in which dormancy is manifested. The specific signal may be due to a continuous production of inhibitory factors, as in apical dominance. The ecodormancy describes dormancy when one or more factors in the basic growth environment are unsuitable for overall growth metabolism. For example, water or nutrient deficiencies or cold could cease plant growth, which is not limited to a specific meristemic tissue. The endodormancy describes dormancy when the initial reaction

leading to growth control is a specific perception of an environmental or endogenous signal in the affected structure alone. The vascular cambium in woody perennials, as shoot meristem does, exhibits different stages of dormancy (Little and Bonga, 1974), that is, endodoramancy (i.e., rest) and ecodormancy (i.e, quiescence). At the end of growing season, typically late summer, the camibium enter dormancy by responding to short daylength (SD). At this stage, cambial dormancy is regulated by internal factors, which has been denoted rest and is characterized by non-responsiveness to indole-3-acetic acid (IAA) under favorable conditions for cambial activity (Little and Bonga, 1974; Little, 1981). Rest is characterized by a requirement for sustained exposure to low temperatures before active growth can resume in the spring (i.e., chilling requirement). The cambium gradually regains the ability to produce xylem in response to IAA as exposure to natural chilling is increased (Christersson, 1978). When fully responsive, the cambium is in the quiescence stage of dormancy, which is imposed solely by adverse external factors. Upon the return of warmer temperatures in spring, plants are released from dormancy and also fully dehardened, primarily by responding to warm temperature in whole shoots of woody species (Fuchigami and Nee, 1987; Smitspinks, et al., 1985).

Cold acclimation is distinguished by typically two stages in woody plants. First stage of cold acclimation together with the onset of dormancy is induced by short daylength under warm temperatures (Weiser, 1970), at which plants can survive temperatures slightly below freezing. The response to SD has an adaptive significance. First, by responding to photoperiod cues, plants are able to begin acclimating to winter conditions prior to the onset of low temperature. Second, photoperiodic cue is much more reliable signal of the approaching winter compared to temperature, which can vary

dramatically from year to year. By responding to SD closely timed with local seasonal time, plants are able to synchronize dormancy induction and cold acclimation with the end of growing season and coming winter. However, higher capacity of cold acclimation, i.e., frost hardiness could be achieved by subsequent cold after dormancy induction by SD, suggesting the possible interrelation between dormancy and cold hardiness (Nissila and Fuchigami, 1978). The idea is supported further by the fact that dormant deciduous genotypes always had higher extent of cold hardiness than non-dormant evergreen genotypes in peach (Arora and Wisniewski, 1994). However, the molecular mechanism underlying the interaction between SD and cold is poorly understood.

Recent studies showed that cold acclimation by low temperature (LT) and SD could be uncoupled in hybrid aspen, suggesting that SD and LT might utilize independent pathways for the cold acclimation (Welling et al., 2002). Olsen et al (1997) reported that cold acclimation and dormancy initiation by SDs were controlled by phytochorme A-involving pathway in *Populus* (Howe et al., 1996). Since the molecules involved in day length perception signaling (e.g., photoreceptors, CONSTANS, circadian proteins) have been well studied and well known as high conservation across distant species, e.g., *Arabidopsis* and rice (Yanovsky and kay, 2003), it is plausible that trees must have functionally homologous genes playing similar roles in their response to daylength changes. Thus, it is interesting to question the mechanism by which the conserved pathway could crosstalk with the cold acclimation pathways. Furthermore, most dormancy/cold hardiness-related traits appear to be inherited in a quantitative manner, displaying continuous distribution in progeny tests (Bradshaw and Stettler, 1995; Howe

et al., 2000). This is indicative of multigenic control. Thereby, these results also justify my genomic approaches to the study of annual growth cycle.

Despite these many observations, we still do not clearly understand the molecular basis of seasonal development. Perhaps our limited understanding of the molecular biology of seasonal development is, in part, due to superimposition of various internal and external events occurring during annual growth cycle. Dormancy development and cold acclimation partially overlap and are interrelated, making it difficult to associate physiological and molecular changes specifically with one or the other phenological events. Thus, to separate dormancy processes from cold hardiness processes has been a fundamental problem in dormancy research. It is important to identify which of the changes is associated with specific events of seasonal growth cycle. Although there were some previous studies (Espinosa-Ruiz et al., 2004; Rowland and Arora, 1997; Welling et al., 2004) that have identified the genes associated with different aspects of dormancy or cold acclimation, they were limited to a few of genes or a short period of study time.

Auxin repressed gene

Since auxin, a plant growth regulator produced mainly in apical buds, plays a key role in a wide variety of growth and developmental processes (Berleth, et al., 2000), the characterization of auxin regulated genes will expand our knowledge on the various biological implication of auxin.

In this thesis, an auxin repressed protein (ARP) from black locust has been characterized. To date, three putative orthologs of ARP were isolated and characterized from pea, strawberry, and tobacco. They were isolated from different tissue/organs (e.g.,

strawberry receptacles, Reddy and Poovaiah, 1990; pea dormant bud, Stafstrom et al., 1998; tobacco pollen, Steiner et al., 2003), however, they all showed transcriptional repression in response to auxin. The previous results suggest that this family of genes may have conserved biological function across species although they were expressed in different organs. Recently, we identified an ARP ortholog (named RpARP) from the expressed sequence tags (ESTs) derived from the trunk wood of a 10-year old black locust (*R. pseudoacacia*) (Yang et al, 2003). The gene was expressed across the trunk wood (bark/cambium, sapwood, and sapwood-heartwood transition zone). The deduced amino acid sequence of RpARP shows high sequence homology with other ARP orthologs. However, the underlying mechanism regulating repression by auxin, and potential biological function that accounts for their expression in various organs or plant species, is not known.

References

Aravind L, Watanabe H, Lipman DJ,Koonin EV (2000) Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci U S A. 97: 11319-24

Arora R, Rowland LJ,Tanino K (2003) Induction and release of bud dormancy in woody perennials: A science comes of age. Hortscience. 38: 911-921

Arora R,Wisniewski ME (1994) Cold-Acclimation in Genetically Related (Sibling) Deciduous and Evergreen Peach (Prunus-Persica L Batsch) .2. A 60-Kilodalton Bark Protein in Cold-Acclimated Tissues of Peach Is Heat-Stable and Related to the Dehydrin Family of Proteins. Plant Physiology. 105: 95-101

Berleth T, Mattsson J,Hardtke CS (2000) Vascular continuity and auxin signals. Trends in Plant Science. 5: 387-393

Bhalerao R, Nilsson O,Sandberg G (2003) Out of the woods: forest biotechnology enters the genomic era. Curr Opin Biotechnol. 14: 206-13

Bradshaw HD, Jr.,Stettler RF (1995) Molecular genetics of growth and development in populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. Genetics. 139: 963-73

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

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

Chen THH, Howe GT,Bradshaw HD (2002) Molecular genetic analysis of dormancy-related traits in poplars. Weed Science. 50: 232-240

Christersson L (1978) Influence of Photoperiod and Temperature on Development of Frost Hardiness in Seedlings of Pinus-Silvestris and Picea-Abies. Physiologia Plantarum. 44: 288-294

Dinus RJ, Payne P, Sewell NM, Chiang VL,Tuskan GA (2001) Genetic modification of short rotation popular wood: Properties for ethanol fuel and fiber productions. Critical Reviews in Plant Sciences. 20: 51-69

Eckardt NA (2000) Sequencing the rice genome. Plant Cell. 12: 2011-7

Espinosa-Ruiz A, Saxena S, Schmidt J, Mellerowicz E, Miskolczi P, Bako L,Bhalerao RP (2004) Differential stage-specific regulation of cyclin-dependent kinases during cambial dormancy in hybrid aspen. Plant J. 38: 603-615

Frugoli J,Harris J (2001) Medicago truncatula on the move! Plant Cell. 13: 458-63

Fuchigami LH,Nee CC (1987) Degree Growth Stage Model and Rest-Breaking Mechanisms in Temperate Woody Perennials. Hortscience. 22: 836-845

Goodman HM, Ecker JR,Dean C (1995) The genome of Arabidopsis thaliana. Proc Natl Acad Sci U S A. 92: 10831-5

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P,Sandberg G (2001) A transcriptional roadmap to wood formation. Proc Natl Acad Sci U S A. 98: 14732-7

Howe GT, Gardner G, Hackett WP,Furnier GR (1996) Phytochrome control of short-day-induced bud set in black cottonwood. Physiologia Plantarum. 97: 95-103

Howe GT, Saruul P, Davis J,Chen THH (2000) Quantitative genetics of bud phenology, frost damage, and winter survival in an F-2 family of hybrid poplars. Theoretical and Applied Genetics. 101: 632-642

Kohler A, Delaruelle C, Martin D, Encelot N,Martin F (2003) The poplar root transcriptome: analysis of 7000 expressed sequence tags. FEBS Lett. 542: 37-41

Lang GA, Early JD, Martin GC,Darnell RL (1987) Endodormancy, Paradormancy, and Ecodormancy - Physiological Terminology and Classification for Dormancy Research. Hortscience. 22: 371-377

Lespinet O, Wolf YI, Koonin EV,Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12: 1048-59

Little CHA (1981) Effect of Cambial Dormancy State on the Transport of 1-C-14 Indol-3-Ylacetic Acid in Abies-Balsamea Shoots. Canadian Journal of Botany-Revue Canadienne De Botanique. 59: 342-348

Little CHA,Bonga JM (1974) Rest in Cambium of Abies-Balsamea. Canadian Journal of Botany-Revue Canadienne De Botanique. 52: 1723-1730

Matsuyama T, Tamaoki M, Nakajima N, Aono M, Kubo A, Moriya S, Ichihara T, Suzuki O,Saji H (2002) CDNA microarray assessment for ozone-stressed Arabidopsis thaliana. Environ Pollut. 117: 191-4

Nissila PC,Fuchigami LH (1978) Relationship between Vegetative Maturity and 1st Stage of Cold-Acclimation. Journal of the American Society for Horticultural Science. 103: 710-711

Oh S, Park S,Han KH (2003) Transcriptional regulation of secondary growth in Arabidopsis thaliana. J Exp Bot. 54: 2709-22

Olsen JE, Junttila O, Nilsen J, Eriksson ME, Martinussen I, Olsson O, Sandberg G,Moritz T (1997) Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. Plant Journal. 12: 1339-1350

Perry TO (1971) Dormancy of Trees in Winter. Science. 171: 29-&

Reddy ASN,Poovaiah BW (1990) Molecular-Cloning and Sequencing of a Cdna for an Auxin-Repressed Messenger-Rna - Correlation between Fruit-Growth and Repression of the Auxin-Regulated Gene. Plant Molecular Biology. 14: 127-136

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR,al. e (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 290: 2105-10

Rowland LJ,Arora R (1997) Proteins related to endodormancy (rest) in woody perennials. Plant Science. 126: 119-144

Samish RM (1954) Dormancy in woody plants. Annual Review of Plant Physiology and Plant Molecular Biology. 5: 183-203

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y,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

Smitspinks B, Swanson BT,Markhart AH (1985) The Effect of Photoperiod and Thermoperiod on Cold-Acclimation and Growth of Pinus-Sylvestris. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere. 15: 453-460

Somerville C,Somerville S (1999) Plant functional genomics. Science. 285: 380-3

Stafstrom JP, Ripley BD, Devitt ML,Drake B (1998) Dormancy-associated gene expression in pea axillary buds. Planta. 205: 547-552

Steiner C, Bauer J, Amrhein N,Bucher M (2003) Two novel genes are differentially expressed during early germination of the male gametophyte of Nicotiana tabacum. Biochim Biophys Acta. 1625: 123-33

Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B,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: 13330-5

Taylor G (2002) Populus: arabidopsis for forestry. Do we need a model tree? Ann Bot (Lond). 90: 681-9

Weiser CJ (1970) Cold resistance and injury in woody plants. Science. 169: 1269-78

Welling A, Rinne P, Vihera-Aarnio A, Kontunen-Soppela S, Heino P,Palva ET (2004) Photoperiod and temperature differentially regulate the expression of two dehydrin genes during overwintering of birch (Betula pubescens Ehrh.). Journal of Experimental Botany. 55: 507-516

Wullschleger SD, Tuskan GA, DiFazio SP (2002) Genomics and the tree physiologist. Tree Physiol. 22: 1273-6

Yang J, Park S, Kamdem DP, Keathley DE, Retzel E, Paule C, Kapur V,Han KH (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: 935-56

Yanovsky MJ, Kay SA (2003) Living by the calendar: how plants know when to flower. Nat Rev Mol Cell Biol. 4: 265-75

CHAPTER 1

Large-Scale Computational Analysis of Poplar ESTs Reveals the Repertoire and Unique Features of Expressed Genes in the Poplar Genome

Abstract

Perennial woody plants differ from annual herbaceous plants in several ways and are expected to have evolved to adopt a unique repertoire and expression profiles of functional genes. Poplar, a model tree species for which a large number of ESTs are publicly available, was used to carry out a large-scale comparative analysis with the expressed sequences of eight plant species. First we have obtained 105,831 poplar ESTs from public databases and identified a set of 25,282 unigenes (i.e., tentative non-redundant sequences). The majority of the unigenes (56%) had significant matches to *Arabidopsis* genes. We then estimated poplar multigene families by counting the tBLASTX matches of each unigene against the poplar unigene dataset itself. Forty-seven percent of the 25,282 unigenes were subsequently organized into 3,481 multigene families 89% of which had less than five copy members. In poplar protein kinases represent the largest family followed by GTP-binding proteins and Myb transcription factors. Several multigene families had a higher copy number in poplar than in *Arabidopsis* hinting potential lineage-specific proliferation of poplar protein families. The protein families with such expansion might be related to the adaptation to high degree of environmental stresses of perennial poplar growth. Comparison of poplar unigenes with the *Arabidopsis* transcriptome revealed that genes involved in transcriptional regulation are the most divergent while metabolism-related genes are most conserved.

Introduction

Tree growth is one of the most important biological processes on the Earth. Its product wood is of primary importance to humans as timber for construction fuelwood and wood-pulp for paper manufacturing. It is also the most environmentally cost-effective renewable source of energy. In order to survive various environmental changes over long periods of time trees have evolved to adopt unique physiological properties such as bud dormancy to overcome unfavorable growth conditions, deciduous habit to avoid deep frost, extended juvenile period, and secondary growth. We are interested in learning about the genetic regulation of the tree-specific developmental processes using *Populus* as a model.

The genus *Populus* belongs to the Salicaceae family and is comprised of about 30 species. Poplar has been widely accepted as a model system for tree biology by many researchers for several reasons: a) a small genome (ca. 550 Mbp only 4x larger than *Arabidopsis* but 40x smaller than pine); b) easy clonal propagation which allows for replication of experiments; c) rapid growth; d) extensive genetic variation; e) high expected synteny with *Arabidopsis*; f) available high-throughput transgenic technology; and g) high quality genetic maps. Furthermore its complete genome sequencing project now has approximately 8x coverage of the genome (Tuskan et al., 2003). It is therefore no surprise that a large number of expressed sequence tags (ESTs) have been produced and analyzed from various tissues of the *Populus* species (Sterky et al., 1998; Hertzberg et al., 2001; Kohler et al., 2003)

Single-pass sequencing of cDNA clones can reveal a substantial portion of the expressed genes of a genome. As a result large numbers of ESTs are being generated

from a wide variety of plant species. Comparative analyses of the expressed sequence tags (ESTs) provide a powerful means for gene discovery as well as the study of molecular basis of plant diversity. First we retrieved a total of 105,831 poplar ESTs from NCBI (http://www.ncbi.nih.gov) and organized them into clusters to generate unigene sets using the StackPACK program. This program has been applied successfully to identify different genes from redundant ESTs of other species such as barley (Michalek et al., 2002) and fungus (Trail et al., 2003). With the identified unigene sets we conducted large-scale computational comparisons against ca. 394,100 expressed sequences of eight plant species that include four dicots, three monocots and one gymnosperm. These species have diverse phylogenetic relationships (Soltis et al., 1999) (Figure 1-1). Of these eight species, two complete plant genome sequences (*Arabidopsis* and rice) were included in the analyses. These analyses provided us with a unique opportunity to examine the repertoire and organization of expressed genes in the poplar genome and to observe the differences in expression profiles between poplar and other plant species. The unigene set generated in this study are accessible on a publicly searchable website (http://www.genomics.msu.edu/ ~han/poplar.html).

Materials and Methods

Assembly and clustering of poplar ESTs

As of January 2003 a total of 105,831 poplar ESTs were available from four *Populus* species and were retrieved using a custom Perl script from dbEST and the non-redundant nucleotide database available at the NCBI (http://www.ncbi.nih.gov). The ESTs were assembled separately by individual species using the StackPACK™ (version 2.2)

clustering system (http://www.egenetics.com) using default settings for EST assembly (Miller et al., 1999). The process includes sequential steps of masking clustering assembly alignment analysis and consensus partitioning. The masking step employed CrossMatch (http://www.genome.washington.edu/uwgc/analysistools/phrap.htm) to mask vector artifacts and repetitive sequences as simple repeat. The masked sequences were clustered based on their relative similarity (the default is greater than 96 % identical over a window of 150 bases) determined by d2 cluster (Hide et al., 1994) that is word-based greedy clustering algorithm. The related but loose clusters were further aligned and assembled using PHRAP (http://www.genome.washington.edu/uwgc/analysistools/-phrap.htm) to improve alignment quality by generating particularly distinct sequences as singletons and highly related sequences as sub-contigs. The aligned sub-contigs were further analyzed using the CRAW alignment analysis tool (Chou and Burke 1999). CRAW is used to analyze sub-contigs for error and alternative expression forms partition the sub-contigs maximize consensus sequence length create final alignments and select the best consensus sequence. For each species, the ESTs were organized into contiguous overlapping consensus ('contigs') while the low-frequency ESTs were not incorporated into contig assemblies and remain as singletons. The resulting singletons and consensus sequences were used as tentative unigene sets in this study. Then the four unigene sets (a total of 41,955 sequences) were assembled again using the same StackPACK clustering system to produce a 29,818 poplar unigene set. In order to eliminate any transcripts that potentially resulted from the same locus this poplar unigene set was re-assembled using the BLASTCLUST program (ftp://ftp.ncbi.nlm.nih.gov/blast/documentations/-README.bcl) with a threshold of 95 % sequence identity covering a 90 % length region.

To avoid potential mosaic sequences of different species the sequence with the longest sequence read in each contig was selected instead of the resulting consensus sequences and used as a representative of the contig regardless of the species origin. The resulting contigs and singletons were further screened to eliminate any sequences shorter than 250-bp. The final 25,282 unigenes that may represent distinct genes were used for further analysis.

Other plant sequence data sets used in the analyses

Computational analyses were performed on several sets of plant gene sequences obtained from public databases. The *Arabidopsis* and rice gene sets as predicted from the entire genomic sequence contain 27,290 and 80,916 sequences respectively (as of March 2003) and are available through TIGR (http://www.tigr.org/tdb/). The minimally-redundant EST sets of three dicots three monocots and one gymnosperm species were obtained from the Gene Indices of TIGR (http://www.tigr.org/tdb/tgi/plant.shtml) which were constructed by first clustering and then assembling EST and annotated gene sequences from GenBank to produce a set of unique high-fidelity virtual transcripts or tentative consensus (TC) sequences (Quackenbush et al., 2000).

Similarity searches

Similarity searches were carried out with the Standalone BLAST programs (Altschul et al., 1997) using executable copies obtained from the NCBI (v 2.2.2). The sequence data sets used here were formatted first as a BLAST searchable database file using Formatdb executable one of the BLAST tools available from NCBI (for a manual see

ftp://ftp.ncbi.nih.gov/blast/documents /README.formatdb). Searches were performed through comparisons of protein sequences with translation of nucleotide query or database sequences using gapped BLASTX or tBLASTX (Altschul et al., 1997). Nucleotide queries were preprocessed with DUST to mask low-complexity regions and protein query sequences (including six-frame translations of ESTs) were filtered with SEG (Wootton and Federhen 1996). Custom Perl scripts (for scripts visit our laboratory web site at http://forestry.msu.edu/biotech/Projects/Projects_Poplar.htm) and relational databases (Microsoft Access) were used to automate searches on large sets of query sequences and to extract summary information (e.g., score and E-value of best hit). Queries were considered to have a significant similarity for E-value cutoff of ≤1.0E-20 a potential similarity for E-value ≤1.0E-10 and no significant similarity for E-value cutoff of >1.0E-5. These user-defined E-values were based on the criteria used in other comparative genomics studies (Rubin et al., 2000; Van der Hoeven et al., 2002).

Results and Discussion

Analysis of Poplar ESTs and Establishment of a Unigene Set

We downloaded a total of 105,831 ESTs (as of January 2003) from public databases that were derived from various tissues of four poplar species (Table 2-1). A hybrid aspen species (*Populus tremula* x *P. tremuloides*) contributed the highest number of ESTs (56,147) followed by *P. balsamifera* ssp. *trichocarpa* (24,050) *P. tremula* (14,091) and *P. tremuloides* (11,543). The ESTs were organized into 25,282 of unigenes as described in method, each of which may represent a potentially distinct gene (i.e., unigene). While redundant transcripts are assembled into contigs, the low-frequency ESTs are not incorporated into contig assemblies and remain as singletons. The resulting 25,282-

unigene set consists of 15,178 singletons and 10,104 contigs with an average length of 476-bp and 692-bp, respectively (Table 2-1). The 15,178 singletons consist of 9,005 from aspen, 2,892 from *P. trichocarpa*, 1,966 from *P. tremula* and 1,317 from tremuloides. The 25,282-unigenes account for about 72 % of an estimated 35,000 poplar genes (Tuskan et al., 2003). A list of the entire poplar unigenes is provided on our project web site at http://www.genomics.msu.edu/~han/poplar.html. However, we should acknowledge that the number of unigenes might have been overestimated mostly due to low quality and short-length attributes of ESTs that has been reported by other studies (Zhu et al., 2003; Rudd, 2003). The attributes may be responsible for the singletons that failed to merge into contigs. Indeed, the singletons of the poplar unigene set had shorter sequence length than did the contigs (Figure 1-2). Fifty-five percent of singletons ranged less than 500-bp while twenty percent of contigs ranged less than 500-bp. The potential drawbacks (i.e. short length and singleton) were considered during interpretation of further analyses. The singletons could have been generated by orthologs less conserved among the four species. To assess the likelihood of orthologs appearing as unigenes, we compared the singletons of three poplar species with the entire unigenes from aspen using tBLASTX algorithm. We defined as the potential orthologs the singletons that had higher similarity with aspen unigenes than with genes of any other non-poplar species (i.e. the eight species used in this study; Figure 1-1) although they could be closely related paralogs. Using the bits score as similarity criteria and less than 1E-20 as a threshold for significance of the similarity, we identified 571 singletons (275 from *P. trichocarpa*, 195 from *P. tremula* and 97 from *P. tremuloides*) as the potential orthologs. This number is

relatively small compared to the 25,282 unigenes, suggesting that the less conserved orthologs are unlikely to affect the overall conclusions at the analysis stringency we used.

A surrogate annotation approach was used to annotate the poplar unigene set. The entire poplar unigene set was searched using the tBLASTX algorithm against the *Arabidopsis* transcriptome predicted from the complete genomic sequence (TIGR Genome Databases http://www.tigr.org/tdb/). Functional categories have been assigned using the Munich Information Center for Protein Sequences (MIPS) *Arabidopsis* database (MATDB http://mips.gsf.de/proj/thal/db/search/search_frame.html) search function. Poplar unigenes with an E-value of ≤1.0E-20 were assigned to the corresponding *Arabidopsis* annotation. This approach is based on the assumption that functionality is transferable based on sequence conservation. The majority of the unigenes (56 %) had significant matches to *Arabidopsis* genes 75 % of which matched to genes of known function but the remaining 25 % were found to be unclassified or of unknown function. Of the assigned functions metabolism-related genes were the most numerous at 18 % followed by cellular organization (9 %) and signal transduction (8 %) (Figure 1-3). Other significant functional categories include transcription regulation (7 %), protein destination (7 %), protein synthesis (4 %), transport facilitation (5 %) and cell division (3 %). Kirst et al., (2003) recently reported that ~90 % of pine ESTs had counterparts in *Arabidopsis* sequences (E value < 1.0E-10). The 56 % match in this study is substantially lower than their estimate. The difference might be explained by the facts that many of the poplar unigenes used in our analysis were not full-length (average length of 572-bp) and different cut-off values (1.0E-10 in pine study and 1.0E-20 in the current study) were used. Indeed, the unmatched unigenes had shorter sequence length with an average of

490-bp than did the matched ones with an average of 645-bp. In addition, this functional classification analysis was performed solely based on ESTs, whose representation is highly dependent on the source libraries and expression level. Genes with low level of expression might not have been represented in this analysis. This notion is supported by the fact that a large number of Arabidopsis ESTs (178,000 ESTs) from various sources representing over 60 distinct tissues, biotic and abiotic challenges, and developmental stages matched to only to 16,115 unigenes (reviewed in Rudd 2003). Likewise, in the current classification analysis the presence of an EST is a reliable attribute while the absence of an EST does not necessarily mean that the gene is not present in the genome and not expressed. Completion of poplar genome sequencing will provide a better opportunity to examine the differences between poplar and other plant species.

Poplar Multigene Families Estimated by tBLASTX Analysis of Unigenes

One of the characteristic features of eukaryotic genomes is that a significant fraction of protein-coding genes belong to multigene families that are likely derived from gene duplication. For example, up to 80 % of the *Arabidopsis* genes are members of multigene families (Lespinet et al., 2002). The main contributing forces for organizational and regulatory diversity in eukaryotes include the changes in domain architectures *via* domain accretion and shuffling gene loss in a particular lineage and lineage-specific proliferation of protein families (Aravind et al., 2000; Lander et al., 2001; Lespinet et al., 2002). Recent advances in structural and comparative genomics have provided valuable information about gene family organization in diverse plant species.

In order to describe the gene family organization in poplar, we first computationally determined the number and size of poplar gene families by counting the tBLASTX matches of each unigene against the poplar unigene set itself with an E-value threshold of ≤1.0E-20. It is important to note that these gene families do not necessarily indicate functional groupings. Of the 25,282 poplar unigenes, 11,909 (47 %) had significant matches with other poplar unigenes at this threshold and were subsequently organized into 3,481 gene families. To determine the potential bias caused by short sequence length or singleton, we performed the same analysis using unigenes greater than 500-bp or excluding singletons (i.e., contigs). Of 15,088 poplar unigenes greater than 500-bp, 9,448 (63 %) were grouped into 2,033 gene families. There were 10,104 poplar contigs (25,282 minus singletons), 5,401 (53 %) of which had significant matches with other poplar unigenes, resulting in 1,995 gene families. The distribution of the gene family sizes is shown in Figure 1-4. Overall, the majority (56-55 %) had two copy members. Gene families with fewer than five copy members accounted for around 90 % of the entire gene families estimated in this study. In order to assess how the copy number of the gene families was conserved between the two species, we have compared the copy numbers of the 14 largest poplar gene families with those of corresponding *Arabidopsis* gene families. The Arabidopsis gene copy number was obtained by subjecting the *Arabidopsis* gene set (TIGR http://www.tigr.org/tdb/) to the same analysis as the poplar tBLASTX. In order to increase the stringency and therefore not to overestimate the poplar gene copy number in the analysis the initially tBLASTX-determined copy numbers of the 14 largest poplar gene families were subjected to several refinement steps. as described in Table 2-2. First, we removed the sequences that had no

matches to the known multigene family proteins in public databases. Then, the sequences that had less than 90 % identity with any member of the multigene family with over 90 % length region were removed. Finally, we removed the sequences that either did not share the conserved region of the protein family due to its short length, or had long gap regions probably resulting from chimera when they were aligned using ClustalW software (Thompson et al., 1994). The resulting alignments are provided as supplemental materials (http://forestry.msu.edu/biotech/Projects/Projects_Poplar.htm). Table 2-2 lists the 14 highest copy number poplar gene families along with their corresponding *Arabidopsis* gene families. In poplar protein kinases represent the largest family followed by Ras-related GTP-binding protein and Myb transcription factor. Several multigene families had lower copy numbers in poplar than in *Arabidopsis* (Table 2-2). However caution should be exercised in interpreting these results because two factors exist that may lead to a substantial underestimate of copy numbers for each gene family. First, many of the 25,282 poplar unigene sequences used in the analysis are partial sequences that generally result in higher E-values (i.e. lower similarity) compared to full-length sequences. It is quite possible that the tBLASTX search with a short coding region of partial EST sequences might miss the consensus motif(s) characteristics of the gene family if they are located in the middle of the protein. Second, the 25,282 poplar unigenes do not represent the entire poplar transcriptome while *Arabidopsis* sequences represent whole transcriptome on its genome. Possibly the genes with low expression levels were not picked up during the EST sequencing resulting in a lower copy number estimation than actual number. This may account for the lower copy number of some multigene families in poplar when compared to *Arabidopsis*.

Five of the 14 largest poplar multigene families had a relatively higher copy number in poplar than in *Arabidopsis* (Table 2-2). These families include ubiquitin-domain protein, ubiquitin-conjugating enzyme, tubulin, peptidylprolyl isomerase and ABA-inducible protein. The proteins of these families had predicted biochemical characteristics that suggest their roles in protein degradation, energy, stress response, and cellular structure. The higher copy number of ubiquitin gene families in poplar may represent an evolutionary adaptation important for high protein turnover throughout different seasons. Interestingly functionally related two multigene families such as ubiquitin-conjugating enzyme and peptidylprolyl isomerase which are involved in protein degradation and folding respectively (Hershko and Ciechanover, 1998; Schiene and Fischer 2000) have a slightly higher copy number in poplar than in *Arabidopsis* (Table 2-2) and indeed these genes were more abundantly expressed in senescent autumn leaves than in the young summer leaves of poplar (Bhalerao et al., 2003). Another example of potential adaptational change in poplar is the tubulin gene family that has 19 members in poplar and 17 in the *Arabidopsis* genome. Perennial nature of poplar growth and development requires adaptive evolution for cold hardiness. The microtubules of higher plants have been suggested to participate in low temperature stress response and adaptation (Nick, 1998; Nyporko et al., 2003). In fact those factors known to increase plant cold hardiness were also involved in the induction of elevated cold-stability in microtubules (Pihakaskimaunsbach and Puhakainen 1995). Also it has been shown that depolymerization of microtubules is related to freezing injury in cotton (*Gossypium hirsutum* L.) (Rikin et al., 1980). The cold-resistance of microtubules appears to be correlated with their cold stability (Detrich 1997; Gupta et al., 2001). Parker and Detrich

(1998) reported that an Antarctic fish (*Notothenia coriiceps*) had a substantial number (~15) of α-tubulin subunits encoded by a multigene family suggesting that the expansion of tubulin gene families might have occurred as a result of adaptive evolution for low temperatures. Considering these observations in the literature it is not surprising that poplar had an increase in copy number for this gene family. Another multigene family with a higher copy number in poplar encodes ABA-inducible protein which is a subfamily of late embryogenesis abundant (LEA) protein family (Wise, 2003). These hydrophilic proteins have been suggested to be involved in the protection of cellular structures from dehydration or hyper-osmotic stress by acting as a hydration buffer by sequestration of ions by direct protection of other proteins or membranes or by re-naturation of unfolded proteins (Bray et al., 1993; Soulages et al., 2002; Koag et al., 2003).

Multigene families with higher copy numbers could imply greater functional diversity providing resources for specific adaptations and evolution of new functional systems. Considering that the poplar genome is approximately four times larger than the *Arabidopsis* genome and only about 72 % of the transcriptome is represented in this analysis it is reasonable to assume that the copy number of each poplar family should be less than or at most equal to that of corresponding *Arabidopsis* families if there is no poplar specific gene family expansion since divergence from the last common ancestor. Nonetheless we noted several multigene families had relatively higher copy numbers in poplar than in *Arabidopsis*. Potential physiological roles of the protein families (as described above) suggest that this kind of multigene family expansion might be common among the perennial temperate tree species. However it should be noted that the copy

numbers of multigene families could also be overestimated considering the high rate of error in EST datasets (Ewing and Green 1998; Hillier et al., 1996), and highly expressed gene families could be overrepresented during clustering analysis.

Comparison of Poplar with Other Plant Species

Arabidopsis is an herbaceous non-woody plant that does not undergo secondary growth under normal conditions. However it has recently been shown that *Arabidopsis* can be induced to express all of the major components of secondary growth (Lev-Yadun, 1994; Zhao et al., 2000; Chaffey et al., 2002). It is therefore of great scientific interest to examine the degree of sequence similarity in expressed gene sequences between *Arabidopsis* an herbaceous plant and poplar a perennial woody species. In order to determine whether certain functional classes of genes represent characteristic differences between the perennial tree species (i.e. polar) and an annual herbaceous plant (i.e. *Arabidopsis*) we carried out computational comparisons of the 25,282 poplar unigenes in all translated frames (tBLASTX algorithm) with *Arabidopsis* transcriptome predicted from the complete genomic sequence using the E-values of tBLASTX searches as an estimate of sequence conservation. Figure 1-5 displays the distribution of E-value matches with regard to functional categories for the poplar unigene set. Of the 25,282 poplar unigenes, 58 % show strong similarity (≤1.0E-20) with their *Arabidopsis* counterparts. Even at the strongest match threshold (≤1.0E-100), eight percent (1,907) of the 25,282 poplar unigenes matched with *Arabidopsis* sequences showing a very high level of conservation. While the frequencies of genes in most functional categories were similar among the five different match stringency (E-value) categories those of genes

involved in a few cellular functions changed significantly as the match stringency increased. For example, the proportion of genes belonging to metabolism category decreased from 29 % at the strongest match stringency (≤1.0E-100) to 12 % when the E-value cutoff was reduced to 1.0E-20 ~ 1.1E-30 suggesting that metabolic functions are more constant across plant species (i.e. more ancestral gene functions) (Figure 1-5). On the other hand, transcription regulation-related genes showed opposite trends. Their frequencies decreased from 9 to 3 % as the E-value stringency changed from 1.0E-20 to 1.0E-100. This along with the fact that *Arabidopsis* can be induced to undergo secondary growth suggests that the major differences between poplar and *Arabidopsis* (e.g. secondary growth perennial growth habit) may be in transcriptional control rather than in structural genes. Genes of unknown function are among the 'less-conserved.' For example 33 % of the poplar unigenes having relatively weak similarity (≤1.0E-20) with their *Arabidopsis* counterparts were categorized as 'unclassified' function while only 9 % of those with the strongest similarity (≤1.0E-100) matched to *Arabidopsis* genes for which no putative functions have been assigned. Genes involved in most other cellular functions had similar frequencies among the five different match stringency categories suggesting that they are relatively fast evolving. The differential distribution of the functional categories based on the match stringencies was confirmed by contingency test (P <1E-5).

In order to investigate whether the highly conserved or divergent genes between poplar and *Arabidopsis* are also conserved or divergent in other vascular plants we computationally compared the poplar unigenes with ca. 394,100 expressed sequences including putative ORFs from the genome sequences of four dicots (*Arabidopsis* barrel

medic, soybean, tomato) three monocots (maize, rice, wheat) and one gymnosperm (pine) (Figure 1-1). For this analysis the 394,100 plant gene sequences were divided into three groups ('dicots', 'monocots' and 'pine'). All sequences from *Glycine max, Medicago truncatula, Arabidopsis thaliana, and Lycopersicon esculentum* were pooled into the 'dicots' group which had a total of 151,832 sequences. Likewise the sequences from *Oryza sativa, Zea mays and Triticum aestivum* went into the 'monocots' group making a total of 220,859 sequences. The 'pine' group represents only one species (*Pinus taeda*) with 21,409 ESTs available. A large number of poplar unigenes did not have any detectable homologues to other plant species, ranging from 6,835 (27 %) (an E-value threshold of ≤1.0E-10) to 9,398 (37 %) (≤1.0E-20). However, about 30 % (7,581) of the 25,282 poplar unigenes had significant matches (≤1.0E-20) in all of the three groups suggesting their involvement with common biological functions in plant species. The proportions in each functional category of the 7,581 unigenes that had counterparts in all of the three groups were very similar to those of the entire poplar unigene set. The largest functional category was 'metabolism' (20 %) followed by 'unclassified proteins' (18 %) and 'signal transduction' (9 %) (Figure 1-6). A total of 7,760 poplar unigenes matched to pine ESTs with E-value cutoff of ≤1.0E-20 while 13,422 and 15,257 poplar unigenes had significant matches to monocot and dicot sequences respectively (≤1.0E-20). The number of poplar unigenes that matched to monocot sequences represents a substantially high percentage (53 %) of the 25,282 poplar unigenes.

Conclusion

About 72 % of the 35,000 estimated poplar genes were identified in the unigene set described in this report. High proportions of the unigenes had counterparts in the

genomes of *Arabidopsis* and other plant species supporting the hypothesis that many of the functional genes are conserved among different plant species. We estimate that at least 47 % of poplar genes belong to multigene families based on the observation that 11,909 of the 25,282 unigenes had significant tBLASTX matches (≤1.0E-20) to other poplar unigenes. Comparison of poplar multigene families with those of *Arabidopsis* suggests that the expansion of poplar multigene families occurred in support of the adaptation to various stresses of poplar growth. Another striking feature of the difference between poplar and *Arabidopsis* transcriptome is that the genes involved in transcriptional regulation are the most divergent while metabolism-related genes are highly conserved. While the findings described in this report increase our understanding of the poplar transcriptome it should be noted that completion of the on-going poplar genome sequencing would provide an invaluable opportunity to verify these findings and advance our knowledge of tree growth.

Acknowledgement

We would like to thank the Genomics Technology Support Facility staff at Michigan State University for their help and consultation with the bioinformatics analysis. This work was supported by USDA Grants to the Eastern Hardwood Utilization Program at Michigan State University (No. 98-34158-5995 00-34158-9236 and 01-34158-11222).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-402

Aravind L, Watanabe H, Lipman DJ, Koonin EV (2000) Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci U S A. 97: 11319-24

Bergelson J, Dwyer G, Emerson JJ (2001) Models and data on plant-enemy coevolution. Annu Rev Genet. 35: 469-99

Bessler JB, Torredagger JZ, Zakian VA (2001) The Pif1p subfamily of helicases: region-specific DNA helicases? Trends Cell Biol. 11: 60-5

Bhalerao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J,Jansson S (2003) Gene expression in autumn leaves. Plant Physiol. 131: 430-42

Bray EA (1993) Molecular Responses to Water Deficit. Plant Physiol. 103: 1035-1040

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

Chou A,Burke J (1999) CRAWview: for viewing splicing variation, gene families, and polymorphism in clusters of ESTs and full-length sequences. Bioinformatics. 15: 376-81

Coleman GD, Banados MP, Chen T (1994) Poplar Bark Storage Protein and a Related Wound-Induced Gene Are Differentially Induced by Nitrogen. Plant Physiol. 106: 211-215

Dean C, Pichersky E, Dunsmuir P (1989) Structure, Evolution, and Regulation of Rbcs Genes in Higher- Plants. Annu. Rev. Plant Physiol. Plant Molec. Biol. 40: 415-439

Detrich HW, 3rd (1997) Microtubule assembly in cold-adapted organisms: functional properties and structural adaptations of tubulins from antarctic fishes. Comp Biochem Physiol A Physiol. 118: 501-13

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8: 186-94

Galili S, Avivi Y, Millet E, Feldman M (2000) RFLP-based analysis of three RbcS subfamilies in diploid and polyploid species of wheat. Mol Gen Genet. 263: 674-80

Gupta ML, Jr., Bode CJ, Dougherty CA, Marquez RT, Himes RH (2001) Mutagenesis of beta-tubulin cysteine residues in Saccharomyces cerevisiae: mutation of cysteine 354 results in cold-stable microtubules. Cell Motil Cytoskeleton. 49: 67-77

Haruta M, Major IT, Christopher ME, Patton JJ, Constabel CP (2001) A Kunitz trypsin inhibitor gene family from trembling aspen (Populus tremuloides Michx.): cloning, functional expression, and induction by wounding and herbivory. Plant Mol Biol. 46: 347-59

Hershko A,Ciechanover A (1998) The ubiquitin system. Annu Rev Biochem. 67: 425-79

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A transcriptional roadmap to wood formation. Proc Natl Acad Sci U S A. 98: 14732-7

Hide W, Burke J,Davison DB (1994) Biological evaluation of d2, an algorithm for high-performance sequence comparison. J Comput Biol. 1: 199-215

Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, Dietrich N, DuBuque T, Favello A, Gish W, Hawkins M, Hultman M, Kucaba T, Lacy M, Le M, Le N, Mardis E, Moore B, Morris M, Parsons J, Prange C, Rifkin L, Rohlfing T, Schellenberg K, Marra M,et al. (1996) Generation and analysis of 280,000 human expressed sequence tags. Genome Res. 6: 807-28

Keeling PJ,Palmer JD (2001) Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. Proc Natl Acad Sci U S A. 98: 10745-50

Koag MC, Fenton RD, Wilkens S, Close TJ (2003) The binding of Maize DHN1 to Lipid Vesicles. Gain of Structure and Lipid Specificity. Plant Physiol. 131: 309-16

Kohler A, Delaruelle C, Martin D, Encelot N,Martin F (2003) The poplar root transcriptome: analysis of 7000 expressed sequence tags. FEBS Lett. 542: 37-41

Koiwa H, Bressan RA, Hasegawa PM (1997) Regulation of protease inhibitors and plant defense. Trends Plant Sci. 2: 379-384

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W,al. e (2001) Initial sequencing and analysis of the human genome. Nature. 409: 860-921

Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12: 1048-59

Levyadun S (1994) Induction of Sclereid Differentiation in the Pith of Arabidopsis-Thaliana (L) Heynh. J. Exp. Bot. 45: 1845-1849

Mandal S, Kundu P, Roy B, Mandal RK (2002) Precursor of the inactive 2S seed storage protein from the Indian mustard Brassica juncea is a novel trypsin inhibitor. Charaterization, post-translational processing studies, and transgenic expression to develop insect-resistant plants. J Biol Chem. 277: 37161-8

Michalek W, Weschke W, Pleissner KP,Graner A (2002) EST analysis in barley defines a unigene set comprising 4,000 genes. Theor Appl Genet. 104: 97-103

Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR,Hide WA (1999) A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Res. 9: 1143-55

Nick P: Signaling to the microtubular cytoskeleton in plants International Review of Cytology - a Survey of Cell Biology, Vol 184, pp. 33-80 (1998).

Nyporko AY, Demchuk ON,Blume YB (2003) Cold adaptation of plant microtubules: structural interpretation of primary sequence changes in a highly conserved region of alpha-tubulin. Cell Biol Int. 27: 241-3

Oguri S, Kamoshida M, Nagata Y, Momonoki YS,Kamimura H (2003) Characterization and sequence of tomato 2S seed albumin: a storage protein with sequence similarities to the fruit lectin. Planta. 216: 976-84

Parker SK,Detrich HW, 3rd (1998) Evolution, organization, and expression of alpha-tubulin genes in the antarctic fish Notothenia coriiceps. Adaptive expansion of a gene family by recent gene duplication, inversion, and divergence. J Biol Chem. 273: 34358-69

Pihakaskimaunsbach K,Puhakainen T (1995) Effect of Cold-Exposure on Cortical Microtubules of Rye (Secale-Cereale) as Observed by Immunocytochemistry. Physiol. Plant. 93: 563-571

Quackenbush J, Liang F, Holt I, Pertea G,Upton J (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. 28: 141-5

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR,al. e (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 290: 2105-10

Rikin A, Atsmon D,Gitler C (1980) Chilling Injury in Cotton (Gossypium-Hirsutum-L) - Effects of Anti-Microtubular Drugs. Plant Cell Physiol. 21: 829-837

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LSnB, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJnM, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XH, Zhong F, Zhong W, Gibbs R, Venter JC, Adams MD,Lewis S (2000) Comparative Genomics of the Eukaryotes. Science. 287: 2204-2215

Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. 8: 321-329

Schiene C, Fischer G (2000) Enzymes that catalyse the restructuring of proteins. Curr Opin Struct Biol. 10: 40-5

Soltis PS, Soltis DE,Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature. 402: 402-4

Soulages JL, Kim K, Walters C,Cushman JC (2002) Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. Plant Physiol. 128: 822-32

Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B, 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: 13330-5

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: 4673-80

Trail F, Xu JR, San Miguel P, Halgren RG, Kistler HC (2003) Analysis of expressed sequence tags from Gibberella zeae (anamorph Fusarium graminearum). Fungal Genet Biol. 38: 187-97

Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S (2002) Deductions about the Number, Organization, and Evolution of Genes in the Tomato Genome Based on Analysis of a Large Expressed Sequence Tag Collection and Selective Genomic Sequencing. Plant Cell. 14: 1441-1456

Wang WX, Pelah D, Alergand T, Shoseyov O,Altman A (2002) Characterization of SP1, a stress-responsive, boiling-soluble, homo-oligomeric protein from aspen. Plant Physiol. 130: 865-75

Wetzel S, Demmers C,Greenwood JS (1989) Seasonally Fluctuating Bark Proteins Are a Potential Form of Nitrogen Storage in 3 Temperate Hardwoods. Planta. 178: 275-281

Wootton JC,Federhen S (1996) Analysis of compositionally biased regions in sequence databases. Methods Enzymol. 266: 554-71

Yoon M, Putterill JJ, Ross GS, Laing WA (2001) Determination of the relative expression levels of rubisco small subunit genes in Arabidopsis by rapid amplification of cDNA ends. Anal Biochem. 291: 237-44

Zardoya R, Ding X, Kitagawa Y, Chrispeels MJ (2002) Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment. Proc Natl Acad Sci U S A. 99: 14893-6

Zhao C, Johnson BJ, Kositsup B, Beers EP (2000) Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol. 123: 1185-96

Zhou J, Monson EK, Teng S, Schulz VP, Zakian VA (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. Science. 289: 771-4

Zhu B, Coleman GD (2001) The poplar bark storage protein gene (Bspa) promoter is responsive to photoperiod and nitrogen in transgenic poplar and active in floral tissues, immature seeds and germinating seeds of transgenic tobacco. Plant Mol Biol. 46: 383-94

Zhu W, Schlueter SD,Brendel V (2003) Refined annotation of the Arabidopsis genome by complete expressed sequence tag mapping. Plant Physiology. 132: 469-484

Web Site References

http://www.tigr.org/tbd/plant.shtml; TIGR Gene Index Databases..

ftp://ftp.ncbi.nih.gov/blast/executables; Stand-alone BLAST program.

ftp://ftp.ncbi.nih.gov/blast/documents/README.formatdb; Documentation for the Formatdb program.

ftp://ftp.ncbi.nih.gov/blast/documents/README.formatdb; Documentation for the BLASTCLUST program.

ftp://ftp.ncbi.nih.gov/blast/db; NCBI Public Databases.

http://www.tigr.org/tdb/; The TIGR Genome Databases

http://mips.gsf.de/proj/thal/db/search/search_frame.html; MATDB Search Page

http://www.egenetics.com; Electric Genetics Home Page.

http://forestry.msu.edu/biotech/Projects/Projects_Poplar.htm The Han Lab Project Web Site

http://www.genomics.msu.edu/~han/poplar.html Searchable web site for the data presented.

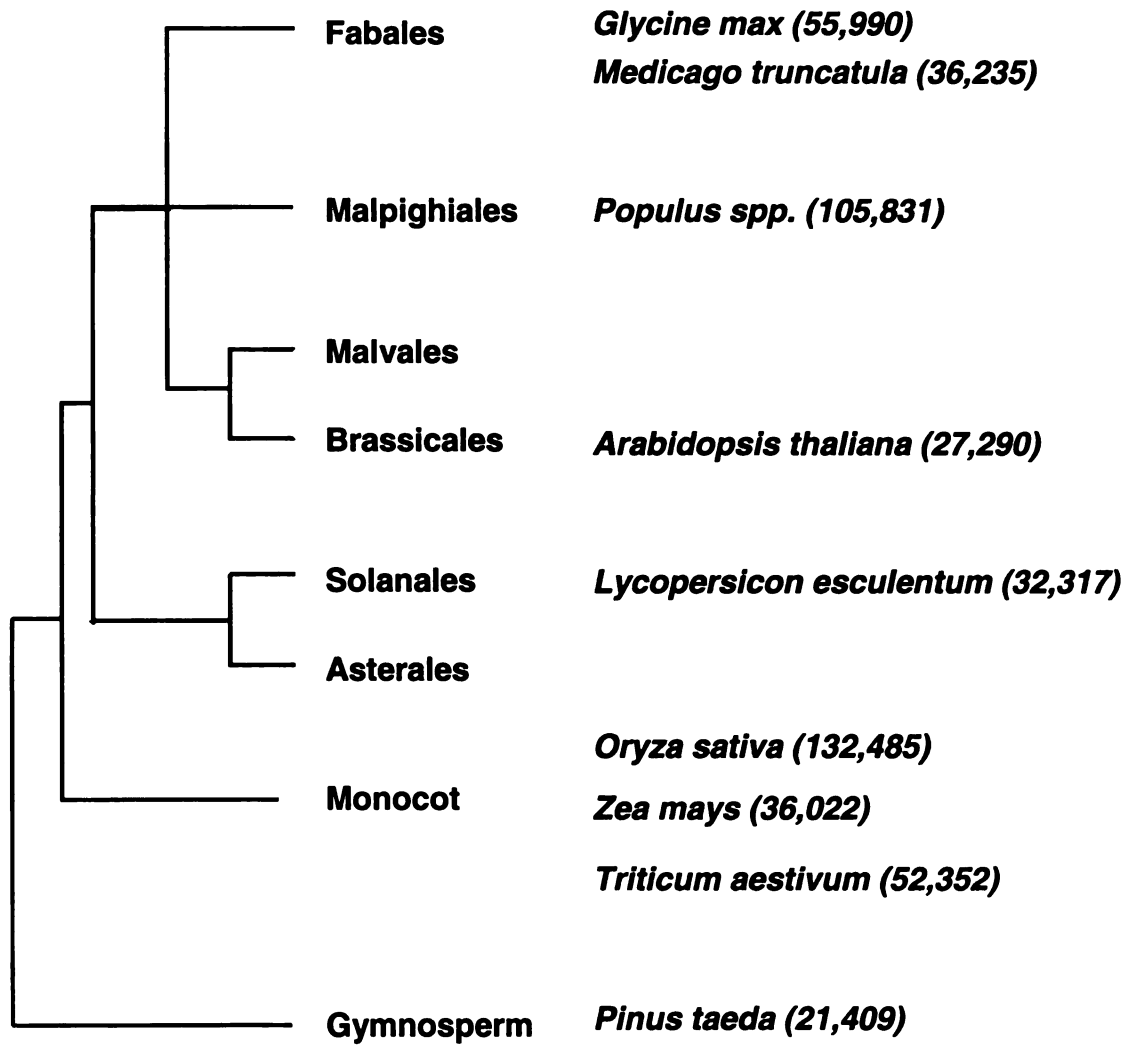


Figure 1-1. Phylogenetic relationships for the nine plant species used in the sequence analyses.
The number of sequences used in the analyses for each species is given in parenthesis. The phylogenetic tree was adapted from Soltis et al. (1999) and abbreviated only to reflect the relative branching order of each species involved.

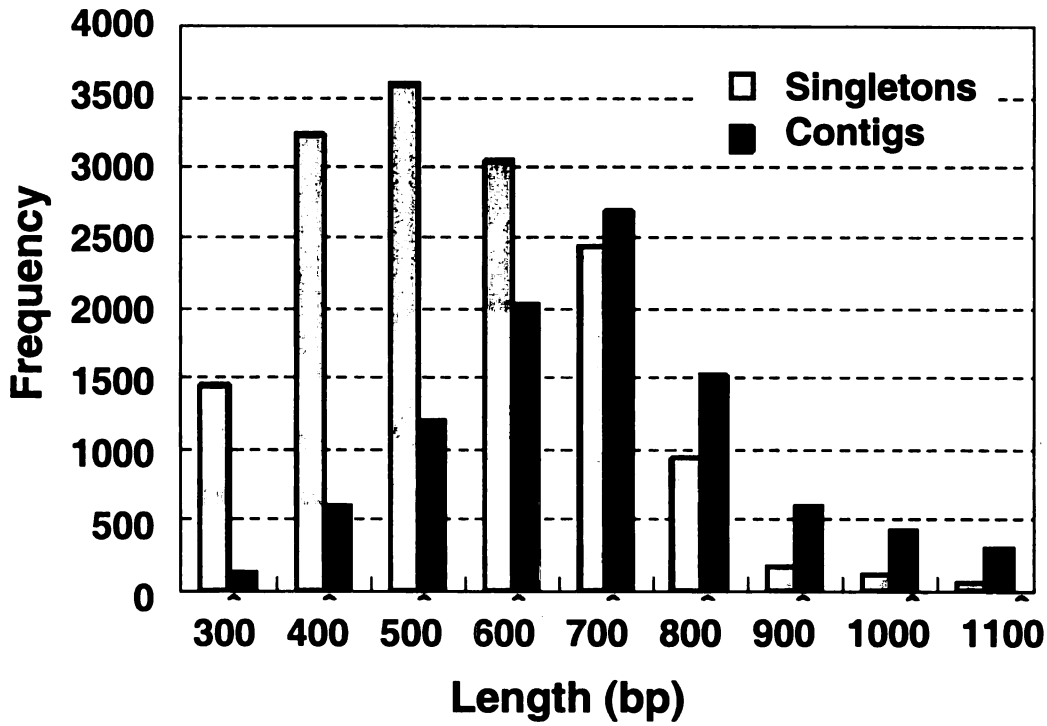


Figure 1-2. Distribution of sequence length (bp) of singletons and contigs that constitute the poplar unigene set.

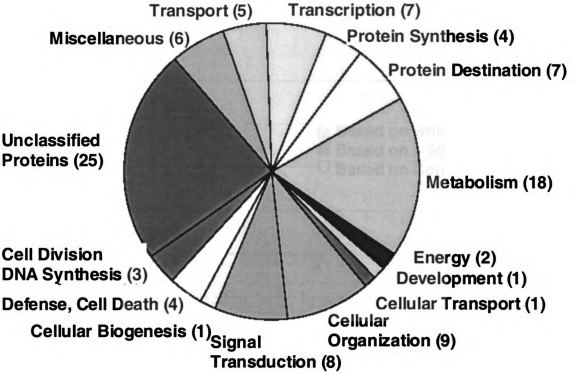


Figure 1-3. Functional categorization of the 14,291 poplar unigenes that had tBLASTX matches to *Arabidopsis* transcriptome with E-value threshold of ≤1.0E-20. The 10,991 unigenes that did not have matches were excluded from the chart. Functional categories are according to the Munich Information Center For Protein Sequence (MIPS http://mips.gsf.de). Percentage of the unigenes in each functional category is given in parenthesis.

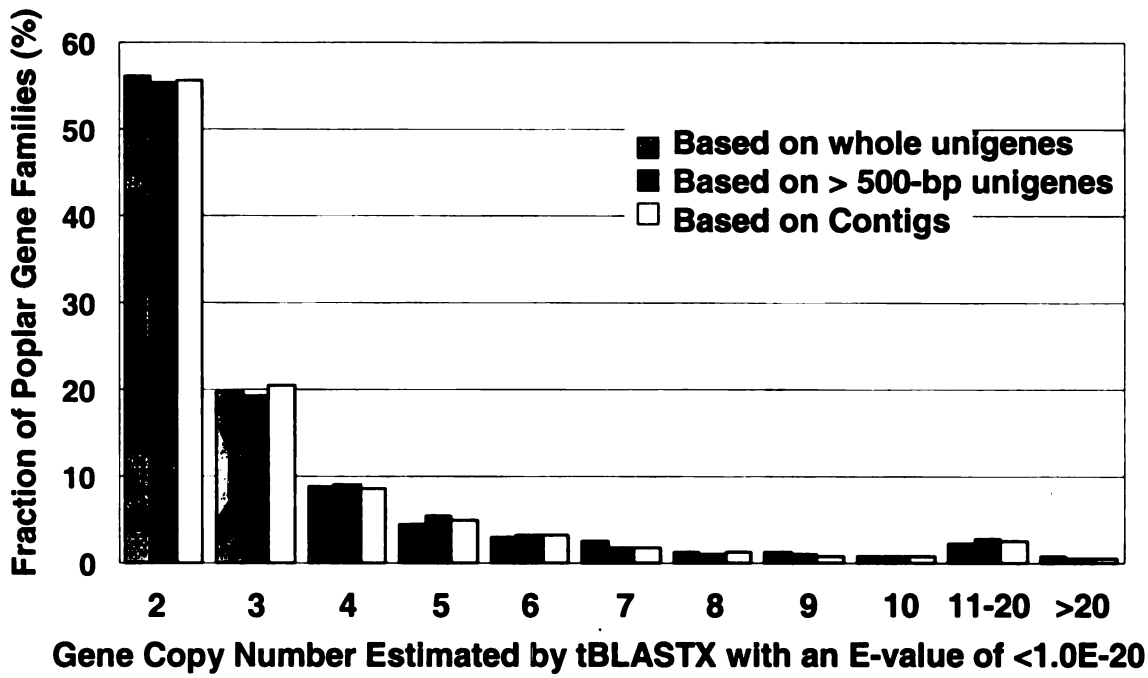


Figure 1-4. Gene copy number distribution of putative poplar gene families estimated by tBLASTX search against the poplar unigene data set itself.
A total of 3,481, 2,303 or 1,995 multigene families were identified based on different stringency, and about 56 % of which had two copy members.

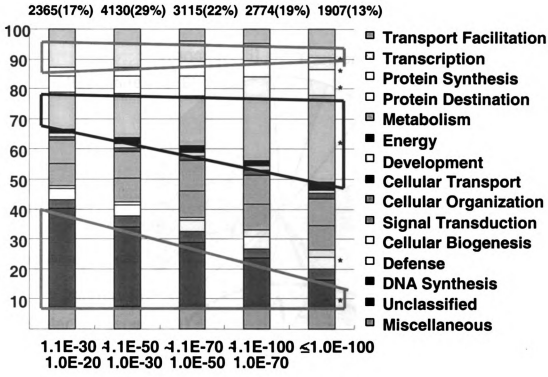


Figure 1-5. Percentage of poplar unigenes belonging to each functional category with different E-value thresholds.

A total of 14,291 sequences had significant tBLASTX matches to the genes of known function according to MIPS (http://mips.gsf.de) functional categorization. The actual number and percentage of sequences matched with ranges of 1.0E-20~1.1E-30 1.1E-30~1.0E-50 1.1E-30~1.0E-50 1.1E-50~1.0E-70 1.1E-70~1.0E-100 and ≤ 1.0E-100 respectively were indicated on top of the graph. Asterisks indicated that the differed distribution of the categories was supported by contingency test ($p < 1E-5$).

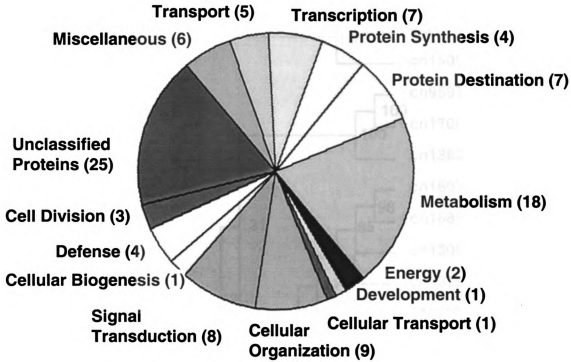


Figure 1-6. Functional categorization of the 7,760 poplar unigenes that had counterparts in all three groups (dicots monocots and pine) of plant species. Percentage of the unigenes in each functional category is given in parenthesis. For this analysis the plant gene sequences were pooled into three groups. All of the sequences from *Glycine max Medicago truncatula Arabidopsis thaliana Lycopersicon esculentum* were pooled into the 'dicots' group which had a total of 151,832 sequences. Likewise, the sequences from *Oryza sativa Zea mays Triticum aestivum* into the 'monocots' group with a total of 220,859 sequences.

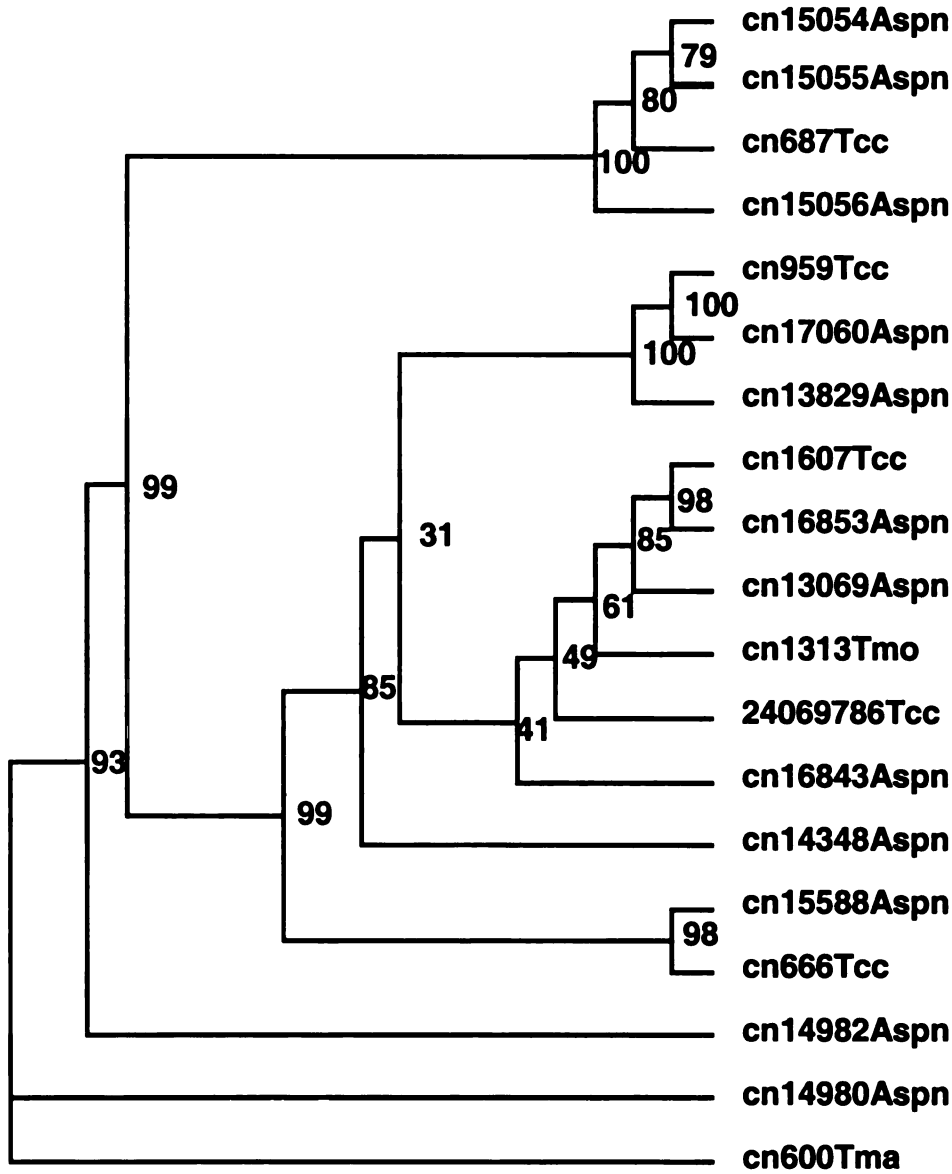


Figure 1-7. Phylogenic tree of DNA sequences of peptidylprolyl isomerase family, constructed using the neighbor-joining method.

To support the gene duplication within the families, phylogenic analysis using peptidylprolyl isomerase protein family were performed as an example. Bootstrap confidence values (%) based on 100 replications are shown for each branch.

Table 1-1. Poplar ESTs and unigene set statistics.

Species	Number of ESTs	Number of Unigenes	Average Length (bp)
P. tremula x P. tremuloides	56147	20864	511
P. trichocarpa	24050	9113	542
P. tremula	14091	7295	461
P. tremuloides	11543	4683	507
Total	**105831**	**41955**	ND [d]
Poplar unigenes identified by clustering the combined 41955 sequences	41955 [a]	29818	ND
Filtering with BLASTCLUST		29644 [b]	ND
Final Poplar Unigene Set		**25282** [c]	**572**

[a] The sequences consists of four species unigene sets.

[b] BLASTCLUST algorithm was used to further remove any sequences with 95 % nucleic acid identity covering 90 % length region.

[c] Any ESTs shorter than 250-bp sequence read were discarded in order to reduce sequence length bias in subsequent analyses. The poplar unigene set consist of 15,178 singletons and 10,104 contigs

[d] ND not determined.

Table 1-2. The 14 Highest copy number poplar unigene families determined by tBLASTX with an E-Value threshold of <1.0E-20.

Poplar Contig (Gene) Family	GenBank Accession Number	Copy No.	Putative Function of the Gene Family	Representative Gene of Corresponding *Arabidopsis* Family	*Arabidopsis* Gene Copy Number
cn726Aspn	BU896558	53[a] (73[b])	Protein kinase	At1g56140	501
cn8047Aspn	BU894782	33 (53)	Ubiquitin-domain protein	At4g02890	19
cn16304Aspn	BU810510	41 (48)	GTP-binding protein	At3g46830	60
24067636Tcc	BU876112	31 (44)	Myb transcription factor	At4g18770	133
cn15003Aspn	BU894972	33 (37)	Ubiquitin-conjugating enzyme	At3g08690	30
cn933Aspn	BU895108	25 (34)	Peroxidase	At2g18150	73
cn6614Aspn	BU831012	24 (33)	Cytochrome P450	At4g37400	152
cn15387Aspn	BU831392	24 (31)	ADP-ribosylation factor	At2g47170	25
cn11888Aspn	BI128682	19 (30)	Tubulin	At1g50010	17
cn14980Aspn	BU831411	20 (29)	Peptidylprolyl isomerase[c]	At4g34870	20
24074413Tcc	BU882889	21 (28)	MADS-box transcription factor	At4g18960	41[d]
cn753Tcc	BU882276	18 (28)	Water channel protein	At3g16240	24
cn51Tcc	BU874797	7 (27)	ABA-inducible protein	At5g38760	3
23534478Tmo	AY095297	18 (26)	Cellulose synthase	At4g39350	26

[a] The copy number estimation was further refined by excluding the genes that 1) had no matches to the known multigene family proteins present in NCBI protein database, 2) were shorter than 500-bp, 3) shared <90 % nucleic acid sequence identity with any member of the multigene family over 90 % length region, 4) did not share the conserved region mostly due to its short length, or 5) had long gap regions probably resulting from chimera when they were aligned using ClustalW software (Thompson et al. 1994).

[b] The copy number determined using tBLASTX with an E-value threshold of ≤1.0E-20.

[c] To support the gene duplication within the families, we performed phylogenic analysis using peptidylprolyl isomerase protein family as an example, which was provided as Figure 1-7.

[d] Actual number of *Arabidopsis* MADS-box genes that can be identified by MADS-box domain search is ~80 (Riechmann et al., 2000). The number 41 reflects the outcome of tBLASTX search using the entire cDNA sequence.

CHAPTER 2

Genetic regulation of annual growth cycle in woody perennial, *Populus*

Abstract

Temperate woody plants have evolved annual growth cycle as an adaptive mechanism for winter survival, which is controlled by interactions between endogenous and environmental factors (e.g., daylength, temperature). To understand the genetic regulation underlying this process, we carried out a series of global transcriptional analysis using stem samples from poplar trees grown in field and under controlled environment conditions. The analysis showed that extensive metabolic switches and alterations in cellular functions occurred during transition from rest to quiescent. The genes involved in the pentose phosphate pathway were elevated and defense-related genes were over-represented during early winter whereas the genes involved in fermentation and fatty acid β-oxidation were upregulated. Moreover, signaling-related genes were over-represented during late winter. Furthermore, clustering of these genes into eight groups has revealed that plants regulate seasonal growth by integrating environmental factors with developmental stages. For example, short daylength (SD) could modulate the cold signal in at least two different ways to achieve maximum cold hardiness. First, short daylength could induce cold associated genes independent of actual cold temperatures, resulting in a higher expression of the genes in both cold and SD than in cold alone. Second, the short daylength itself did not induce some of the genes associated with cold acclimation, but did enhance their expression in response to cold temperature. This mechanism may help

plants to control cold hardiness more closely timed with seasonality than unpredictable temperature alone does.

Introduction

In the temperate zone, woody perennial plants exhibit an annual cycle alternating between growth and dormancy. By integrating information from multiple environmental cues, they can coordinate their development with seasonal changes, and ultimately, control their growth capacity. During the warm summer season, woody plants grow actively, and at the end of summer they enter dormancy and initiate cold acclimation. At this stage, cambial dormancy is regulated by internal factors, which has been denoted rest and characterized by non-responsiveness to indole-3-acetic acid (IAA) under favorable conditions for cambial activity (Little and Bonga 1974, Little 1981). Short daylength (SD) has been known to be one of the most important environmental factors controlling the onset of dormancy and cold acclimation (Greet et al., 1989; Olsen et al., 1997). By responding to SD concordant with local seasonal change, plants are able to synchronize dormancy development and cold acclimation with the coming winter cold. Previous studies have showed that cold acclimation facilitated by low temperature (LT) could be uncoupled from dormancy, suggesting that SD and LT might utilize independent pathways for the cold acclimation (Welling et al., 2002). However, LT and dormancy development would be needed for plants to achieve the higher degree of cold hardiness, i.e., second stage of cold hardiness (Anisko et al., 1994; Christersson, 1978). Indeed, dormant deciduous genotypes always reach a higher extent of cold hardiness than the non-dormant evergreen genotypes, as discovered in peach (Arora and Wisniewski, 1994). In nature, dormant plants could induce a second stage of cold hardiness by responding to the subsequent low or freezing temperatures that occur in late fall. The facts indicated the possible interaction between dormancy development and cold acclimation. Cold is also

required to break dormancy, after which the cambium gradually regains the ability to produce xylem in response to IAA, referred to as a chilling requirement (Little and Bonga, 1974). When fully responsive, the cambium is in the quiescence stage of dormancy, which is imposed solely by adverse external factors, typically by low temperature. Upon the return of warmer temperatures in the spring, plants are gradually dehardened, primarily by responding to the rising temperature, and the cambial cells are released from dormancy (Mellerowicz et al., 1992). The periodicity regulated timely with seasonal changes is a dominant factor in the distribution, growth, and survival of woody plants (Samish, 1954; Weiser, 1970).

Despite these early observations, the molecular basis of seasonal development remains a mystery. Perhaps this lack of understanding of the molecular mechanism underlying seasonal development has been, in part, due to superimposition of various internal and external events occurring during annual growth cycle, which make it difficult to associate molecular changes with specific events. In nature, dormancy development and cold acclimation/deacclimation partially overlap and are interrelated. Thus, it is important to identify which of the changes is associated with specific events in the seasonal growth cycle. Although there were some previous studies that have identified the genes associated with different aspects of dormancy or cold acclimation, they were limited to a few of genes or a short period of study time (Espinosa-Ruiz et al., 2004; Rowland and Arora, 1997; Welling et al., 2004).

To understand the molecular basis of seasonal development, we sought to obtain global transcriptional profiles of eastern cottonwood (*Populus deltoides*) on an annual cycle as well as under controlled environmental conditions. We identified 714 and 743 of

genes that were differentially regulated during seasons in bark and xylem, respectively. Clustering and functional analysis of the genes showed that various functional gene sets were differentially expressed from season to season, which provided us with a global characterization of seasonally regulated biological processes in woody plants. Furthermore, we analyzed seasonal behaviors of the genes in context of the responsiveness to environmental factors. These approaches showed that the cold-associated genes are under both developmental and environmental control, and thus provided new insights into the interrelation between cold acclimation and dormancy development.

Materials and Methods

Plant Materials and Growth Conditions.

Populus deltoides (clone ill-129) was used throughout this work. First, stem cuttings were planted and grown in a green house under long day (LDW, 18 h) or short day (SDW, natural photoperiod) conditions for 13 weeks (Aug. 15 – Nov. 22, 2002). The temperature was kept at 23°C and the plants were watered three times a week. Then half of the plants from each photoperiod group were transferred to lower temperature (5 °C, LDC and SDC) conditions for three weeks. For a drought treatment (LDD), we took some of the plants that had been grown under LD and discontinued watering them for 40 days. Second, for the plant samples grown under the natural environmental conditions, stem cuttings were planted in field soil at the Tree Research Center on the campus of Michigan State University (East Lansing, MI). Those plants were harvested in the indicated months

(i.e., July 2002 to June 2003). Bark and xylem tissue samples were collected separately from the subject plants and stored at – 80 °C until used in the microarray experiments.

Experimental Design and Data Analysis.

The first set of experiments included a 12-month comparison between the bark and xylem tissues. The second set of experiments compared the bark and xylem tissues of plants grown under controlled environmental conditions, i.e., LDW, LDD, LDC, SDW, and SDC (Figure 2-1). To avoid variability due to sample preparation, the bark and xylem tissue samples used in all of the experiments were pooled from at least three independent plants. To eliminate potential variability due to aRNA amplification, we duplicated the amplification process. Each one of the duplicate amplifications was labeled with Cy5 or Cy3 using a dye swap strategy. Therefore, four data sets were derived for each experimental unit. A woven loop design was used for both sets of experiments. In the first set of experiments, RNA from each treatment plant group was competitively hybridized to the RNA in each of the other plant treatment groups. In the second set of experiments, RNA from the plants at N-th month was competitively hybridized to RNA from the plants at n-1^{th}, n+1^{th}, n+4^{th} and n+8^{th} month (Figure 2-1). For statistical analysis, systemic errors were corrected using lowess method after a log2-transformation of raw data. The normalized signal values were analyzed with ANOVA to evaluate the sources of variability. This method was previously described in Yang et al (2002) and implemented in MAANOVA package (http://www.jax.org/staff/churchill/labsite/-software/anova/rmaanova) in the statistical language R. The MAANOVA package provides three statistics tests (called F1, F2 and F3). The F1 is based on the error variance

of a single gene, the F3 is based on common error variance across all genes, and the F2 is based on a hybrid of the other two tests (Chui and Churchill, 2003). Probes for which there were no significant treatment effects ($p > 0.001$, > 0.05, and > 0.05 for F1, F2 and F3 for bark and $p > 0.0001$, > 0.005, and > 0.005 for xylem) were removed from the dataset. Finally, the signal values were further normalized by subtracting the variances caused by factors other than the treatment effects, thus generating a population with the relative expression values of each gene across the treatments. The relative expression values are treatment-based normalized data in the sense that effects due to the array and dye have been removed (Cui and Churchill 2003). Hierarchical and k-means clustering were performed on the normalized data by using CLUSTER and dendograms and k-means groups were visualized with TREEVIEW (Eisen et al., 1998).

Preparation of the cDNA Arrays

The six cDNA libraries were created using RNA prepared from either bark or xylem tissues harvested on Sep-06-01, Oct-30-01 and Apr-09-02. Total RNAs were extracted according to the modified Hu, *et al* (2002), and poly(A)+ enriched RNAs were prepared using the Poly(A)Purist™ MAG (Ambion, Austin, TX). cDNAs were synthesized using a SMART cDNA Library Construction Kit (Clontech, Palo Alto, CA), directionally ligated to pBluescript which was modified to contain *sfi* I site, and transformed to ElectroMAX™ DH12S™ competent cells (Invitrogen, Carlsbad, CA) by electroporation. The six cDNA libraries were subtracted by the randomly selected 1,920 sequences from the September bark and xylem cDNA libraries following Soares et al. (1994) with a few of modifications. The driver RNAs were prepared using Megascript T3 system (Ambion)

and BrightStar Psoralen-Biotin Nonisotopic Labeling Kits (Ambion). Target phagemid ssDNAs were prepared using M13K07 as a helper phage. The subtracted ssDNAs by the driver antisense RNAs were repaired to dsDNA using Klenow DNA polymerase and transformed to ElectroMAX™ DH12S™ to make the subtracted cDNA libraries. For the cDNA array, the 960 clones were randomly selected from each of the six subtracted libraries, which made a total of 7,680 clones including 1,920 of the driver cDNA clones. The cDNA inserts were amplified with T7 and T3 primer directly from the selected colonies. The 7,680 cDNA inserts were spotted on SuperAmine glass slides (Telechem, Sunnydale, CA) in a format containing 32 blocks of 240 inserts using an Omnigridder robot (Gene Machines, San Carlos, CA). After spotting, the slides were UV light cross-linked, washed and denatured as described by the manufacturer. Hybridized microarrays were scanned using an Affymetrix 428 array scanner (Affymetrix, Santa Clara, CA) and analyzed with GenePix Pro 4.0 software (Axon Instruments, Union City, CA). The 4,965 probes that showed significant signals when compared to the background were selected, sequenced and then clustered by StackPack program (Miller et al., 1999) resulting in 1,953 contigs and singletons that were supposed to represent unique genes.

RNA amplification, Probe Preparation and Hybridization.

For hybridization, total RNAs were prepared according to the modified Hu, et al (2002). The amplification of total RNA was performed based on a previously described method (Wang, et al., 2000). The amplified aRNA was cleaned up using an RNeasy mini kit (Qiagen, Valencia, CA). Two micrograms of aRNA were labeled with aminoallyl-dUTP (Sigma, St. Louis, MO) during reverse transcription using random hexamers and

Superscript™ II (Invitrogen). The labeled probes were purified using a QIAquick PCR purification kit (Qiagen). After purification, the probes were coupled with Cy3 or Cy5 monoreactive dye (Amersham, Piscataway, NJ). The labeled cDNA probes were then hybridized to cDNA arrays at 42°C for 12-18 hours in 40μl of hybridization buffer (50 % formamide, 5X SSC, 0.1 % SDS and 5 μg yeast tRNA). Following hybridization, the arrays were washed with 1X SSC, 0.1% SDS at 42°C for 5 min, 0.1X SSC, 0.1 % SDS at room temperature for 5 min and deionized water for 5 min. Hybridized microarrays were scanned using an Affymetrix 428 array scanner (Affymetrix) and analyzed with GenePix Pro 4.0 software (Axon Instruments)

Northern Blotting

Seven micrograms of each RNA sample were subjected to electrophoresis through 1.1% formaldehyde-agarose gels and transferred to Hybond-N nylon membranes (Amersham). Hybridization was performed in UltraHyb™ (Ambion) following the manufacturer's protocol. Probes were ^{32}P-dATP-labeled with Strip-EZ DNA kit (Ambion) and Northern blots were analyzed via QUANTITY ONE 4.1.1 software (Bio-Rad, Hercules, CA).

Results

Relationship between annual growth capacity and local environmental change.

To provide an assessment of relationship between the annual growth capacity of poplars and local seasonal changes, we measured the length and diameter growth of poplars in monthly intervals throughout the experiment year, from July 2002 to June 2003, and correlated the results to local seasonal changes, i.e., day length and temperature (Figure

2-2). Length and diameter growth were highly correlated indicating that shoot apical and cambial growth are correlated. Measurable diameter growth was first detected between April and May, correlating with bud break in mid-April, and reached its peak between July and August. The growth stopped around mid-September with the setting of terminal bud. As expected, plants entered dormancy in mid-September under relatively warm temperatures, around an average of 18 °C, and about 12 hr of daylength, implying that poplar plants perceive SD to be less than 12 hr. Re-growth of poplars in the spring was likely to be correlated with rising temperature rather than photoperiod because the plants did not resume their growth despite the increase in daylength, to longer than 13 h, that occurred in April. Based on these observations and previous studies (Chen et al., 2002; Howe et al., 1995) we defined the growing season as the period from May to August, September being the onset of dormancy, and the period from October to the following April as the non-growing season during which the cambium could transition from rest to quiescence along with the accumulation of chilling. This definition was further used to interpret results from the microarray experiment.

Major alterations in gene expression during winter reflect the different stages of cambial dormancy.

Seasonal growth of temperate woody plants is regulated by poorly understood interactions between endogenous and external environmental factors, which are orchestrated by multi-genetic factors (Chen et al., 2002). To understand the genetic regulation of seasonal growth, we investigated the global transcript profiles for an annual cycle using cDNA microarrays. The microarrays carrying 1,953 tentative unigenes were

used to determine the levels of mRNAs in bark and xylem tissues from field-grown poplar plants in monthly intervals as shown in Figure 2-2. During the annual growth cycle, 714 of the genes in bark and 743 from xylem were identified to show significant differential expressions, which had been determined by ANOVA as described in the method section. These selected genes were further clustered hierarchically, based on the similarity of their monthly expressions. The genes in bark were grouped largely into three clusters (Figure 2-3A) while the genes in xylem, grouped into two clusters (Figure 2-4A). The distinctive expression profiles in bark within the non-growing season were consistent with previous studies showing the existence of the two stages of cambial dormancy, i.e., rest and quiescence (Little and Bonga, 1974). Thus, we speculated that cluster A and cluster B might represent rest and quiescence, respectively. This speculation was further supported by the functional correlation of the genes in the clusters to the corresponding dormancy states as discussed later. Furthermore, the fact that the temperatures in Jan were similar to those in Feb, when cluster A/B genes were up- or down-regulated, suggests that the genes might be regulated not only by external environmental factors but also by internal factors, e.g., dormancy development.

Up-regulation of the genes associated with cold hardiness during early winter.

In bark, a major fraction of the genes (443) were identified as up-regulated in the growing season (Figure 2-3, Cluster C), representing 22 % of the total genes on the array. That group was followed by cluster A, identified as up-regulated in rest and representing 7 % (138), as well as cluster B, identified as up-regulated in quiescence and representing 6.8 % (133). On the other hand, in xylem, 34 % (669) of genes in the growing season, and

9.5 % (186) of the genes in non-growing season were identified as being up-regulated. The major biological activities represented by each cluster may be indicated by the over- or under-representation of categories within the clusters. For functional categorization, the genes of each cluster were sorted into ten functional groups according to MIPS (http://mips.gsf.de) categorization (Figure 2-3B, 3-4B). The genes that could be assigned to any group due to unknown function or had no significant matches in the current database were assigned to Unclassified or No-Hit groups, respectively. In cluster A of bark and xylem, the cell defense/rescue-associated genes and No-Hits were over-represented, and signal transduction and protein synthesis-associated genes were under-represented, indicating activation of cell protection mechanisms with reduced cellular communication and activity. The cell defense/rescue category codes for detoxification-, cold/drought- and pathogenesis-associated proteins (Figure 2-5). Many of them were reportedly induced during cold acclimation in arabidopsis and winter annuals, suggesting their roles in protection of cells from cold damages. For example, the antioxidant proteins may be involved in the detoxification of reactive oxygens probably caused by excess excitation of photosystem and imbalanced respiratory metabolism (Fowler and Thomashow 2002). Additionally, pathogenesis-associated proteins have been known to have antifreeze activity (Hon et al., 1995). However, unlike transiently up-regulation of the antioxidant genes in arabidopsis, the genes in poplar were induced throughout winter, suggesting that they could have more functions in woody plants. Indeed, the antioxidants have been speculated to be associated with breaking dormant bud based on a correlation of increased capacity to scavenge free radicals with chilling accumulation (Wang et al 1991; Wang and Faust, 1994). Down-regulation of these genes under freezing

temperatures in Feb supported this notion for cambial dormancy as well. One intriguing result is the up-regulation of the putative histone deacetylase gene, suggesting possible involvement of epigenetic regulation in transition from rest to quiesecence. During winter, temperate woody species have been known to considerably modify their metabolic processes, including the synthesis of proteins and sugars and membrane lipids. For example, the proportion of phospholipids and unsaturated fatty acids increases, while free sterol decrease (Wang and Faust 1988), and the pentose phosphate (PP) pathway is elevated (Sagisaka, 1974). The biochemical alterations of plasma membranes enable plants to maintain permeability of plasma membrane by increasing fluidity of membranes at low temperatures and the PP pathway could provide the reducing power needed for desaturation (Thomashow, 1999). Consistent with previous studies, our results showed an up-regulation of genes encoding enzymes associated with PP pathway, lipid desaturation, and cell wall modification during rest (Figure 2-5). Additionally, transcripts encoding galactinol synthases, the enzyme that catalyzes the first committed step in the synthesis of raffinose, and invertase were up-regulated, indicating the accumulation of soluble sugars, which are related to cold hardiness (Wanner and Junttila, 1999).

Elevation of signaling-associated genes occurred during quiescent dormanc.

For bark cluster B upregulated during late winter and early winter, signal transduction-associated genes were over-represented while cellular organization- and protein synthesis-associated genes were under-represented, indicating the activation of cellular signaling with reduced cellular activity during quiescence (Figure 2-3B). These results are consistent with the previous observation in which cambial cells could restore their

capability of responsiveness to growth hormones without any visible growth after being exposed to some period of chilling (Little and Bonga, 1974). The reactivation of the cellular signaling was reflected by receptor kinases, intracellular signaling proteins and the genes implicated in cytokinin and auxin signaling. Cytokinin as well as auxin has been known to be involved in breaking bud dormancy (Cutting et al 1991; Hewett and Wareing, 1974), suggesting the potential role of this hormone in the cambial transition to quiescence (Figure 2-6). Additionally, up-regulation of the putative cyclin gene further supports restoration of growth-competency after fulfillment of chilling requirements. Compared to rest dormancy, biological activities during quiescence dormancy have been less studied. We found that there were considerable changes in metabolism during late winter (i.e., quiescence). The transcripts encoding fermentation, (e.g., pyruvate decarboxylases and alcohol dehydrogenase) were up-regulated, suggesting that plants might be under oxygen deficiency stress, probably caused by ice encasement formed during freeze-thaw cycles in late winter (Bertland et al., 2003). We also found that transcripts encoding fatty acid oxidizing enzymes, e.g., acyl-CoA dehydrogenase, fatty acid CoA ligase, and fatty acid β-oxidation multifunctional protein were induced. Interestingly, induction of fatty acid oxidation is consistent with previous histochemical observations showing a gradual decline of lipid bodies in bark according to proceeding toward reactivation of cambium after transition to quiescence (Farrar and Evert, 1997; Riding and Little, 1984; Arend and Formm, 2003). The hydrolysis of lipids might have an effect on the freeing of water, which increased gradually with the end of bud endodormancy (Erez et al., 1998) and could allow plants to preserve the cryoprotective sugars during late winter. In addition, lignin synthesis-, phospholipid- and sterol-

associated genes were found to be up-regulated during quiescence. Although we could not determine the roles of these genes in cold hardiness or quiescence, these findings may reflect that, with dormancy break, there are changes in cell wall and membrane composition allowing increased permeability of solutes and water to the cytoplasm. Overall, the functional analysis suggests that plants have different biochemical characteristics between rest and quiescence, and the difference is represented in transcription levels.

The dissection of different stages of cold acclimation and dormancy under controlled growth conditions.

To dissect the effects of temperature and photoperiod, plants were grown under five different environmental conditions as described in methods and used for microarray experiments. In our growth system, LDC-treated plants did not show any visible change of the shoot but reduced vessel size in xylem, indicating reduced activity of cambial cells (Figure 2-7). LDD- or SDW-treated plants induced terminal bud and showed compressed layers of cambial zone, indicating that they were in dormancy. However, cambial cells in SDW plants had thickened cell wall that is characteristic of the rest (Chaffey et al., 1998). Based on these observations and previous studies, we speculated that LDC would induce cold acclimation without dormancy; LDD, quiescent dormancy; SDW, both first stage of cold acclimation and rest dormancy; SDC, second stage of cold acclimation and partly quiescent dormancy. Thus, comparison between them could lead us to associate the genes with various aspects of seasonal development. The speculation was further supported by hierarchical clustering based on the controlled environment and monthly expression

profiles (Figure 2-8). In bark, SDW were the closest to September; SDC, the closest to December; LDD and LDC are largely grouped to growing and non-growing months, respectively. On the other hand, in xylem, SDC is close to Oct/Nov while SDW and LDD, LDC is largely grouped to growing and non-growing months. The difference between bark and xylem implies that they might be under different regulation mechanism.

The modulation of cold signal by SD-induced dormancy.

For the next level of analysis, we performed k-means clustering based on seasonal and controlled environmental expression profiles of the genes selected in Figure 2-8. These genes were placed into eight groups (Figure 2-9). The genes of group G1, G2, G3 and G4 were largely expressed during the growing season, the genes of G5, G6, G7 and G8 were expressed during the non-growing season. The differences in responsiveness to controlled treatments was of interest since they may suggest what treatment(s) was likely to be a main regulatory group factor(s) during the annual cycle. Furthermore, the variety of responsiveness suggests that multiple pathways might be involved in regulation of the genes. For example, all of G1 – 4 groups have responded negatively to SDC while they have responded differently to LDD or SDW. G1 was up-regulated during the active growing season and responded negatively to four treatments, suggesting that they were involved mostly in active growth. Consistent with this view, G1 includes active growth-related genes, i.e., photosynthesis-, cell division-, and auxin signaling-associated genes (Table 3-1). G2 showed the positive responsiveness only to LDD, indicating that their genes might be associated with drought stress. G3 responds positively to LDD and SDW

and G4 responds positively only in SDW, indicating the association with onset of dormancy. The positive response to SDW appeared to be correlated with a relatively high level of expression in September. Interestingly, G3 included starch synthesis- and amino acid- as well as dehydration- related genes (Table 3-1), which have been reported to occur before onset of dormancy (Perry, 1971). Moreover, the fact that they could respond to both SD and drought might provide a molecular insight on how drought stress could hasten dormancy (Rinne et al., 1994). On the other hand, genes in G4 were up-regulated solely by SDW rather than LDC and had a higher expression in September when compared to the genes in G3. This result suggested that G4 genes might be involved in the onset of dormancy rather than cold acclimation. G4 includes many cell wall-associated genes, suggesting that cell wall thickness induced during dormancy development could be regulated primarily by SD signal. Finally, the negative responsiveness to LDC and SDC may be responsible for down-regulation of G1-4 gene groups during winter.

The groups up-regulated during late fall/winter, G5 – 8 (Figure 2-9), are of the most immediate interest for an integrative view of environmental and developmental factors in dormancy/cold hardiness. It has been known that SD and cold could induce cold acclimation independently (Welling et al., 2002). Consistent with the fact, many of genes (G5 -7) were up-regulated by SD as well as cold, suggesting that these genes should be under independent regulation of cold or SD. Woody perennials have been well known to require both cold and SD to achieve a higher capacity of cold hardiness. Consistently, G6 and 7 showed much a higher expression in SDC than LDC or SDW alone. Furthermore, the different response to cold between, before and after dormancy

induction suggested that dormancy development might modulate cold signal. Interestingly, a few of the genes up-regulated in early spring (i.e., bark cluster B) were grouped into G8 that showed up-regulation in LDD and LDC. The result was consistent with the fact that plants were imposed only by an adverse environment during quiescence. Moreover, the fact that most of cluster C genes were excluded from this analysis due to a non-significant difference between treatments suggested that the cluster C genes might be regulated by the unknown internal factors (e.g., hormones) rather than external environment (e.g., day length and temperature)

Validation of microarray results

Nothern blots were performed to confirm microarray results. Some examples are shown in Figure 2-10, which represents genes from G6 or G7 (Figure 2-9). The selected gene results followed the overall patterns obtained with microarrays.

Discussion

Annual growth cycle in woody perennials is a complex developmental process that is controlled by interactions between environmental factors and internal factors (Arora et al., 2003; Dennis, 1994; Lang, 1987). To understand the molecular basis of seasonal growth processes, we designed microarray experiments by which we could associate gene expressions with various environmental factors or growth stages. Using this approach, it was possible to show that seasonal growth was regulated through interaction between environmental and developmental signals at the molecular level.

Cell wall modification is associated with dormancy while the accumulation of energy reserves is associated with both growth cessation and dormancy.

Growth cessation has been known to be a prerequisite of cold acclimation since actively growing woody plants could not acclimate to cold. Moreover, treatments that stop growth (e.g., nutrient deficiency, water stress and chemicals) could increase hardiness to a limited extent (Weiser, 1970; Perry, 1971). However, the effect of growth cessation on cold hardiness is not clearly understood in woody plants. Since SD and drought could induce growth cessation, G3 responding positively to both drought and SD could provide insight on how the growth cessation promotes cold hardiness. Interestingly, G3 includes starch synthesis-associated genes, i.e., granule-bound starch synthase and starch branching enzymes (Table 3-1). The starch accumulation is necessary for cold hardiness because woody plants severely depleted in photosynthetic reserves cannot acclimate (Weiser, 1970). The coordinated up-regulation of plastidial forms of aldolase and Glyceraldehyde 3 phosphate dehdrogenase and cytosolic forms of triosephosphate isomerase implies that bark tissue may obtain carbon sources by its photosynthesis as

well as importing them from source organs, e.g., leaves. Thus, growth cessation by water stress appears to trigger accumulation of photosynthates that could contribute to promoting cold hardiness. In addition, the amino acid synthesis and ribosomal genes were found in G3, which might be related to nitrogen and sulfur mobilization. The coordinated expression of starch and amino acid synthesis associated genes suggests that growth cessation triggered by the prolonged exposure to adverse environments could activate the pathway to accumulate the energy reserves, e.g., starch and storage proteins, which could be used in return of a favorable environment. On the other hand, based on microscopic observations (Figure 2-7), SDW-treated plants were in rest dormancy while LDD-treated plants were in growth cessation (i.e., quiescence). Thus, we speculated that G4 responding solely to SD might be associated with rest dormancy. G4 included many cell wall and lignin/flavonoid associated genes. During cambial dormancy, it has been well known that extensive cell wall thickening and modification occur (Ermel et al., 2000). Thus, the cell wall related genes in G4 might reflect the difference between dormancy and growth cessation. Assuming that the structural changes are associated with rest dormancy while the metabolic changes are associated with growth cessation/quiescent dormancy, it is interesting to hypothesize that the biochemical differences could account in part for the different reversibility between them to active growth state in return of favorable environmental conditions.

The transition from rest to quiescence

The transition from rest to quiescence in the cambium is gradual and is accomplished by chilling (Little and Bonga 1974). However, the mechanism of chilling induced-transition

to quiescence is little understood. One of the activities associated with satisfaction of chilling requirement is a restoration of cell-to-cell signaling networks among individual cells of the meristem that had been disrupted during dormancy induction. Resumption of cell-to-cell communication, regulated by plasmodesmata (PD), may allow for symplastic movement of small signaling molecules, hormones, or proteins responsible for dormancy release. Rinne *et al.* (2001) found that the restoration of the symplastic organization of birch apical meristem was correlated with removing of 1,3-β-glucan, which is digested by 1,3-β-glucanase. This may be the case on cambial dormancy. Putative 1,3-β-glucanase highly increased by chilling and decreased in time supposed to be in quiescence (Figure 2-4A). Moreover, Schrader *et al.* (2004) reported up-regulation of 1,3-β-glucan synthase in poplar dormant cambium, which can produces 1,3-β-glucan to block PD. In addition, we found up-regulation of pectin methylesterase and pectin glucuronyltransferase during early winter (Figure 2-4A). Recently, pectin glucuronyltransferase has been reported to be essential for intercellular attachment of meristemic cells, of which the disruption could disconnect cell signaling via the plasmodesmata (Iwai et al., 2002). The up-regulation of the genes during chilling led us to assume that these enzymes could be associated with symplastic restoration of cambium. An intriguing finding was the upregulation of a popoar homolog of histone deacetylase during rest (Figure 2-4A). Many characteristics are shared between vernalization and chilling requirement, e.g., the range of effective temperature, the experienced period and the site of perception, and the role of histone deacetylation on transition from vegetative to reproductive stage by prolonged exposure to cold (i.e., vernalization) has been well known in arabidopsis (He et al., 2003). Thus, the up-regulation of the gene provides the possibility that chilling requirement could be

regulated by transcriptional regulation through the modification of chromosomal structure.

Woody plants to maintain cold hardiness differently depending on growth stages. Woody plants are hardly dehardened during rest, whereas quiescent plants are rapidly dehardened in response to rising temperature (Perry, 1971, Anisko et al., 1994). However, there is little data on the nature of enzymes involved in this process. Our data suggested that the difference on dehardening rate could be accounted by reactivation of signaling-related genes (e.g., receptor kinases, calmodulin) during late winter. In addition, the coordinated up-regulation of phospholipid (i.e., diacylglycerol kinase and phosphoesterase) and sterol modification associated genes (i.e., cholesterol acyltransferase, sterol reductase and sterol sulfotransferase) could imply that the molecules might play a role as intra/intercellular signaling messengers. Recently, the significant role of sterols on vascular patterning and differentiation were reported (Carland et al., 2002; Schrick et al., 2004). Since dehardening process is correlated with re-activation of cambium, thus, it is plausible that sterols could act on reactivation of the cambium.

The modulation of cold signal by SD

It has been well known that both low temperature and SD signal are required for maximum cold hardiness. However, how SD signal could modulate the response to cold at the molecular level is less understood. Our expression profiles may provide a molecular glimpse of these phenomena. According to our results, there were at least two mechanisms. First, SD enhanced the expression of the cold-associated genes as well as

SD itself induced the genes independently of cold. As shown at G6 of k-means clustering (Figure 2-9), cold associated genes had much a higher expression in SDC than LDC. Chen and Li (1978) proposed that LT, SD and water stress trigger independent frost hardening mechanisms, and total hardiness would be a sum of these. It might be reflected by G6. SD and cold signal could induce the genes independently, resulting in the higher expression of the genes than cold signal alone. Second, SD signal itself could not induce the cold-associated genes but it enhanced the genes much higher than cold signal alone when plants had cold signal, which was supported by G7 (Figure 2-9). Thus, the responsiveness to cold signal appeared to depend on how the dormancy development proceeded. Moreover, the fact that both groups of genes were up-regulated in first stage of cold acclimation, (i.e., LDC) suggested that the higher level of cold hardiness (i.e. SDC) could be achieved by a higher expression of the same battery of genes as with the first stage of cold acclimation. In annual herbaceous plants, the expression level of cold-associated genes is regulated primarily by responding to the degree of temperature, i.e., non-freezing temperature or freezing temperature (Zhu et al., 2000) while in woody plants, they are induced differentially depending on different development stages as shown here. This fact suggests that woody plants may evolve different mechanisms to attain extreme freezing tolerance. The ultimate survival of woody plants is dependent on not only the maximal capacity of cold hardening, but also on the timing and rate of cold acclimation and the stability of cold hardiness against unseasonably warm periods during winter (Weiser, 1970). Hence, the successful performance of a woody species in a particular locality implies synchronization of the annual development of cold hardiness with the seasonal temperature changes. By controlling cold hardiness via developmental

stage, which is regulated by predictable photoperiod, plants are able to distinguish between unseasonable and seasonable cold and keep cold hardiness in accordance with local seasonality.

References

Anisko T, Lindstrom OM, Hoogenboom G (1994) Development of a Cold-Hardiness Model for Deciduous Woody-Plants. Physiologia Plantarum. 91: 375-382

Arora R, Rowland LJ, Tanino K (2003) Induction and release of bud dormancy in woody perennials: A science comes of age. Hortscience. 38: 911-921

Arora R, Wisniewski ME (1994) Cold-Acclimation in Genetically Related (Sibling) Deciduous and Evergreen Peach (Prunus-Persica L Batsch) .2. A 60-Kilodalton Bark Protein in Cold-Acclimated Tissues of Peach Is Heat-Stable and Related to the Dehydrin Family of Proteins. Plant Physiology. 105: 95-101

Bertrand A, Castonguay Y, Nadeau P, Laberge S, Michaud R, Belanger G, Rochette P (2003) Oxygen deficiency affects carbohydrate reserves in overwintering forage crops. J Exp Bot. 54: 1721-30

Carland FM, Fujioka S, Takatsuto S, Yoshida S, Nelson T (2002) The identification of CVP1 reveals a role for sterols in vascular patterning. Plant Cell. 14: 2045-58

Chaffey NJ, Barlow PW, Barnett JR (1998) A seasonal cycle of cell wall structure is accompanied by a cyclical rearrangement of cortical microtubules in fusiform cambial cells within taproots of Aesculus hippocastanum (Hippocastanaceae). New Phytologist. 139: 623-635

Chen HH, Li PH (1978) Interactions of Low-Temperature, Water Stress, and Short Days in Induction of Stem Frost Hardiness in Red Osier Dogwood. Plant Physiology. 62: 833-835

Chen THH, Howe GT, Bradshaw HD (2002) Molecular genetic analysis of dormancy-related traits in poplars. Weed Science. 50: 232-240

Christersson L (1978) Influence of Photoperiod and Temperature on Development of Frost Hardiness in Seedlings of Pinus-Silvestris and Picea-Abies. Physiologia Plantarum. 44: 288-294

Cui X, Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. Genome Biol. 4: 210

Cutting JGM, Strydom DK, Jacobs G, Bellstedt DU, Vandermerwe KJ, Weiler EW (1991) Changes in Xylem Constituents in Response to Rest-Breaking Agents Applied to Apple before Budbreak. Journal of the American Society for Horticultural Science. 116: 680-683

Dennis FG (1994) Dormancy - What We Know (and Dont Know). Hortscience. 29: 1249-1255

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 95: 14863-8

Erez A, Couvillon GA, Kays SJ (1980) Effect of Oxygen Concentration on the Release of Peach Leaf Buds from Rest. Hortscience. 15: 39-41

Erez A, Faust M, Line MJ (1998) Changes in water status in peach buds on induction, development and release from dormancy. Scientia Horticulturae. 73: 111-123

Ermel FF, Follet-Gueye ML, Cibert C, Vian B, Morvan C, Catesson AM, Goldberg R (2000) Differential localization of arabinan and galactan side chains of rhamnogalacturonan 1 in cambial derivatives. Planta. 210: 732-740

Espinosa-Ruiz A, Saxena S, Schmidt J, Mellerowicz E, Miskolczi P, Bako L, Bhalerao RP (2004) Differential stage-specific regulation of cyclin-dependent kinases during cambial dormancy in hybrid aspen. Plant J. 38: 603-615

Farrar JJ, Evert RF (1997) Seasonal changes in the ultrastructure of the vascular cambium of Robinia pseudoacacia. Trees-Structure and Function. 11: 191-202

Fowler S, Thomashow MF (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell. 14: 1675-90

Greer DH, Stanley CJ, Warrington IJ (1989) Photoperiod Control of the Initial Phase of Frost Hardiness Development in Pinus-Radiata. Plant Cell and Environment. 12: 661-668

He Y, Michaels SD, Amasino RM (2003) Regulation of flowering time by histone acetylation in Arabidopsis. Science. 302: 1751-4

Hewett EW, Wareing PF (1973) Cytokinins in Populus by Robusta - Changes During Chilling and Bud Burst. Physiologia Plantarum. 28: 393-399

Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DS (1995) Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. Plant Physiol. 109: 879-89

Howe GT, Hackett WP, Furnier GR, Klevorn RE (1995) Photoperiodic Responses of a Northern and Southern Ecotype of Black Cottonwood. Physiologia Plantarum. 93: 695-708

Hu C, Honda C, Kita M, Zhang Z, Tsuda T, Moriguchi T (2002) A Simple Protocol for RNA isolation from Fruit Trees Containing High Levels of Polysaccharides and Polyphenol Compounds. Plant Molecular Biology Reporter. 20: 69a-69g

Iwai H, Masaoka N, Ishii T, Satoh S (2002) A pectin glucuronyltransferase gene is essential for intercellular attachment in the plant meristem. Proc Natl Acad Sci U S A. 99: 16319-24

Jiang M, Zhang J (2002) Water stress-induced abscisic acid accumulation triggers the increased generation of reactive oxygen species and up-regulates the activities of antioxidant enzymes in maize leaves. J Exp Bot. 53: 2401-10

Kim HJ, Kim YK, Park JY, Kim J (2002) Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in Arabidopsis thaliana. Plant J. 29: 693-704

Lang GA, Early JD, Martin GC, Darnell RL (1987) Endodormancy, Paradormancy, and Ecodormancy - Physiological Terminology and Classification for Dormancy Research. Hortscience. 22: 371-377

Little CHA (1981) Effect of Cambial Dormancy State on the Transport of 1-C-14 Indol-3-Ylacetic Acid in Abies-Balsamea Shoots. Canadian Journal of Botany-Revue Canadienne De Botanique. 59: 342-348

Little CHA, Bonga JM (1974) Rest in Cambium of Abies-Balsamea. Canadian Journal of Botany-Revue Canadienne De Botanique. 52: 1723-1730

Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR, Hide WA (1999) A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Res. 9: 1143-55

Nissila PC, Fuchigami LH (1978) Relationship between Vegetative Maturity and 1st Stage of Cold-Acclimation. Journal of the American Society for Horticultural Science. 103: 710-711

Olsen JE, Junttila O, Nilsen J, Eriksson ME, Martinussen I, Olsson O, Sandberg G, Moritz T (1997) Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. Plant Journal. 12: 1339-1350

Perry TO (1971) Dormancy of Trees in Winter. Science. 171: 29-&

Riding RT, Little CHA (1984) Anatomy and Histochemistry of Abies-Balsamea Cambial Zone Cells During the Onset and Breaking of Dormancy. Canadian Journal of Botany-Revue Canadienne De Botanique. 62: 2570-2579

Rinne P, Saarelainen A, Junttila O (1994) Growth Cessation and Bud Dormancy in Relation to Aba Level in Seedlings and Coppice Shoots of Betula-Pubescens as Affected by a Short Photoperiod, Water-Stress and Chilling. Physiologia Plantarum. 90: 451-458

Rinne PLH, Kaikuranta PM, van der Schoot C (2001) The shoot apical meristem restores its symplasmic organization during chilling-induced release from dormancy. Plant Journal. 26: 249-264

Rowland LJ, Arora R (1997) Proteins related to endodormancy (rest) in woody perennials. Plant Science. 126: 119-144

Sagisaka S (1974) Transition of metabolisms in living poplar bark from growing to wintering stages and vice versa. Plant Physiol. 54: 544-549

Samish RM (1954) Dormancy in woody plants. Annual Review of Plant Physiology and Plant Molecular Biology. 5: 183-203

Schrader J, Baba K, May ST, Palme K, Bennett M, Bhalerao RP, Sandberg G (2003) Polar auxin transport in the wood-forming tissues of hybrid aspen is under simultaneous control of developmental and environmental signals. Proceedings of the National Academy of Sciences of the United States of America. 100: 10096-10101

Schrader J, Moyle R, Bhalerao R, Hertzberg M, Lundeberg J, Nilsson P, Bhalerao RP (2004) Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome. Plant Journal. In press

Schrick K, Fujioka S, Takatsuto S, Stierhof YD, Stransky H, Yoshida S, Jurgens G (2004) A link between sterol biosynthesis, the cell wall, and cellulose in Arabidopsis. Plant J. 38: 227-43

Thomashow MF (1999) Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology. 50: 571-599

Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM (2000) High-fidelity mRNA amplification for gene profiling. Nat Biotechnol. 18: 457-9

Wang SY, Faust M (1988) Changes of fatty acids and sterols in apple buds during bud break induced by a plant bioregulator, thidiazuron. Physiologia Plantarum. 72: 115-120

Wang SY, Faust M (1994) Changes in the Antioxidant System Associated with Budbreak in Anna Apple (Malus-Domestica Borkh) Buds. Journal of the American Society for Horticultural Science. 119: 735-741

Wang SY, Jiao HJ, Faust M (1991) Changes in Ascorbate, Glutathione, and Related Enzyme-Activities During Thidiazuron-Induced Bud Break of Apple. Physiologia Plantarum. 82: 231-236

Wanner LA, Junttila O (1999) Cold-induced freezing tolerance in Arabidopsis. Plant Physiol. 120: 391-400

Weiser CJ (1970) Cold resistance and injury in woody plants. Science. 169: 1269-78

Welling A, Moritz T, Palva ET, Junttila O (2002) Independent activation of cold acclimation by low temperature and short photoperiod in hybrid aspen. Plant Physiology. 129: 1633-1641

Welling A, Rinne P, Vihera-Aarnio A, Kontunen-Soppela S, Heino P, Palva ET (2004) Photoperiod and temperature differentially regulate the expression of two dehydrin genes during overwintering of birch (Betula pubescens Ehrh.). Journal of Experimental Botany. 55: 507-516

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30: e15

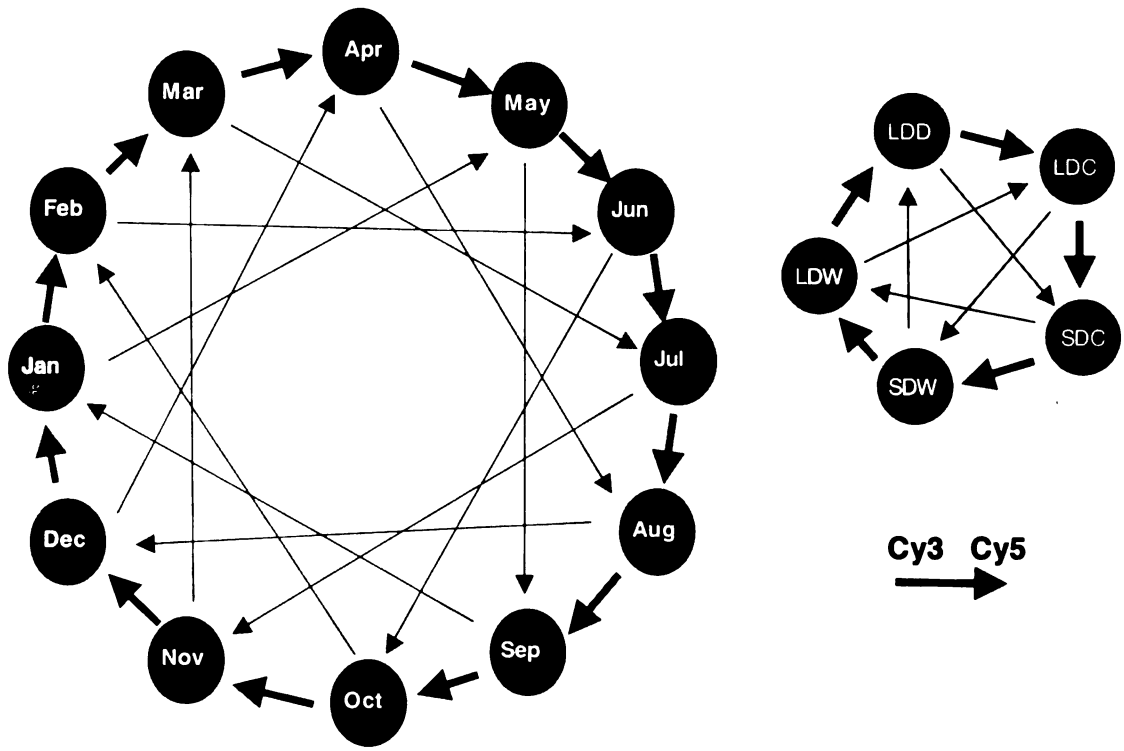


Figure 2-1. Microarray experimental design.
For a particular target sample, the pooled material was divided into two batches; total RNAs were amplified before labeling. These batches were split in two and labeled with Cy5 or Cy3 dye. The labeled probes were randomly assigned to each of the four hybridizations, respectively. Arrows represent hybridizations between the target samples. The sample at the tail of the arrow is labeled with Cy3, and the sample at the head of the arrow is labeled with Cy5. Four array experiments were performed based on each target sample. LDW represent long day (16hr) in warm temperature (25 °C) condition; LDD, long day in warm temperature (25 °C) with drought treatment; LDC, long day in cold temperature (5 °C); SDW, short day (~9hr) in warm temperature; SDC, short day in cold temperature.

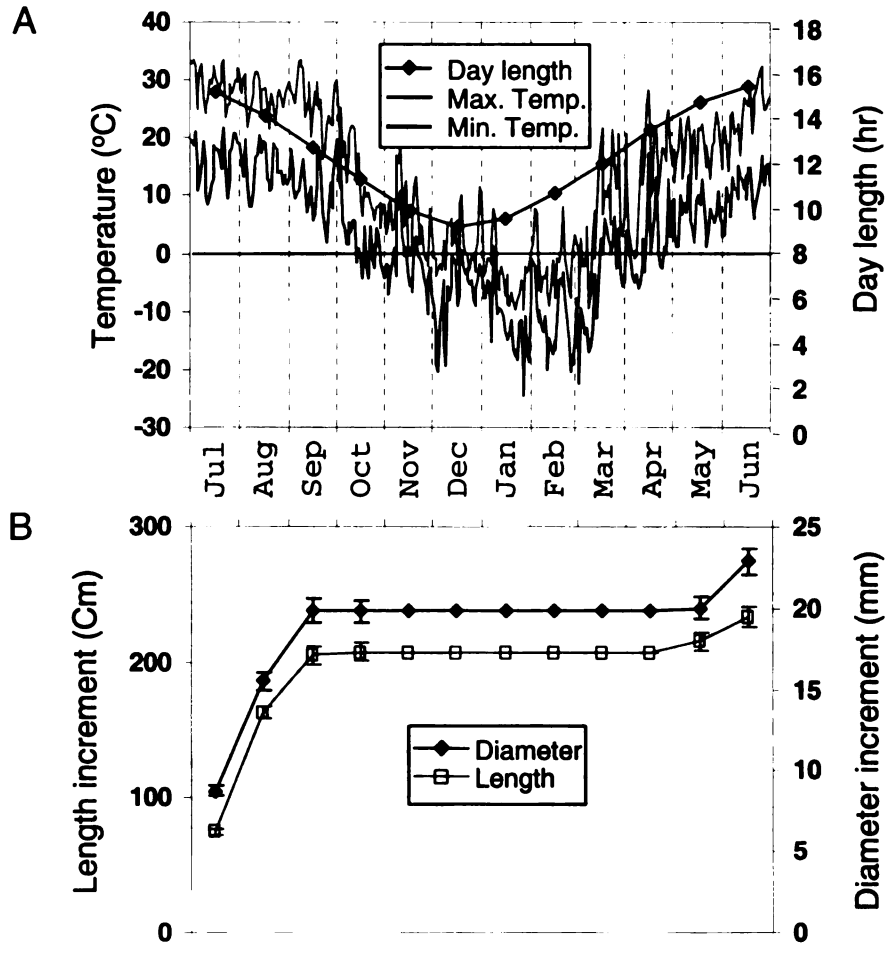


Figure 2-2. The temperature, daylength and tree growth change during the experiment period (i.e., July 2002 to June 2003).

(A) Daily minimum and maximum temperatures are illustrated throughout the experiment period. Photoperiod is represented by the length of daylight on 15th day of each month at experimental site in East Lansing, MI. (B) The length or volume growth of primary stems is an average of 20 independent measurements, and the tree materials used for both the measurements and the microarray experiments were from the same field. Error bar represent ±S.E.

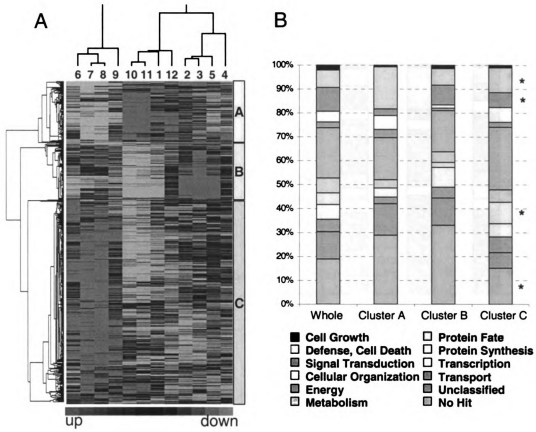


Figure 2-3. Hierarchical clustering (A) and Functional categorization (B) of genes based on the seasonal variation of bark.

The 714 genes from the bark samples were selected based on ANOVA. In hierarchical clustering, months are indicated numerically by columns, and genes are indicated by rows. In functional categorization, all 1,953 of the genes on the array were used for comparison reference, and referred to as Whole in the graph. The asterisks indicate the differed distribution of the functional groups between clusters, which was supported by a contingency test (p <0.001).

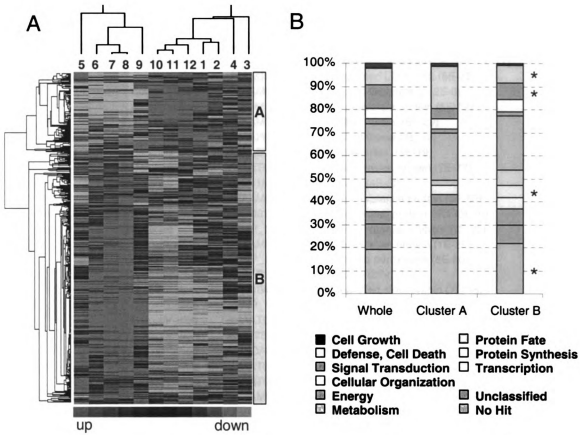


Figure 2-4. Hierarchical clustering (A) and Functional categorization (B) of genes based on seasonal variation of xylem.
The 743 genes were selected based on ANOVA in xylem. Hierarchical clustering and functional categorization were analyzed as described in Figure. 2.

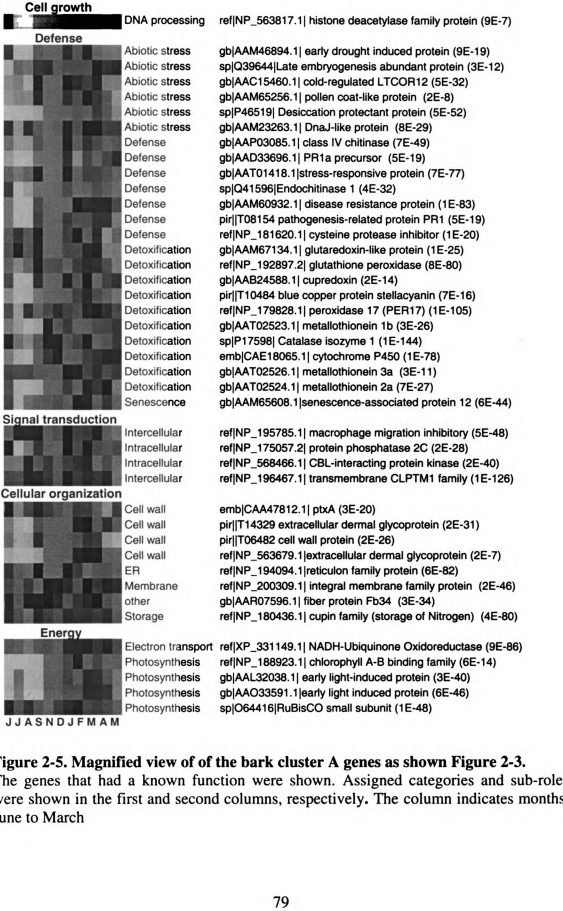
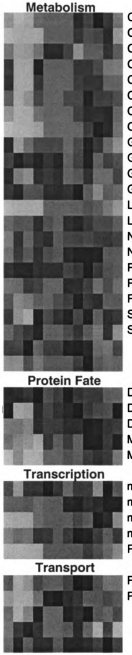


Figure 2-5. Magnified view of of the bark cluster A genes as shown Figure 2-3. The genes that had a known function were shown. Assigned categories and sub-roles were shown in the first and second columns, respectively. The column indicates months, June to March

Figure 2-5 (cont'd)



Metabolism		
	Carbohydrate	pir||T50664 methylenetetrahydrofolate dehydrogenase (1E-107)
	Carbohydrate	gb|AAM96867.1| galactinol synthase (1E-118)
	Carbohydrate	gb|AAM22411.1| cell-wall invertase (2E-79)
	Cell wall	emb|CAD29733.1| pectin methylesterase (4E-45)
	Cell wall	pir||T16976 pectinesterase (2E-99)
	Cell wall	dbj|BAD15422.1|pectin-glucuronyltransferase (6E-41)
	Cell wall	gb|AAP87281.1| beta-1,3-glucanase (1E-104)
	Cell wall	pir||T50680 beta-1,3 glucanase (1E-141)
	Glycolysis	dbj|BAA02729.1| cytoplasmic aldolase (4E-16)
	Glycolysis	gi|17380672| phosphoglycerate dehydrogenase (0E0)
	Glycolysis	ref|NP_850759.1| fructose-bisphosphate aldolase (1E-124)
	Glycolysis	gi|22296820| pyruvate kinase (1E-103)
	Lipid	ref|NP_200316.1| GDSL-motif lipase/hydrolase family (6E-74)
	Lipid	gb|AAN87573.1| delta 12 oleic acid desaturase (1E-131)
	NAD salvage	gb|AAN74808.1|Nicotinate phosphoribosyltransferase (2E-79)
	Nucleic acid	emb|CAD37200.1|guanine deaminase (7E-70)
	PPP	pir||T07790 transaldolase (1E-102)
	PPP	ref|NP_198892.1| glucose-6-P dehydrogenase (0E0)
	Rubber syn.	ref|NP_176904.1| rubber elongation factor family (4E-31)
	Sterol	ref|NP_595138.1| oxosterol binding domain (3E-41)
	Sterol	emb|CAD87012.1|progesterone 5-beta-reductase (1E-136)
		ref|NP_172904.2| oxygenase (1E-116)
		ref|NP_190468.1| AMP-dependent synthetase (8E-73)
Protein Fate		
	Degradation	ref|NP_850481.1| F-box family protein (1E-82)
	Degradation	pir||S42882 cysteine proteinase (9E-64)
	Degradation	ref|NP_564919.1| F-box family protein (2E-90)
	Modification	ref|NP_922744.1|palmitoyl-protein thioesterase (1E-86)
	Modification	ref|NP_194414.1| ubiquitin-like protein (2E-40)
Transcription		
	mRNA syn.	ref|NP_179571.1| zinc finger (2E-75)
	mRNA syn.	gb|AAT37999.1|RNA polymerase sigma factor (5E-97)
	mRNA syn.	ref|NP_194461.1| zinc finger (5E-13)
	mRNA syn.	ref|NP_180326.1| zinc finger (AN1-like) family protein (7E-43)
	RNA processing	emb|CAD48198.1|RNA-binding protein (4E-42)
Transport		
	Protein	ref|NP_849394.1| mitochondrial membrane translocase (3E-60)
	Proton	pir||T14564 inorganic H+ diphosphatase (2E-23)
		ref|NP_567178.1| plasma membrane intrinsic protein (1E-129)
		pir||JQ2288 nodulin-26 (3E-39)
		ref|NP_565876.2| SOUL heme-binding family protein (2E-31)

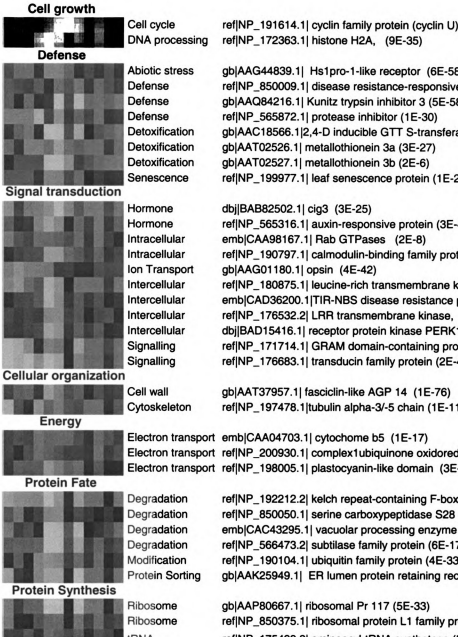
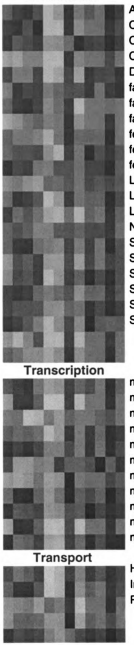


Figure 2-6. Magnified view of the genes of the bark cluster B as shown Figure 2-3.

Figure 2-6 (cont'd)



Metabolism	
Amino acid	ref|NP_197318.1| glutamate dehydrogenase 1 (6E-71)
Carbohydrate	ref|NP_180697.1| pfkB-type carbohydrate kinase family (7E-51)
Cell wall	ref|NP_180377.2|alpha-L-fucosidase, (4E-24)
Cell wall	ref|NP_195212.1| beta-fructofuranosidase (1E-102)
Detoxification	ref|NP_563973.1| lactoylglutathione lyase (glyoxalase I) (1E-12)
fatty acid	dbj|BAC83542.1| acyl-CoA dehydrogenase (6E-41)
fatty acid	pir||T07929long-chain-fatty-acid-CoA ligase (2E-86)
fatty acid	ref|NP_194630.1| fatty acid oxidation multifunctional (1E-23)
fermentation	ref|NP_014051.1| NADPH alcohol dehydrogenase (2E-38)
fermentation	sp|P93436|Alcohol dehydrogenase class III (1E-86)
fermentation	sp|P51850|Pyruvate decarboxylase isozyme 1 (1E-120)
Lipid	ref|NP_172203.1| phosphoesterase family protein (2E-16)
Lipid	ref|NP_200577.1| diacylglycerol kinase, (2E-16)
Lipid	gb|AAP23033.1| sphingolipid delta-8 desaturase (8E-63)
Nucleic acid	ref|NP_179967.3| 5' nucleotidase family protein (1E-34)
Secondary	sp|P93711|Caffeoyl-CoA O-methyltransferase (1E-124)
Secondary	gb|AAK58908.1| 4-coumarate:CoA ligase 3 (6E-38)
Secondary	emb|CAB65335.1| ferulate-5-hydroxylase (1E-114)
Sterol	gb|AAP47635.1| lecithine cholesterol acyltransferase (8E-84)
Sterol	gb|AAF81279.1| C-14 sterol reductase (8E-91)
Sterol	dbj|BAB11159.1| steroid sulfotransferase (2E-95)
	ref|NP_171840.1|oxoglutarate-dependent dioxygenase (1E-24)
	ref|NP_180811.1| dienelactone hydrolase family (6E-54)
Transcription	
mRNA syn.	ref|NP_564690.1| Kin17 DNA-binding protein (7E-57)
mRNA syn.	ref|NP_194573.1| DNA-binding protein (5E-40)
mRNA syn.	ref|NP_567546.1| DNA-binding family protein (1E-38)
mRNA syn.	dbj|BAD19748.1| helix-loop-helix DNA-binding protein (1E-7)
mRNA syn.	ref|NP_920230.1| transcription initiation factor (1E-5)
mRNA syn.	ref|NP_188169.1| no apical meristem (2E-76)
mRNA syn.	gb|AAM09320.2|Homeobox-containing protein (1E-20)
mRNA syn.	ref|NP_196202.1| zinc finger (C3HC4-type) family (3E-42)
mRNA syn.	ref|NP_200445.1| zinc finger (C3HC4-type) (4E-7)
mRNA syn.	pir||T01735 homeobox protein NTH15 (2E-79)
mRNA syn.	ref|NP_565774.1| DNA-binding protein (7E-47)
Transport	
Hormone	sp|O04011|Auxin-binding protein ABP20 (1E-47)
Intra-transport	ref|NP_195616.2| kinesin-related protein (5E-27)
Phosphate	gb|AAK38196.1| phosphate transporter 1 (1E-107)
	pir||T00435 mitochondrial carrier protein (6E-12)
	gb|AAQ91200.1| glutathione transporter (1E-131)

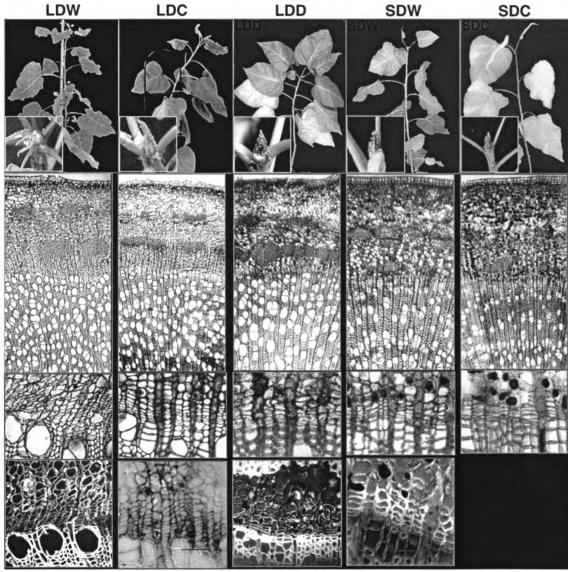


Figure 2-7. The whole plant picture and stem sections of the poplar trees treated under controlled environments.

The tree materials used were the same as for the microarray experiments. Colored pictures of stem sections (second and third rows) were taken on microscopy. Black and white pictures at the fourth row were taken on confocal microscopy. Scale bar represents 50 um. Long daylength and warm (LDW); Long daylength and drought (LDD); Long daylength and cold (LDC); Short daylength and warm (SDW); Short daylength and cold (SDC)

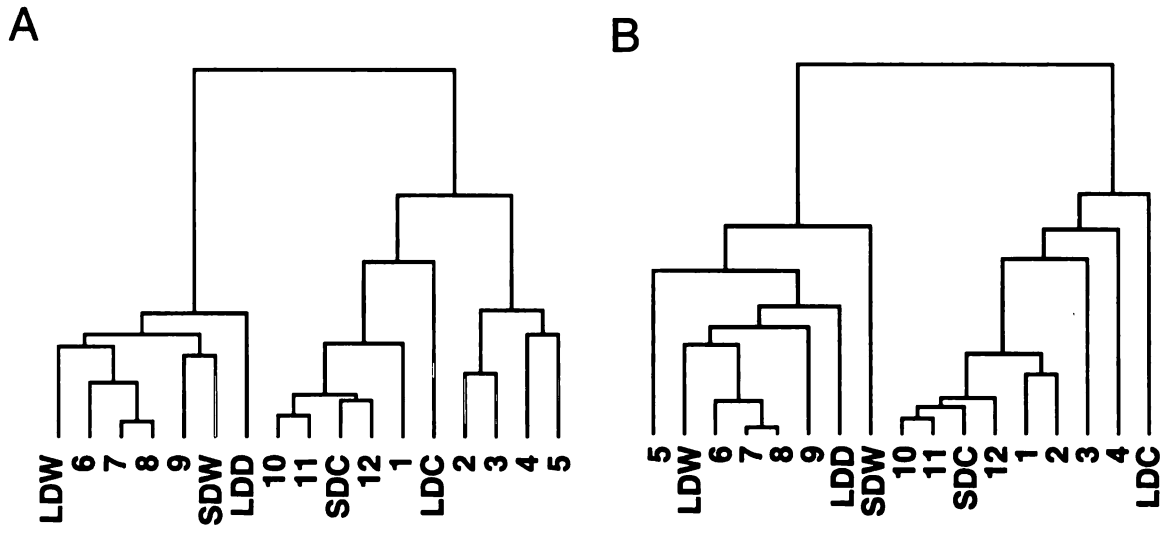


Figure 2-8. Dendrogram showing relationships between seasonal changes and controlled environmental treatments in bark (A) and xylem tissue (B).
Months were indicated numerically. For dendrogram, sets of genes were selected based on $P < 0.001$, 0.05, 0.05 for bark or $P < 0.0001$, 0.005, 0.005 (for xylem) of MAANOVA at both seasonal changes and controlled treatments.

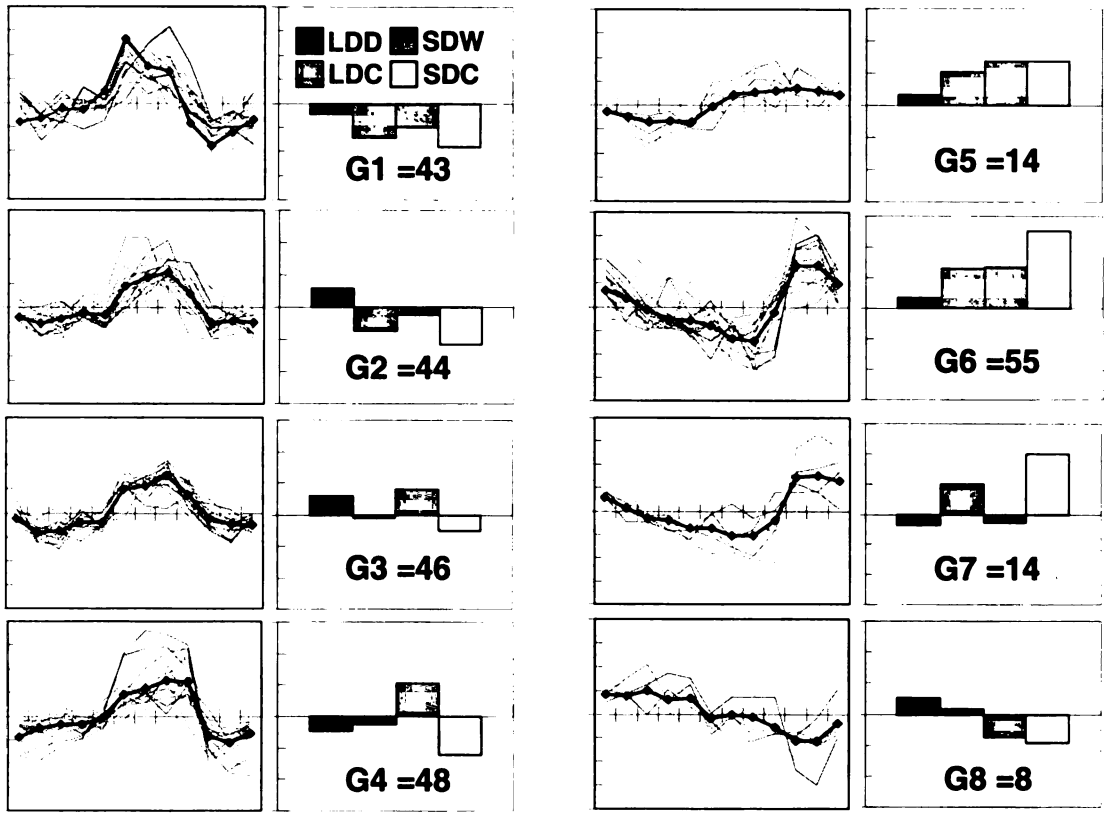


Figure 2-9. K-means clusters based on expression profiles of both annual cycle and controlled environments.

A total of 272 genes were individually clustered into eight groups. Each group is represented by a pattern of seasonal variation (line graphs, left) and environment responsiveness (bar graphs, right) for genes in the group. The number of genes in each group is indicated at the bottom of the bar graph. For the seasonal variation graph, the x axis represents twelve months (i. e., Jan to Dec) and the y axis represent the normalized value ranging between 4 and -4 in log 2 space as described in method section. The thick black line indicates the means value. In the environmental responsiveness graph, the means fold changes ranging between 3 and –3 for each of the four treatments (i.e., LDD, LDC, SDW, and SDC) over LDW were shown as the y axis in log 2 scale.

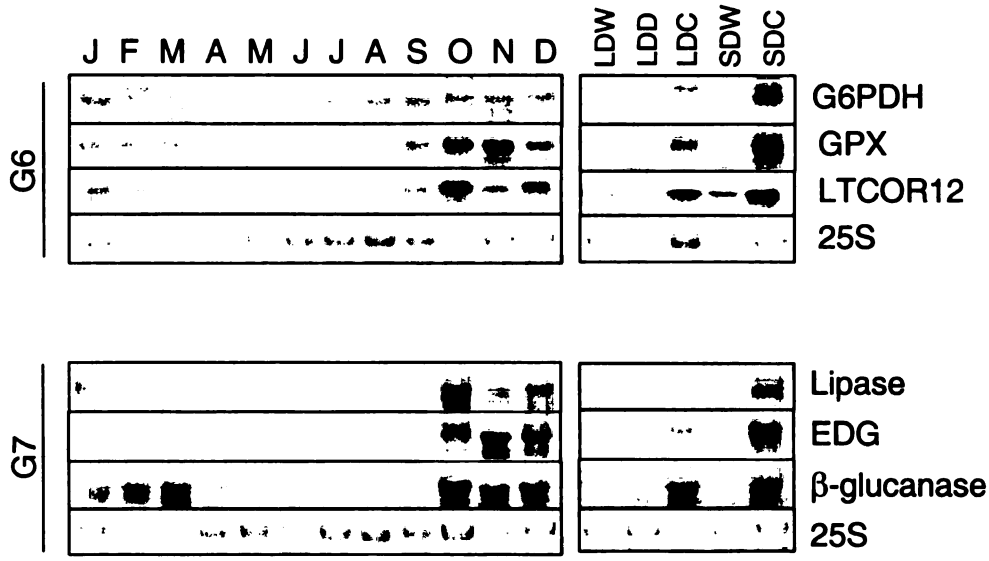


Figure 2-10. Validation of microarray results by Northern blot analysis.
RNAs were prepared from the bark tissues of trees grown under the indicated environment conditions or in out-field. 25S rRNA was used as a loading control. G6PDH, glucose 6-phosphate dehydrogenase; GPX, glutathione peroxidase; LTCOR15, Cold-regulated COR15; Lipase, GDSL-motif lipase; EDG, Extracellular dermal glycoprotein; β-glucanase, endo-1,3-β-glucanase.

Table 2-1. Eight groups of genes identified by K-means clustering as shown in Figure 2-9.

CloneID	Description (BlastX *E*-val)	Category	Sub-role
GROUP 1			
L2P10d10	dbj|BAD00018.1|histone 1 (6E-19)	CELL GROWTH	DNA processing
B9SP10b05	gb|AAK01359.1|dehydration stress-induced protein (8E-57)	CELL RESCUE, DEFENSE	Abiotic stress
B9P02g09	gb|AAS58469.1|ultraviolet-B-repressible protein (1E-28)	CELL RESCUE, DEFENSE	Abiotic stress
B9P07f10	gb|AAR09168.1|alpha-expansin 1 (1E-112)	CELLULAR ORGANIZATION	Cell wall
B9SP09a02	gb|AAF78903.1|proline-rich protein (2E-34)	CELLULAR ORGANIZATION	Cell wall
X9P08b06	emb|CAD58680.1|microtubule associated protein (8E-44)	CELLULAR ORGANIZATION	Cytoskeleton
L1P08f11	gb|AAP32191.1|alpha-tubulin (1E-126)	CELLULAR ORGANIZATION	Cytoskeleton
X9P09d02	emb|CAB66336.1|alpha-tubulin (1E-107)	CELLULAR ORGANIZATION	Cytoskeleton
B9P07d01	gb|AAP03873.1|photosystem I reaction center (3E-44)	ENERGY	Photosynthesis
B9SP08d05	ref|NP_176940.1|photosystem II core complex(5E-25)	ENERGY	Photosynthesis
X9SP07c12	ref|NP_192772.1|chlorophyll A-B binding protein (9E-78)	ENERGY	Photosynthesis
L1P10c10	pir||S04125 chlorophyll a/b-binding protein type III (1E-36)	ENERGY	Photosynthesis
X9P07e05	sp|P34105|NADP Malic enzyme (1E-158)	METABOLISM	Carbohydrate
B9SP09a05	ref|NP_198236.1|NAD-epimerase/dehydratase family(1E-103)	METABOLISM	Carbohydrate
X9SP06f01	gb|AAF80591.1|xyloglucan endotransglycosylase (1E-22)	METABOLISM	Cell wall
X9P10c09	ref|NP_973975.1|lipase class 3 family protein (3E-55)	METABOLISM	Lipid
L3P02b09	sp|O23787|Thiazole biosynthetic enzyme (1E-119)	METABOLISM	
X9SP04f03	gb|AAM51555.1|metacaspase 1(protease) (2E-50)	PROTEIN FATE	Degradation
B9P03e01	gb|AAB97006.1|Gibberellin regulated protein(4E-23)	SIGNAL TRANSDUCTION	Hormone
X9SP05e05	dbj|BAA33810.1|phi-1(phosphate induced) (1E-130)	SIGNAL TRANSDUCTION	Intracellular
X9P05f04	ref|NP_195130.1|zinc finger (C3HC4-type) (3E-56)	SIGNAL TRANSDUCTION	Intracellular
B9SP06h06	pir||T07086 acid phosphatase (8E-91)	SIGNAL TRANSDUCTION	Intracellular
B9P10a12	gb|AAM29182.1|Aux/IAA protein (2E-64)	TRANSCRIPTION	mRNA synthesis
L3P04b02	emb|CAC84706.1|aux/IAA protein (1E-127)	TRANSCRIPTION	mRNA synthesis
B9P03f11	ref|NP_178396.1|zinc finger (C3HC4-type)(9E-38)	TRANSCRIPTION	mRNA synthesis
X9P05d09	pir||T10807 annexin 2 (8E-87)	TRANSPORT	
X9P05g06	emb|CAC82712.1|major intrinsic protein 1(1E-115)	TRANSPORT	
PIN	gi|10441744|pttPIN1 (0E-0)	TRANSPORT	Hormone

Table 2-1 (cont'd)

B9SP07h06	pir||T48886 aquaporin (1E-115)	TRANSPORT	Water
X9SP01h07	ref|NP_921277.1|Transposase of Tn10 (0E0)	UNCLASSIFIED	
X9SP04f06	ref|NP_199968.1|phosphate-responsive 1 family (1E-112)	UNCLASSIFIED	
B9P06b11	gb|AAM28295.1|PVR3-like protein (6E-20)	UNCLASSIFIED	
GROUP 2			
L4P02c08	sp|P09189|dnaK-type molecular chaperone hsp70 (1E-133)	CELL RESCUE, DEFENSE	Abiotic stress
X9P10a05	emb|CAC35772.3|dhn1 (4E-42)	CELL RESCUE, DEFENSE	Abiotic stress
B9P03e05	gb|AAD01605.1|copper/zinc-superoxide dismutase (6E-83)	CELL RESCUE, DEFENSE	Detoxification
L1P09a01	sp|O65768|Superoxide dismutase (2E-12)	CELL RESCUE, DEFENSE	Detoxification
L2P04e07	emb|CAC33845.1|cytosolic CuZn-superoxide dismutase (5E-85)	CELL RESCUE, DEFENSE	Detoxification
L3P09g11	ref|NP_194825.1|CBL-interacting protein kinase (4E-67)	SIGNAL TRANSDUCTION	Intracellular
B9P09g06	ref|NP_188015.1|hydroxyproline-rich glycoprotein (4E-71)	CELLULAR ORGANIZATION	Cell wall
L2P09h10	gb|AAG61120.1|RuBisCO activase 1 (1E-119)	ENERGY	Photosynthesis
X9SP06a05	sp|P32869|Photosystem I reaction center subunit II (7E-88)	ENERGY	Photosynthesis
L1P06e05	pir||JC5876 early light-inducible protein (2E-11)	ENERGY	Photosynthesis
B9SP04h05	ref|NP_192705.1|oxidoreductase family protein (3E-58)	METABOLISM	
L4P07h02	sp|Q9LW96|Inositol-3-phosphate synthase (1E-124)	METABOLISM	Carbohydrate
X9P03f10	sp|O78328|1-deoxy-D-xylulose 5-phosphate synthase (1E-140)	METABOLISM	Carbohydrate
L1P02a09	ref|NP_566550.1|UDP-glucoronosyl transferase (2E-56)	METABOLISM	Cell wall
L2P08h01	gb|AAM64731.1|nitrilase associated protein (2E-10)	METABOLISM	Nitrogen
L3P09e05	gb|AAM64577.1|nitrilase-associated protein (2E-11)	METABOLISM	Nitrogen
X9P08a10	gb|AAM95643.1|WD-repeat anthocyanin biosynthesis (1E-131)	METABOLISM	Secondary
B9SP07e03	emb|CAA73220.1|isoflavone reductase (1E-108)	METABOLISM	Secondary
L3P05b05	gb|AAS93803.1|progesterone 5-beta-reductase (1E-78)	METABOLISM	Sterol
L3P05b03	emb|CAC80137.1|progesterone 5-beta-reductase (8E-98)	METABOLISM	Sterol
X9SP03e06	gb|AAM65243.1|carboxyl-terminal peptidase (1E-129)	PROTEIN FATE	Degradation
L1P07b05	ref|NP_179037.2|proline iminopeptidase (5E-91)	PROTEIN FATE	Degradation
X9P02g09	gb|AAO63777.1|cyclophilin (2E-83)	PROTEIN FATE	Folding
X9SP05g09	sp|O49884|ribosomal protein L30 (1E-53)	PROTEIN SYNTHESIS	Ribosome
B9SP07g09	sp|P19951|40S RIBOSOMAL PROTEIN S14 (2E-49)	PROTEIN SYNTHESIS	Ribosome
X9P01h07	gb|AAR01683.1|ribosomal protein L13a (9E-78)	PROTEIN SYNTHESIS	Ribosome

Table 2-1 (cont'd)

X9SP05e04	gb|AAQ08196.1|eukaryotic translation initiation factor (3E-84)	PROTEIN SYNTHESIS	Translation
B9P04b03	sp|O48650|TRANSLATION FACTOR SUI1 (5E-50)	PROTEIN SYNTHESIS	Translation
B9P05d07	gb|AAQ08191.1|eukaryotic translation initiation factor (7E-86)	PROTEIN SYNTHESIS	Translation
X9P10g11	gb|AAC61786.1|glycine-rich RNA-binding protein (2E-38)	TRANSCRIPTION	mRNA processing
X9P02h02	dbj|BAB84582.1|transcription factor LIM (1E-117)	TRANSCRIPTION	mRNA synthesis
X9SP06g05	ref|NP_567050.1|remorin family protein (6E-23)	TRANSCRIPTION	mRNA synthesis
L4P07f05	gb|AAQ84334.1|zinc-finger protein (2E-61)	TRANSCRIPTION	mRNA synthesis
B9P02g05	sp|Q43019|Nonspecific lipid-transfer protein (1E-29)	TRANSPORT	Lipid
X9P04d05	emb|CAE53882.1|aquaporin (1E-149)	TRANSPORT	Water
L4P07f09	sp|Q9ZRX0|Translationally controlled tumor protein (5E-9)	UNCLASSIFIED	
B9P05b07	sp|P35681|Translationally controlled tumor protein (5E-65)	UNCLASSIFIED	
GROUP 3			
X9P02e09	ref|NP_193537.2|dehydration-responsive protein (1E-87)	CELL RESCUE, DEFENSE	Abiotic stress
B9P07a09	gb|AAQ56599.1|chitinase-like protein (9E-51)	CELL RESCUE, DEFENSE	Defense
X9SP06e04	gb|AAL71857.1|dehydroascorbate reductase (2E-96)	CELL RESCUE, DEFENSE	Detoxification
B9SP05b06	gb|AAT02522.1|metallothionein 1a (4E-27)	CELL RESCUE, DEFENSE	Detoxification
B9SP05b04	gb|AAQ74889.1|Al-induced protein (1E-113)	CELL RESCUE, DEFENSE	Detoxification
B9P01b02	pir||JQ2252 peroxidase (4E-87)	CELL RESCUE, DEFENSE	Detoxification
X9P10d02	gb|AAK84479.1|auxin growth promotor protein (1E-70)	SIGNAL TRANSDUCTION	Hormone
L2P10b11	dbj|BAD19898.1|ABI3-interacting protein 2 (2E-41)	SIGNAL TRANSDUCTION	Hormone
X9SP03b06	emb|CAA72330.1|shaggy-like kinase (3E-32)	SIGNAL TRANSDUCTION	Intracellular
X9P11e02	sp|Q9FVI1|Actin-depolymerizing factor 2 (1E-66)	CELLULAR ORGANIZATION	Cytoskeleton
X9P07f07	gb|AAP73462.1|actin (5E-68)	CELLULAR ORGANIZATION	Cytoskeleton
X9P03h11	ref|NP_566065.1|reticulon family protein (7E-86)	CELLULAR ORGANIZATION	ER
X9SP01b03	ref|NP_196738.1|NADH-ubiquinone oxidoreductase (8E-83)	ENERGY	Electron Transfer
X9P02c11	emb|CAC86995.1|ATP citrate lyase a-subunit (1E-141)	METABOLISM	Acetyl-CoA
B9SP08h07	pdb|1QGN|Cystathionine Gamma-Synthase (1E-121)	METABOLISM	Amino acid
B9SP06b10	gb|AAF26735.1|methionine synthase (1E-107)	METABOLISM	Amino acid
B9P03a03	ref|NP_193129.1|glycine/serine hydroxymethyltransferase (8E-66)	METABOLISM	Amino acid
L4P07g09	gb|AAB61597.1|glutamine synthetase (1E-100)	METABOLISM	Amino acid

Table 2-1 (cont'd)

L3P10f05	ref|NP_178224.1|fructose-bisphosphate aldolase/Plastid (3E-91)	METABOLISM	Carbohydrate
L1P10c05	sp|P12858|Glyceraldehyde 3-p dehydrogenase/Plastid (1E-103)	METABOLISM	Carbohydrate
X9P05f02	sp|P21820|Triosephosphate isomerase (1E-89)	METABOLISM	Carbohydrate
B9P01a06	dbj|BAB85987.1|Acyl-CoA-binding protein (6E-32)	METABOLISM	fatty acid
L1P09e05	sp|Q43784|Granule-bound starch synthase (1E-118)	METABOLISM	Starch
X9P04a01	emb|CAA54308.1|1,4-alpha-glucan branching enzyme (1E-157)	METABOLISM	Starch
B9P03h05	emb|CAA54308.1|1,4-alpha-glucan branching enzyme (6E-47)	METABOLISM	Starch
X9P07f06	gb|AAF63239.1|alpha-amylase (1E-66)	METABOLISM	Starch
B9SP05a09	ref|NP_175460.1|dehydrocholesterol reductase (1E-74)	METABOLISM	Sterol
X9SP09h03	gb|AAR12194.1|molecular chaperone Hsp90-2 (2E-47)	PROTEIN FATE	Folding
B9SP06c05	gb|AAM78180.1|polyubiquitin (6E-55)	PROTEIN FATE	Modification
X9SP04g10	dbj|BAA85750.1|polyubiquitin (7E-7)	PROTEIN FATE	Modification
X9P11f02	emb|CAA54603.1|pentameric polyubiquitin (1E-111)	PROTEIN FATE	Modification
L2P08g09	gb|AAF34767.1|60S acidic ribosomal protein (1E-128)	PROTEIN SYNTHESIS	Ribosome
B9SP07d03	sp|P17093|ribosomal protein S11 (2E-69)	PROTEIN SYNTHESIS	Ribosome
X9P06a06	dbj|BAD07825.1|60S ribosomal protein L9 (4E-64)	PROTEIN SYNTHESIS	Ribosome
X9P02g02	gb|AAP80650.1|elongation factor (3E-31)	PROTEIN SYNTHESIS	Translation
L1P10g02	dbj|BAA34348.1|elongation factor-1 alpha (1E-162)	PROTEIN SYNTHESIS	Translation
X9P11h08	sp|O49169|ELONGATION FACTOR 1-ALPHA (1E-69)	PROTEIN SYNTHESIS	Translation
B9P05d10	gb|AAD56020.1|elongation factor-1 alpha 3 (4E-75)	PROTEIN SYNTHESIS	Translation
L2P10c05	ref|NP_056176.2|R3H domain (2E-13)	TRANSCRIPTION	mRNA synthesis
X9SP06g11	ref|NP_191705.1|zinc finger (C3HC4-type RING finger) (2E-11)	TRANSCRIPTION	mRNA synthesis
B9SP03f05	pir||T14332 homeotic protein (6E-52)	TRANSCRIPTION	mRNA synthesis
B9SP03g02	pir||T14563 inorganic diphosphatase (1E-178)	TRANSPORT	Proton
B9P04h05	pir||S54172 inorganic diphosphatase (2E-91)	TRANSPORT	Proton
X9SP03c01	ref|NP_178443.1|nodulin-related (1E-25)	TRANSPORT	
L4P02f09	dbj|BAB09103.1|endosomal protein-like (1E-174)	TRANSPORT	
GROUP 4			
X9SP07d02	ref|NP_567575.1|dehydration stress protein (ERD3) (1E-133)	CELL RESCUE, DEFENSE	Abiotic stress
X9P07d04	gb|AAC24588.1|late embryogenesis-like protein (6E-68)	CELL RESCUE, DEFENSE	Abiotic stress
X9SP04a12	dbj|BAC78212.1|thaumatin/PR5-like protein (1E-83)	CELL RESCUE, DEFENSE	Defense
X9SP09c10	gb|AAD25355.1|pathogenesis-related protein (1E-64)	CELL RESCUE, DEFENSE	Defense

Table 2-1 (cont'd)

B9P07b10	gb|AAF65767.1|glutathione S-transferase (2E-79)	CELL RESCUE, DEFENSE	Detoxification
B9SP01f05	gb|AAK52084.1|peroxidase (2E-96)	CELL RESCUE, DEFENSE	Detoxification
B9P04g09	pir||T10790 peroxidase (1E-137)	CELL RESCUE, DEFENSE	Detoxification
B9P07f03	ref|NP_177592.1|ABA-responsive protein (2E-9)	CELL RESCUE, DEFENSE	Hormone
B9SP10c06	gb|AAM76226.1|germin E protein precursor (2E-78)	CELLULAR ORGANIZATION	Cell wall
B9SP09d08	gb|AAT37955.1|fasciclin-like AGP 12 (1E-89)	CELLULAR ORGANIZATION	Cell wall
B9SP08c05	pir||T09546 extensin (3E-27)	CELLULAR ORGANIZATION	Cell wall
X9P10c03	sp|P47916|S-adenosylmethionine synthetase (1E-134)	METABOLISM	Amino acid
B9SP04d10	dbj|BAA89049.1|sucrose synthase (0E0)	METABOLISM	Carbohydrate
B9SP04b08	gb|AAS45400.1|endo-1,4-beta-glucanase (1E-108)	METABOLISM	Cell wall
B9SP06f12	gb|AAK66161.1|pectate lyase (3E-14)	METABOLISM	Cell wall
B9SP06d11	ref|NP_564059.1|glycosyl hydrolase family protein 17 (5E-20)	METABOLISM	Cell wall
X9SP01b07	gb|AAD39086.1|xyloglucan endo-transglycosylase (1E-102)	METABOLISM	Cell wall
X9SP09b09	pir||T09870 endo-xyloglucan transferase (3E-92)	METABOLISM	Cell wall
B9SP01b06	emb|CAE76634.1|pectin methylesterase (3E-76)	METABOLISM	Cell wall
B9SP05g08	ref|NP_197666.1|glycosyl transferase family 2 (1E-15)	METABOLISM	Cell wall
B9SP10c11	pir||T10054 pyruvate kinase/Plastid (6E-75)	METABOLISM	Glycolysis
B9SP05e02	dbj|BAD09193.1|PrMC3 Esterases and lipases (4E-58)	METABOLISM	Lipid
X9P08c05	gb|AAG30576.1|nitrate reductase (2E-64)	METABOLISM	Nitrogen
X9P01f08	gb|AAN52279.1|phenylalanine ammonia-lyase (1E-107)	METABOLISM	Secondary
X9P01c10	gb|AAQ74878.1|phenylalanine ammonia lyase (5E-80)	METABOLISM	Secondary
X9P09f01	gb|AAN52280.1|phenylalanine ammonia-lyase (1E-93)	METABOLISM	Secondary
X9P08c12	sp|P45730|Phenylalanine ammonia-lyase (1E-114)	METABOLISM	Secondary
X9P07a12	gb|AAF60951.1|O-methyltransferase (0E0)	METABOLISM	Secondary
B9SP04c06	dbj|BAA21643.1|phenylalanine ammonia-lyase (1E-105)	METABOLISM	Secondary
B9SP10d05	gb|AAP31058.1|flavonoid 3',5'-hydroxylase (1E-109)	METABOLISM	Secondary
X9SP05h08	gb|AAC49929.1|flavanone 3beta-hydroxylase (1E-104)	METABOLISM	Secondary
B9SP07f06	sp|O65333|Chalcone--flavonone isomerase (1E-88)	METABOLISM	Secondary
B9P05a10	ref|NP_178637.2|glycosyl hydrolase family (2E-12)	METABOLISM	
L1P10d05	pir||S34285 polyubiquitin (1E-129)	PROTEIN FATE	Modification
X9SP05g06	gb|AAN03468.1|bZIP transcription factor ATB2 (2E-44)	TRANSCRIPTION	mRNA synthesis

Table 2-1 (cont'd)

B9P10g03	ref|NP_568932.2|YT521-B-like family (6E-15)	TRANSCRIPTION	RNA processing
X9SP07d01	gb|AAF28385.1|lipid-transfer protein (3E-13)	TRANSPORT	Lipid
X9P07d05	ref|NP_200019.1|BAG domain-containing protein (6E-21)	UNCLASSIFIED	
B9SP06c02	ref|NP_566712.1|protease inhibitor/storage/lipid transfer (4E-24)	UNCLASSIFIED	
X9SP02a11	ref|NP_179803.3|C2 domain-containing protein (2E-95)	UNCLASSIFIED	
GROUP 5			
L3P09e07	gb|AAL67991.1|dehydration-induced protein RD22 (9E-29)	CELL RESCUE, DEFENSE	Abiotic stress
B9P05d04	gb|AAL26909.1|dehydration-responsive protein RD22 (1E-20)	CELL RESCUE, DEFENSE	Abiotic stress
B9P07f06	ref|NP_195785.1|macrophage migration inhibitory family (5E-48)	SIGNAL TRANSDUCTION	Intercellular
B9SP09d05	emb|CAD66637.1|phytocyanin protein, PUP2 (4E-19)	ENERGY	Electron transport
X9SP02f06	sp|O64416|RuBisCO small subunit (1E-48)	ENERGY	Photosynthesis
L2P10a02	sp|P26301|phosphoglycerate dehydratase/cytosolic (1E-125)	METABOLISM	Glycolysis
L4P02e09	gi|22296820|pyruvate kinase/cytosolic (0E0)	METABOLISM	Glycolysis
B9SP07a03	gi|17380672|phosphoglycerate dehydrogenase (0E0)	METABOLISM	Glycolysis
X9SP03c10	ref|NP_568818.1|eukaryotic translation initiation factor (2E-34)	PROTEIN SYNTHESIS	Translation
B9P08e05	gb|AAM77753.1|translation initiation factor B04 (1E-41)	PROTEIN SYNTHESIS	Translation
L2P09c03	dbj|BAD04010.1|tonoplast intrinsic protein (2E-29)	TRANSPORT	
GROUP 6			
L3P03e02	gb|AAC15460.1|cold-regulated LTCOR12 (5E-32)	CELL RESCUE, DEFENSE	Abiotic stress
X9P08e03	gb|AAM23263.1|DnaJ-like protein (8E-29)	CELL RESCUE, DEFENSE	Abiotic stress
B9SP02a02	gb|AAM46894.1|early drought induced protein (9E-19)	CELL RESCUE, DEFENSE	Abiotic stress
L1P07c04	sp|Q39644|Late embryogenesis abundant protein (3E-12)	CELL RESCUE, DEFENSE	Abiotic stress
L2P08d11	gb|AAD33696.1|PR1a precursor (5E-19)	CELL RESCUE, DEFENSE	Defense
L4P05f06	pir||T08154 pathogenesis-related protein PR1 (5E-19)	CELL RESCUE, DEFENSE	Defense
L2P09a06	gb|AAT01418.1|stress-responsive protein (7E-77)	CELL RESCUE, DEFENSE	Defense
L3P03d02	sp|Q41596|Endochitinase 1 (4E-32)	CELL RESCUE, DEFENSE	Defense
L1P10g09	ref|NP_179828.1|peroxidase 17 (PER17) (1E-105)	CELL RESCUE, DEFENSE	Detoxification
L1P10f07	pir||T10484 blue copper protein stellacyanin (7E-16)	CELL RESCUE, DEFENSE	Detoxification
L1P06e08	gb|AAB24588.1|cupredoxin (2E-14)	CELL RESCUE, DEFENSE	Detoxification

Table 2-1 (cont'd)

B9SP03b11	ref|NP_192897.2|glutathione peroxidase (8E-80)	CELL RESCUE, DEFENSE	Detoxification
L2P08c05	sp|P17598|Catalase isozyme 1 (1E-144)	CELL RESCUE, DEFENSE	Detoxification
L1P01h03	ref|NP_175057.2|protein phosphatase 2C (2E-28)	SIGNAL TRANSDUCTION	Intracellular
L1P07a08	pir||T06482 probable cell wall protein (2E-26)	CELLULAR ORGANIZATION	Cell wall
L2P03f02	ref|NP_194094.1|reticulon family protein (6E-82)	CELLULAR ORGANIZATION	ER
L3P03f10	gb|AAO33591.1|early light induced protein (6E-46)	ENERGY	Photosynthesis
L2P04h07	gb|AAL32038.1|early light-induced protein (3E-40)	ENERGY	Photosynthesis
L1P02f10	ref|NP_188923.1|chlorophyll A-B binding family (6E-14)	ENERGY	Photosynthesis
L3P04g06	pir||T16976 pectinesterase (2E-99)	METABOLISM	Cell wall
L1P04c11	emb|CAD29733.1|pectin methylesterase (4E-45)	METABOLISM	Cell wall
L4P04f05	pir||T50680 beta-1,3 glucanase (1E-141)	METABOLISM	Cell wall
L3P03e07	ref|NP_850759.1|fructose-bisphosphate aldolase/cytosolic (1E-124)	METABOLISM	Glycolysis
L1P05d01	gb|AAN87573.1|delta 12 oleic acid desaturase (1E-131)	METABOLISM	Lipid
L1P04b06	ref|NP_198892.1|glucose-6-P dehydrogenase (0E0)	METABOLISM	PPP
X9P08e06	ref|NP_176904.1|rubber elongation factor family (4E-31)	METABOLISM	Rubber syn.
L2P10b07	emb|CAD87012.1|progesterone 5-beta-reductase (1E-136)	METABOLISM	Sterol
L4P01g09	gb|AAN74808.1|Nicotinate phosphoribosyltransferase (2E-79)	METABOLISM	
L2P01b01	pir||S42882 cysteine proteinase (9E-64)	PROTEIN FATE	Degradation
L1P06c05	ref|NP_194414.1|ubiquitin-like protein (2E-40)	PROTEIN FATE	Modification
L3P08d02	ref|NP_194461.1|zinc finger (5E-13)	TRANSCRIPTION	mRNA synthesis
L1P06a02	emb|CAD48198.1|RNA-binding protein (4E-42)	TRANSCRIPTION	RNA processing
X9P11b09	gb|AAM66114.1|nifU-like protein (2E-53)	TRANSPORT	Nitrogen
L3P08e06	pir||JQ2288 nodulin-26 (3E-39)	TRANSPORT	
B9P05c12	ref|NP_849394.1|mitochondrial membrane translocase (3E-60)	TRANSPORT	
L1P08e03	ref|NP_567178.1|plasma membrane intrinsic protein (1E-129)	TRANSPORT	
L3P03d05	sp|Q07469|BARK STORAGE PROTEIN (3E-11)	UNCLASSIFIED	Storage
L2P10e07	ref|NP_568160.1|protease inhibitor/storage/lipid transfer(7E-38)	UNCLASSIFIED	
L1P10g05	gb|AAS55470.1|little protein 1 (1E-16)	UNCLASSIFIED	
L4P07a02	ref|NP_194476.2|CBS domain-containing protein (9E-39)	UNCLASSIFIED	
GROUP 7			
L4P04f03	gb|AAM60932.1|disease resistance protein (1E-83)	CELL RESCUE, DEFENSE	Defense
L1P01e05	gb|AAP03085.1|class IV chitinase (7E-49)	CELL RESCUE, DEFENSE	Defense

Table 2-1 (cont'd)

L1P03d10	pir||T14329 extracellular dermal glycoprotein (2E-31)	CELLULAR ORGANIZATION	Cell wall
X9P03c07	gb|AAR07596.1|fiber protein Fb34 (3E-34)	CELLULAR ORGANIZATION	
L1P04d01	ref|NP_200316.1|GDSL-motif lipase/hydrolase family (6E-74)	METABOLISM	Lipid
L2P05h10	gb|AAM96867.1|galactinol synthase (1E-118)	METABOLISM	Carbohydrate
L4P04c02	gb|AAP87281.1|beta-1,3-glucanase (1E-104)	METABOLISM	Cell wall
B9P04b01	dbj|BAA02729.1|cytoplasmic aldolase (4E-16)	METABOLISM	Glycolysis
B9P04f06	emb|CAD37200.1|guanine deaminase (7E-70)	METABOLISM	Nucleic acid
L1P06d04	ref|NP_190468.1|AMP-binding synthetase (CoA ligase) (8E-73)	METABOLISM	
L1P04e02	ref|NP_850481.1|F-box family protein (1E-82)	PROTEIN FATE	Degradation
L1P09h10	ref|NP_179571.1|zinc finger (2E-75)	TRANSCRIPTION	mRNA synthesis
GROUP 8			
L2P10h02	gb|AAT02527.1|metallothionein 3b (2E-6)	CELL RESCUE, DEFENSE	Detoxification
L2P01a05	gb|AAT02526.1|metallothionein 3a (3E-27)	CELL RESCUE, DEFENSE	Detoxification
L1P06a05	dbj|BAB82502.1|cytokinine inducible gene3 (3E-25)	SIGNAL TRANSDUCTION	Hormone
X9SP07d12	sp|P51850|Pyruvate decarboxylase isozyme 1 (1E-120)	METABOLISM	fermentation
L1P06f10	ref|NP_172203.1|phosphoesterase family protein (2E-16)	METABOLISM	Lipid
X9P06g05	ref|NP_566473.2|subtilase family protein (6E-17)	PROTEIN FATE	Degradation
L2P05f11	ref|NP_188169.1|no apical meristem (2E-76)	TRANSCRIPTION	mRNA synthesis
X9P05b12	gb|AAQ91200.1|glutathione transporter (1E-131)	TRANSPORT	

CHAPTER 3

An auxin-repressed gene (RpARP) from black locust (*Robinia pseudoacacia*) is posttranscriptionally regulated and negatively associated with shoot elongation.

Abstract

The plant hormone auxin regulates various plant growth and developmental processes by controlling the expression of the auxin-response genes. While many genes up-regulated by auxin have been characterized, less is known about the genes that are down regulated by auxin. Here, we isolated and characterized an auxin-repressed gene (RpARP) from the tree legume, *Robinia pseudoacacia* L. A sequence similarity search in public databases showed that RpARP gene has homologues in various higher plants including monocots and dicots. The deduced amino acid sequences are highly conserved among them (up to 85% identity). Northern blot analysis showed that auxin can repress RpARP gene expression and the repression was dependent on metabolizable sugar and protein synthesis. In addition, cold treatment abolished the auxin-mediated repression of RpARP gene expression. Results from the transgenic plant analyses suggest that the expression of RpARP gene is post-transcriptionally regulated by auxin and the untranslated regions (UTRs) are required for the regulation. Sequence analysis of the promoter region (-70bp and -500bp upstream of putative transcription initiation site) of RpARP gene identified four sucrose-repressible-response elements (TATCCAT-motifs, Huang et al. 1990), suggesting that the *cis*-elements responsible for the regulation by sucrose are located in the promoter region. In fact, the expression of transgenic RpARP gene was not affected

by sucrose when driven by CaMV35 promoter. We also present evidence that RpARP gene expression is negatively associated with hypocotyl elongation.

Introduction

Auxin, a plant growth regulator produced mainly in apical buds, plays a key role in a wide variety of growth and developmental processes such as lateral root formation, apical dominance, tropism, and differentiation of vascular tissue (Berleth et al. 2000, Hamann 2001, Muday 2001). There have been two major experimental approaches to understand the molecular basis of auxin action in such a variety of developmental processes. One is to identify and analyze mutants that lack normal auxin responses. Two related families of proteins, the AUX/IAA proteins and auxin response factors (ARFs), were identified as key regulators of auxin-modulated gene expression (Ulmasov et al. 1997, Rouse et al. 1998). Those protein families function as transcriptional regulators and are thought to activate or repress the auxin-induced gene expression through interaction between them. Recently, the protein degradation by ubiquitin-mediated process was found to be important in auxin response (Ruegger et al. 1998). The other approach involves the use of molecular biology tools to identify and characterize the genes regulated by auxin signals. This molecular approach has led to the identification of several classes of early auxin response genes and auxin responsive *cis*-elements within the promoters of these genes (Guilfoyle et al.1998). The expression level of the early (or primary) auxin response genes increases within minutes of auxin application, independent of *de novo* protein synthesis (Abel and Theologis 1996, Walker and Estelle 1998). Although many mutants and cDNA clones that are associated with auxin signaling have been isolated and characterized, our knowledge of the mechanisms by which auxin regulates various biological functions is still limited. This lack of information may be partially due to the

fact that most of the molecular studies were focused on the 'primary response genes' that are up-regulated by auxin signal (Hagen and Guilfoyle 2002).

Auxin-represssed protein (ARP) genes and their role for plant growth and development are relatively understudied. So far, three orthologs of ARP were isolated and characterized from pea, strawberry, and tobacco. SAR5 was first isolated by differential screening from auxin-deprived strawberry receptacles (Reddy and Poovaiah 1990). They showed that SAR5 gene was repressed by auxin and the repression of the gene was positively correlated with strawberry fruit maturation. Another ortholog PsDRM1 has been reported as bud dormancy-related genes from pea. The abundance of PsDRM1 transcripts in axillary buds declines quickly, within six hours of the removal of the apical buds. Then, they re-accumulate when the axillary buds become latent again and the newly established apical bud becomes active (Stafstrom et al. 1998). Therefore, the gene was named "dormancy-related protein (DRP)." However, it is not known how the gene is regulated at the molecular level, or how the two different plant developmental processes are related with ARP expression. Steiner et al. (2003) reported a new member of auxin-repressed protein family from tobacco, whose transcripts were high during pollen maturation but rapidly declined to minimal levels in germinating pollen.

Recently, we identified an ARP ortholog (named RpARP) from the expressed sequence tags (ESTs) derived from the trunk wood of a 10-year old black locust (*R. pseudoacacia*) (Yang et al., 2003). The gene was expressed across the trunk wood (bark/cambium, sapwood, and sapwood-heartwood transition zone). The deduced amino acid sequence of RpARP shows high sequence homology with other ARP orthologs. Repression of RpARP gene by auxin was dependent on sucrose and post-transcriptionally

regulated. The expression of this auxin-repressed gene is negatively associated with shoot elongation.

Materials and Methods

Plant materials

Black locust seeds were germinated as previously described (Han et al. 1993). Seven-day old hypocotyls (~2.5 cm long) from the seedlings were used. *Arabidopsis thaliana* (L.) Heynh. plants, used in this study, were grown under sterile conditions, the seeds of ecotype Columbia were surface sterilized with 1.5 % (v/v) sodium hypochlorite solution containing 0.02 % (v/v) triton-X100 for 5 min with vigorous shaking, and then washed several times with sterile water. Seeds were chilled at 4°C for 2 to 4 days in water and then were plated onto MS medium (Murashige and Skoog 1962) containing 0.8% agar and 1% (w/v) sucrose (pH 5.7-5.8). The plants were grown at 25°C under a 16h/day photoperiod.

Isolation of full-length RpARP gene and its promoter region

The RpARP gene (Genbank AY009094) was isolated through the black locust EST sequencing project (Yang et al., 2003). Total RNAs were purified from both the cambial and heartwood region of 10-year old black locust trees and used to construct directional cDNA libraries in the pTripleEx plasmid vector (Clontech, Palo Alto, CA). The resulting libraries were applied to EST sequencing. Using a computational contig search, approximately 15 RpARP clones were detected out of the 4000 sequenced clones. Presumed whole transcript of RpARP was determined through multiple alignments. The

RpARP gene promoter was cloned using Genome Walker Kit (Clontech, Palo Alto, CA) following a manufacturer's instruction with 2.5 μg of genomic DNA isolated from a black locust seedling.

Construction of binary vectors and plant transformation

The 700bp fragment of RpARP cDNA including 5'UTR and 3'UTR was prepared by restriction digestion with *Sma*I and *Sac*I from pTripleARP. The fragment was inserted into the corresponding sites of the binary vector pBI121 under the control of CaMV 35S promoter to produce the resulting recombinant vector, pBIARP (Figure 1-7A). For the construction of pBIARPGUS, a PCR reaction with primers containing linker sites, *Xba*I (5') and *Sma*I (3') was performed on pTripleARP to yield 400bp of RpARP coding region. The resulting PCR product was cloned to pGEM-T (Promega, Madison, WI) to produce plasmid pGEMARP (Figure 1-7A). The *Sma*I-*Xba*I fragment of pGEMARP was cloned to the corresponding sites of pBI101 (Genbank U12639.1) to yield pBIARPGUS. *Arabidopsis* ecotype Columbia was transformed using recombinant A. *tumefaciens* by the vacuum infiltration method (Bechhold et al. 1998). Transformants were selected on MS supplemented with 0.7 % (w/v) agar medium containing 1 % (w/v) sucrose and 50 mg/L kanamycin.

Elongation measurement

Black locust seedlings that had uniform hypocotyl lengths (~ 25 mm) were selected and used for elongation measurements. Hypocotyls were cut into 5-mm segments and immersed in MS+2 % (w/v) sucrose medium (Figure 3-9) or in liquid MS medium in

absence or presence of 2 % sucrose, 10 μM NAA, 20 μM Fusicoccin, 2 % mannose, or low temperature (4°C) (Figure 3-5). Incubation was carried out at 25°C with a 16h/day photoperiod on a shaker at 30 cycles per min. Elongation ratios were determined by measuring the increase in segment length after 24 hours of incubation.

Hormone and chemical treatment

Hypocotyl segments (5 mm in length) were cut from below the apical meristem of 7-day old seedlings and incubated in liquid MS medium supplemented with indicated hormones or chemicals with gentle agitation on a shaker at 30 cycles per min at 25°C. All hormones and chemicals used were purchased from Sigma-Aldrich. Working and stock concentrations of hormones and chemicals were used as follows unless otherwise noted: 10 μM Fusicoccin (FC, diluted from 10mM stock solution in DMSO); 50 μM cycloheximide (CHX, from 200 mM in DMSO); 20 μM abscisic acid (ABA, from 10mM in EtOH); 20 μM 2-chloroethyl phosphonic acid (EP, from 100mM in EtOH); 20 μM gibberellic acid (GA, from 50mM in H_2O); 20 μM jasmonic acid (JA, from 200mM in EtOH); 10 μM α-Naphthalene acetic acid (NAA, from 1mM in H_2O). After the treatments, the tissues were frozen in liquid nitrogen and stored at -80°C until use for RNA analysis.

Northern blot analysis

For northern-blot analysis, total RNA was isolated from black locust hypocotyl segments or from whole tissues of transgenic *Arabidopsis* plant using Trizol reagent (Invitrogen, San Diego). Six micrograms of total RNA were separated on formaldehyde-agarose gels,

transferred to Hybond-N membranes. Hybridization of northern blots was performed using UltraHyb solution (Amersham-Pharmacia, Piscataway, NJ) at 42°C according to the manufacturer's instruction. RpARP or GUS DNA labeled with ^{32}P-dCTP in a random primed reaction was used as probe. DNAs used for the probes were obtained by PCR using gene-specific primers. Hybridization signals were detected by autoradiography. For quantification of northern signal, the autoradiography films were scanned and analyzed using ImageQuant program (Amersham-Pharmacia)

Results

The ARP gene family is higher plant-specific.

The BLAST search against NCBI dbEST and The Institute For Genomic Research (TIGR) Plant Gene Indices (http://www.tigr.org/tdb/tgi) identified homologous sequences of RpARP, all of which are from higher plant species (four monocots and eight dicots). No significant hits were obtained from other organisms such as animal and fungal species using a 1.0E-5 E-value cut-off value. Multiple alignments of the deduced amino acid sequences of the orthologs revealed the presence of highly conserved amino acid domains at both N-terminal and C-terminal (Figure 3-1A). The deduced amino acid sequence of RpARP is highly conserved among the plant genes (up to 85% identity at amino acid level). In order to examine the evolutionary relatedness of the 12 putative ARP proteins, we generated an unrooted phylogenic tree with the deduced amino acid sequences (Figure 3-1B). The tree showed two distinct clades of plant ARPs that are separated into dicot and monocot. The two clades are consistent with the evolutionary distance of monocot and dicot (Soltis et al. 1999), suggesting that the ARP gene was attained before monocot

and dicot diverged from the last common ancestor. Furthermore, sequences from all of the legume species (black locust, pea, and soybean) were branched together, indicating that RpARP (from a legume tree) is structurally closer to that of other legume species (pea and soybean) than to that of a tree species poplar. The fact that ARP genes are present only in higher plants and down-regulated by auxin suggests that the gene products play a role in plant growth and development processes specific to higher plants. While the RpARP was repressed rapidly by exogenous auxin as in the previous studies with pea and strawberry ARP genes, it is not known whether these genes are functionally conserved.

Suppression of RpARP gene expression by auxin.

The expression of the RpARP gene in response to exogenous auxin was analyzed using northern blot analysis (Figure 3-2). Total RNA was extracted from hypocotyl fragments of black locust seedlings after treatment with 10 μM NAA or pre-treatment with 10 μM NAA for 24 hours, followed by incubation in auxin-deprived MS medium. The transcript level was found to decline within 3 hours of the application of exogenous auxin, reaching level undetectable by northern blot after 6 hours. However, the RpARP transcripts re-accumulated rapidly when the tissues were transferred to auxin-free MS media (Figure 3-2A). This auxin-mediated repression of the gene was effectively achieved using as low as 0.1 μM NAA (data not shown). To determine if the auxin-mediated reduction of RpARP transcripts was dependent on *de novo* protein synthesis, we carried out the same experiment with a protein synthesis inhibitor, cycloheximide. The cycloheximide treatment retarded the reduction of RpARP transcripts by auxin and also blocked the re-

accumulation of the transcripts in the auxin-free medium (Figure 3-2B), indicating that the protein synthesis was required for the suppression of the gene expression by auxin. Our next question was whether other hormones or abiotic stress signaling was involved in the regulation of the ARP gene expression. To answer that question, we examined the effects of various growth regulators such as ABA, BA, ethylene and GA, and abiotic stresses such as salt stress and cold treatment with or without exogenous auxin (Figure 3-3). Cold treatment seemed to increase RpARP gene expression with or without auxin treatment, while salt slightly depressed the repression by auxin. On the other hand, GA treatment resulted in loss of the transcript even in the absence of auxin. In the presence of sucrose, auxin effectively repressed RpARP gene expression regardless of the treatments with the exception of cold stress.

RpARP gene is suppressed in response to fusicoccin and acidic pH.

According to the acid growth hypothesis (Kutschera 1994), auxin begins by stimulating proton pumping, which results in a hyperpolarization of the plasma membrane, thereby inducing cell elongation through acid-dependent cell wall loosening (Senn and Goldsmith 1988, Rayle and Cleland 1992). To determine the relationship between the physiological action of auxin and RpARP gene suppression, we treated the samples with fusicoccin (FC), which is known to induce rapid cell elongation through the activation of the transport protein H+ATPase in the plasma membrane (Blatt 1988), causing strong acidification of the cell walls (Marre 1979, Kutschera and Schopfer 1985). Figure 3-4 shows that RpARP gene expression was suppressed by both fusicoccin and acidic pH treatments. These results suggest that auxin may suppress the RpARP gene through the

acidification of the cells. Interestingly, under FC treatment, RpARP was suppressed in one hour. Furthermore, FC and low pH treatments appear to be less effective in depressing the RpARP transcript than is auxin. The RpARP transcript was not detectable after 2 hours of auxin treatment (Figure 3-2), while it is clearly detectable even after 6 hours of treatments with FC or low pH (Figure 3-4). In addition, the transcript level was increasing again after 9 hours of FC treatment, suggesting that FC effect might be transient.

Effect of sucrose and low temperature on auxin suppression of RpARP mRNA expression. Our next question was whether the factors associated with the auxin-mediated growth can also affect the expression of RpARP gene. We examined the effects of sucrose starvation and cold treatment (4°C). Sucrose has been known to affect the auxin-mediated elongation and growth (Stevenson and Cleland 1981, Seyedin et al. 1982, Gray et al. 1998) and low temperature has known as a general inhibitor of auxin transport and metabolism (George et al. 1967). In fact, sucrose-deprivation and low temperature conditions inhibit the auxin-modulated elongation of hypocotyl segments (Figure 3-5A). Sucrose was required for auxin-mediated RpARP gene suppression. In addition, sucrose deprivation increased the RpARP expression and under low temperature treatment, the expression of RpARP gene was not affected by exogenous auxin (Figure 3-6A).

In order to gain some insights on the physiological role of sucrose in auxin-mediated suppression of the gene, we tested the effects of sugar analogs, mannose and mannitol on the expression of RpARP gene (Figure 3-6B). Mannitol is known for its impermeability of plasma membrane and mannose as a non-metabolizable sugar by most

plant systems (Gibson 2000). We analyzed the time course expression (0.5, 1, 3, 6, 9, 24 h) of the gene in MS medium containing 2% sucrose, 2% mannose, or 2% mannitol. Mannitol, the impermeable form of sugar, had no effect on the expression of the gene. On the other hand, mannose, permeable but non-metabolizable sugar in most plant systems (Gibson 2000), had an intermediate effect between those of control and sucrose. These results indicate that sucrose is needed for both intracellular osmotic regulation and metabolic energy for the regulation of RpARP gene expression by auxin. This is consistent with previous findings. Auxin can induce an initial growth (Senn and Goldsmith 1988, Rayle and Cleland 1992), but after 4 to 6 hrs the initial growth rate begins to decrease without supply of exogenous sucrose, which is needed for osmotic regulator and new cell wall compounds (Cheung and Cleland 1991, Inouhe and Yamamoto 1991).

Regulation of RpARP expression through mRNA stability

In order to investigate the regulation mechanism and biological function of RpARP, we generated transgenic *Arabidopsis* plants harboring T-DNA containing a full-length RpARP cDNA, with or without the untranslated regions (UTRs), driven by the CaMV 35S promoter (Figure 3-7A). The transgenic plants showed a very similar phenotype to that of the wild type. The expression of RpARP (with UTRs) was repressed in response to exogenous auxin in pBIARP transgenic plants, even though the gene is under the control of constitutive CaMV promoter (Figure 3-7B). On the other hand, the level of RpARP transcipts was not affected by exogeous auxin in the transgenic plants harboring the *35S::RpARPorf:GUS* construct (pBIARPGUS, Figure 3-7A) which contains only 350

bp of RpARP coding region but no UTRs. These results suggest that auxin suppression of the recombinant RpARP transcript is through post-transcriptional regulation and that the elements responsible for the auxin suppression are located on the untranslated regions of RpARP. However, the expression of the transgenic RpARP was not affected by sucrose in transgenic plants transformed with either pBIARP or pBIARPGUS constructs. This suggests that the *cis*-element(s) for sucrose signaling is likely to be located in the promoter region. We sequenced the 2 kb genomic region upstream of the RpARP coding region and searched for putative *cis*-elements. Four sucrose repressible response element (Huang et al. 1990) were found between 70bp and 500bp upstream from the putative transcription start site (Figure 3-8).

RpARP gene expression is negatively associated with hypocotyls elongation

Previous studies indicated that the expression level of ARP genes were low in actively growing tissue (Reddy and Poovaiah 1990, Stafstrom et al. 1998). We examined the expression levels of RpARP gene in the segments of elongating hypocotyls that have different elongation rates. Hypocotyls (25 mm in length) were cut into 5-mm segments and incubated on MS media for 24 hours. Then, percent elongation and transcript levels of RpARP were measured at different regions. The hypocotyl segments showing higher percent elongation had a lower level of RpARP expression and *vice versa* (Figure 3-9). To further confirm that the expression of RpARP gene was negatively associated with hypocotyl elongation, we examined the effects of various treatments that affect shoot elongation. In general, relatively lower levels of the gene expression were observed in the treatments that increased shoot elongation, with exception of SU+FC+ and Mn+N+

treatments (Figure 3-5). This may be due to the transient nature of FC and Mn effects on the repression of the gene. Those results suggest that RpARP gene products might play a role in the cellular processes that are characteristic to non-actively growing tissues.

Discussion

It has long been appreciated that the plant growth regulator, auxin, is an important regulator of such developmental processes in higher plants as tropism, apical dominance, and rhizoid initiation. Auxin can regulate the developmental processes through modification of the expression patterns of the genes involved in the processes (Theologis 1989, Worley et al. 2000). This auxin-mediated regulation mechanism is often highly conserved. For instance, despite the differences in the chemical composition of cell walls, the physiological response of all seed plants to auxin and acid is very similar and the key protein mediating the rapid elongation process, expansin, is highly conserved structurally and functionally among seed plants (Shcherban et al. 1995, Cooke et al. 2002). The ARP gene represents another such example. Here, we isolated one ARP ortholog (RpARP) from a tree legume, black locust (*Robinia pseudoacacia*). The fact that the RpARP gene was also repressed rapidly by exogenous auxin, is consistent with previous studies, supporting the idea that ARP genes might have common patterns of expression and hormonal regulation. A sequence homology search with dbESTs indicates that ARP gene is present in many plant species and its amino acid sequences are highly conserved. The high level of structural conservation between monocotyledonous and dicotyledonous plants suggests that 1) the ARP gene family might have formed before the evolutionary divergence of monocots and dicots, 2) the ARP protein has strict functional constraints

that limit structural modification while maintaining function, and 3) the function of the protein is important to normal development or physiology of the higher plants. Furthermore, it is notable that ARP homologous sequences are found only in higher plants, suggesting its role in higher plant-specific processes.

The RpARP gene is regulated by sucrose as well as auxin. However, these two signals have different regulation mechanisms. For example, the transcript level of RpARP was effectively reduced by exogenous auxin in transgenic plants, transformed with *35S::RpARP* construct that contains *RpARP* structural gene with UTRs, but not in transgenics carrying *35S::RpARPorf:GUS* construct that has the coding region but no UTRs. On the other hand, sucrose had no effect on either construct. These results suggest that auxin regulates the ARP gene posttanscriptioally and the responsible *cis*-elements are likely to be located in the UTR regions. A number of mRNA instability sequences have been identified in UTR regions. For example, AU-rich elements (AREs) known as the archetypal mRNA instability determinant were found in 3' UTR of such mammalian genes as lymphokine and cytokine (Chen and Shyu 1995). Down stream (DST) element which is responsible for mRNA instability in plant, was found in the 3' UTR of the small auxin-up RNA (SAUR) transcripts (Gil and Green 1996). However, despite the fact that the UTRs are involved in auxin-mediated regulation of RpARP gene, no known regulatory element was found in the UTRs of the gene. Therefore, the auxin-mediated regulation of RpARP's transcript stability might represent a novel mechanism for post-transcriptional regulation of gene expression. Fusicoccin, known as H^+-ATPase activator, also repress RpARP gene expression (Figure 3-4). Whether H^+-ATPase activation is involved in this auxin-mediated repression of RpARP is not known.

Cis-elements responsible for sucrose signal might be located at genomic regions upstream of RpARP gene. Four copies of conserved TATCCA *cis*-elements were present in the promoter region of RpARP. This sucrose repressible response element was reported to be responsible for sugar repression of alpha-amylase gene in rice (Lu et al. 1998, Toyofuku et al. 1998). Lu et al. (1998) identified nuclear proteins binding to the TATCCA in a sequence-specific and sugar dependent manner. Sucrose is needed both osmotically and metabolically for auxin- mediated RpARP repression. We tested various sugar analogs for their impacts on RpARP expression. Mannitol (non-permerable through plasma membrane) mimicked the effect of the no sugar control treatment, while mannose (permeable but non-metabolizable sugar; Gibson 2000) showed an intermediate effect between the no sugar control and sucrose treatment. Those results indicate that sucrose is needed for both intracellular osmotic regulation and metabolic energy. Sugars function as important regulators of various processes associated with plant growth and development as well as metabolic resources and structural constituents of cells. In recent studies, sugar was found to favor the expression of growth-related genes while repressing the expression of the stress-related gene (Ho et al. 2001). Both auxin and sucrose are required to completely repress the expression of RpARP gene. This is consistent with the auxin-induced elongation process. According to prevailing acid growth hypothesis, auxin can induce an initial growth through acidification (Senn and Goldsmith 1988, Rayle and Cleland 1992), but after 4 to 6 h the rate begins to decrease without supply of exogenous sucrose in which sucrose is needed for osmotic regulator and new cell wall compound (Cheung and Cleland 1991, Inouhe and Yamamoto 1991). These results led us to a

hypothesis that RpARP gene repression is associated with the elongation process. In other words, RpARP gene needs to be down-regulated for auxin-mediated elongation.

Here, we present two pieces of evidence in support of the hypothesis. First, RpARP mRNA expression was lower at upper positions of hypocotyl which has more enhanced elongation ability than at lower positions. Second, the physiological conditions stimulating elongation can repress the RpARP mRNA expression effectively. Indeed, RpARP gene is expressed at various developmental stages (e.g. from seedling to mature wood) or in various tissue (i.e. leaves, root, hypocotyl and stem), and most abundantly in non-growing tissues. It is quite feasible that RpARP gene needs be repressed for cell growth probably to loose cell wall, modification of cell wall, or increase elasticity of cytoskeleton. This seems to corroborate with previous results with two ARP genes that were isolated and characterized from pea (psDRM1; Stafstrom et al. 1998) and strawberry (SAR5; Reddy and Poovaiah 1990). The repression of SAR5 was associated with strawberry fruit maturation, while the expression level of PsDRM1 gene was high in dormant axillary buds. These two different plant developmental processes are related with cell elongation or expansion process regulated by auxin.

When considering all the data we gathered, RpARP gene is likely to play a role in biological processes that are characteristic for non-growing (i.e., dormant) conditions or stress conditions such as cold and sucrose-starvation. Subcellular localization of the protein, and its interaction with other proteins, and analysis of transgenic plants over-expressing or suppressing the gene will help understand the biological function of the gene. We are in the process of generating transgenic plants for such studies.

References

Abel S,Theologis A (1996) Early genes and auxin action. Plant Physiol. 111: 9-17

Berleth T, Mattsson J,Hardtke CS (2000) Vascular continuity and auxin signals. Trends Plant Sci. 5: 387-393

Blatt MR (1988) Mechanisms of Fusicoccin Action - a Dominant Role for Secondary Transport in a Higher-Plant Cell. Planta. 174: 187-200

Castle LA, Errampalli D, Atherton TL, Franzmann LH, Yoon ES,Meinke DW (1993) Genetic and Molecular Characterization of Embryonic Mutants Identified Following Seed Transformation in Arabidopsis. Mol Gen Genet. 241: 504-514

Chen CY,Shyu AB (1995) AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci. 20: 465-70

Cheung SP,Cleland RE (1991) Galactose Inhibits Auxin-Induced Growth of Avena Coleoptiles by 2 Mechanisms. Plant Cell Physiol. 32: 1015-1019

Cooke TJ, Poli D, Sztein AE,Cohen JD (2002) Evolutionary patterns in auxin action. Plant Mol Biol. 49: 319-38

Gibson SI (2000) Plant sugar-response pathways. Part of a complex regulatory web. Plant Physiol. 124: 1532-1539

Gil P,Green PJ (1996) Multiple regions of the Arabidopsis SAUR-AC1 gene control transcript abundance: the 3' untranslated region functions as an mRNA instability determinant. Embo J. 15: 1678-86

Gray WM, Ostin A, Sandberg G, Romano CP,Estelle M (1998) High temperature promotes auxin-mediated hypocotyl elongation in Arabidopsis. P Natl Acad Sci USA. 95: 7197-7202

Grunwald C, Deutsch F, Eckstein D,Fladung M (2000) Wood formation in rolC transgenic aspen trees. Trees-Struct Funct. 14: 297-304

Guilfoyle T, Hagen G, Ulmasov T,Murfett J (1998) How does auxin turn on genes? Plant Physiol. 118: 341-347

Hagen G,Guilfoyle T (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol Biol. 49: 373-385

Hamann T (2001) The role of auxin in apical-basal pattern formation during Arabidopsis embryogenesis. J Plant Growth Regul. 20: 292-299

Han KH, Keathley DE, Davis JM,Gordon MP (1993) Regeneration of a Transgenic Woody Legume (Robinia-Pseudoacacia L, Black Locust) and Morphological Alterations Induced by Agrobacterium-Rhizogenes-Mediated Transformation. Plant Sci. 88: 149-157

Ho S, Chao Y, Tong W,Yu S (2001) Sugar Coordinately and Differentially Regulates Growth- and Stress-Related Gene Expression via a Complex Signal Transduction Network and Multiple Control Mechanisms. Plant Physiol. 125: 877-90

Huang N, Sutliff TD, Litts JC,Rodriguez RL (1990) Classification and characterization of the rice alpha-amylase multigene family. Plant Mol Biol. 14: 655-68

Inouhe M,Yamamoto R (1991) Effects of 2-Deoxygalactose on Auxin-Induced Growth and Levels of Udp-Sugars in Higher-Plants. Plant Cell Physiol. 32: 433-438

Kotake T, Nakagawa N, Takeda K,Sakurai N (2000) Auxin-induced elongation growth and expressions of cell wall-bound exo- and endo-beta-glucanases in barley coleoptiles. Plant Cell Physiol. 41: 1272-1278

Kurata T,Yamamoto KT (1998) petit1, a conditional growth mutant of Arabidopsis defective in sucrose-dependent elongation growth. Plant Physiol. 118: 793-801

Kutschera U (1994) The Current Status of the Acid-Growth Hypothesis. New Phytol. 126: 549-569

Kutschera U,Schopfer P (1985) Evidence for the Acid-Growth Theory of Fusicoccin Action. Planta. 163: 494-499

Liscum E,Reed JW (2002) Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol Biol. 49: 387-400

Lu CA, Lim EK,Yu SM (1998) Sugar response sequence in the promoter of a rice alpha-amylase gene serves as a transcriptional enhancer. J Biol Chem. 273: 10120-31

Marre E (1979) Fusicoccin - Tool in Plant Physiology. Annu Rev Plant Phys. 30: 273-288

Muday GK (2001) Auxins and tropisms. J Plant Growth Regul. 20: 226-243

Muday GK,DeLong A (2001) Polar auxin transport: controlling where and how much. Trends Plant Sci. 6: 535-542

Reddy ASN,Poovaiah BW (1990) Molecular-Cloning and Sequencing of a Cdna for an Auxin-Repressed Messenger-Rna - Correlation between Fruit-Growth and Repression of the Auxin-Regulated Gene. Plant Mol Biol. 14: 127-136

Rouse D, Mackay P, Stirnberg P, Estelle M,Leyser O (1998) Changes in auxin response from mutations in an AUX/IAA gene. Science. 279: 1371-1373

Ruegger M, Dewey E, Gray WM, Hobbie L, Turner J,Estelle M (1998) The TIR1 protein of Arabidopsis functions in auxin response and is related to human SKP2 and yeast Grr1p. Gene Dev. 12: 198-207

Senn AP,Goldsmith MHM (1988) Regulation of Electrogenic Proton Pumping by Auxin and Fusicoccin as Related to the Growth of Avena Coleoptiles. Plant Physiol. 88: 131-138

Seyedin N, Burris JS, Lamotte CE,Anderson IC (1982) Temperature-Dependent Inhibition of Hypocotyl Elongation in Some Soybean Cultivars .1. Localization of Ethylene Evolution and Role of Cotyledons. Plant Cell Physiol. 23: 427-431

Shcherban TY, Shi J, Durachko DM, Guiltinan MJ, McQueen-Mason SJ, Shieh M,Cosgrove DJ (1995) Molecular cloning and sequence analysis of expansins--a highly conserved, multigene family of proteins that mediate cell wall extension in plants. Proc Natl Acad Sci U S A. 92: 9245-9

Soltis PS, Soltis DE,Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature. 402: 402-4

Stafstrom JP, Ripley BD, Devitt ML,Drake B (1998) Dormancy-associated gene expression in pea axillary buds. Planta. 205: 547-552

Stevenson TT,Cleland RE (1981) Osmoregulation in the Avena-Coleoptile in Relation to Auxin and Growth. Plant Physiol. 67: 749-753

Takahashi Y, Ishida S,Nagata T (1995) Auxin-Regulated Genes. Plant Cell Physiol. 36: 383-390

Theologis A (1989) Auxin-regulated gene expression in plants. Biotechnology. 12: 229-43

Toyofuku K, Umemura T,Yamaguchi J (1998) Promoter elements required for sugar-repression of the RAmy3D gene for alpha-amylase in rice. FEBS Lett. 428: 275-80

Uggla C, Mellerowicz EJ,Sundberg B (1998) Indole-3-acetic acid controls cambial growth in Scots pine by positional signaling. Plant Physiol. 117: 113-121

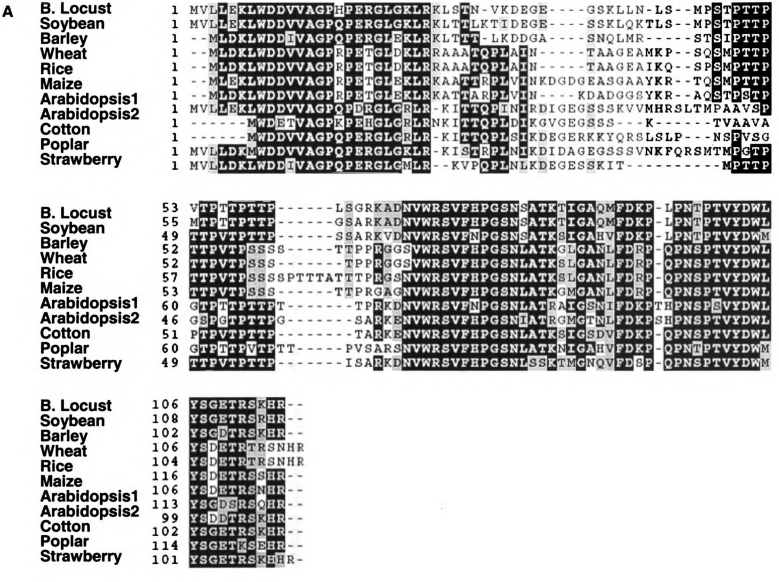
Uggla C, Moritz T, Sandberg G,Sundberg B (1996) Auxin as a positional signal in pattern formation in plants. P Natl Acad Sci USA. 93: 9282-9286

Ulmasov T, Hagen G,Guilfoyle TJ (1997) ARF1, a transcription factor that binds to auxin response elements. Science. 276: 1865-1868

Walker L,Estelle M (1998) Molecular mechanisms of auxin action. Curr Opin Plant Biol. 1: 434-439

Walker L,Estelle M (1998) Molecular mechanisms of auxin action. Curr Opin Plant Biol. 1: 434-439

Worley CK, Zenser N, Ramos J, Rouse D, Leyser O, Theologis A,Callis J (2000) Degradation of Aux/IAA proteins is essential for normal auxin signalling. Plant J. 21: 553-62


A
B. Locust
Soybean
Barley
Wheat
Rice
Maize
Arabidopsis1
Arabidopsis2
Cotton
Poplar
Strawberry

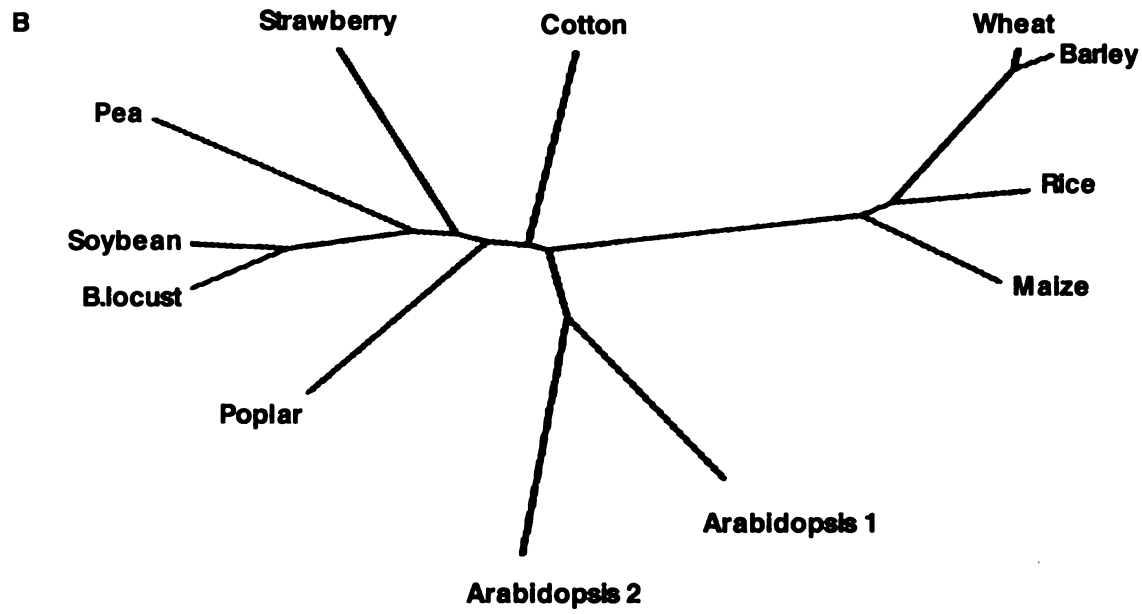


Figure 3-1. Comparison and evolutionary tree of the predicted amino acid sequence of RpARP with other plant ARPs.

(A) Sequences of Arabidopsis1 (At1g28330)a, Arabidopsis2 (At2g38330)a, Barley (TC26321)b, Cotton (TC9854)b, Maize (TC138763)b, Rice (TC82951)b, Soybean (TC132184)b, Wheat (TC40040)b, Pea (AF029242)c, and Strawberry (L44142)c, Poplar (BI136544)c are compared. Reverse type indicate amino acid identities, and shaded gray indicates similar residues. The sequence alignment and sequence identity/similarity was displayed using the Clustal W and BOXSHADE programs of the General Computer Group Package. (B) An unrooted phylogenic tree of the putative ARP proteins shown in (A). The unrooted tree was generated using CTREE program based on the Neighbor-Joining algorithm at A National Laboratory for Compuational Science and Engineering at the University of California San Diego (http://biowb.sdsc.edu). aAGI Number, bTIGR Number, cGenbank Number.

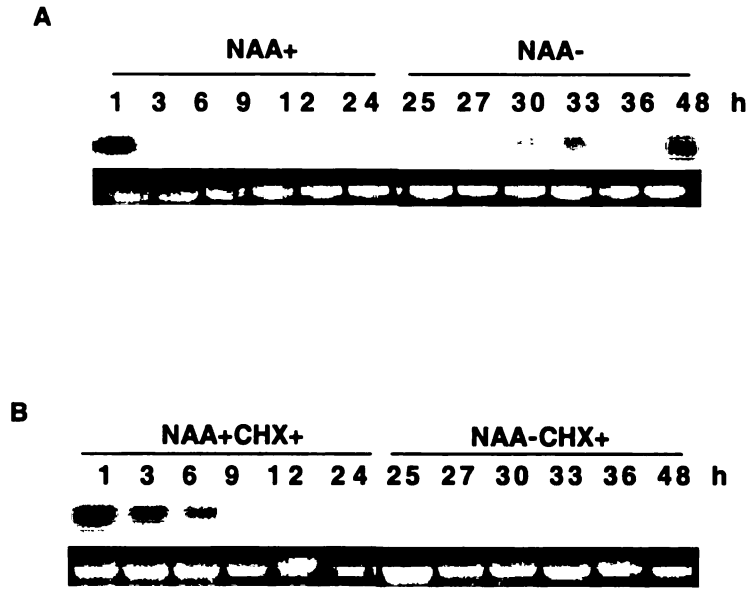


Figure 3-2. Analysis of RpARP mRNA expression in response to auxin and cycloheximide.

(A) The hypocotyl segments of black locust seedling were treated with MS + sucrose (2%,w/v) medium containing 10 μ M NAA (NAA+) for indicated times, or with the NAA free MS-sucrose medium (NAA-) after 24 h pre-treatment with 10 μ M NAA. (B) The hypocotyl segments were treated with the same condition as (A) except for adding 40 μ M cycloheximide (CHX). Total RNA was extracted in indicated times of each blot and loaded 6μ g of total RNA per lane. The blots were probed with a random-primer-labeled RpARP cDNA. Ethidium bromide stained ribosomal RNA is shown in the bottom of each panel as a loading control.

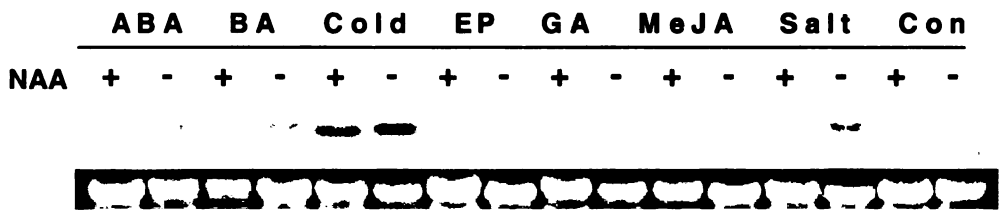


Figure 3-3. Comparison of RpARP transcript levels under various hormonal and abiotic conditions.

Total RNA was extracted from black locust hypocotyl segments immersed for 24 h in liquid MS + sucrose (w/v, 2%) medium containing indicated hormones, abscisic acid (ABA), benzyladenine (BA), ethephon (EP), gibberellin (GA), and methyl jasmonate (MeJA) (each of 20 μ M) or indicated abiotic conditions, 4oC (Cold) and 300 mM NaCl (Salt) in the absence or presence of 10 μ M NAA as indicated (+/-). As control, the samples were incubated in only MS + sucrose medium (Con). 6μ g RNA was loaded per lane and hybridized with RpARP cDNA. Ethidium bromide stained ribosomal RNA is shown in the bottom as a loading control.

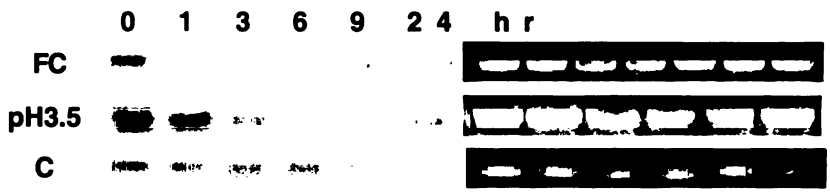


Figure 3-4. Fusicoccin and acidic pH can repress RpARP mRNA expression.
Excised hypocotyl segments of black locust seedlings (7 day old) were treated with MS + 2% sucrose (w/v) medium containing 10 μ M Fusicoccin (FC), buffered with 5 mM MES (pH 3.5), or with no fusicoccin (C) as a control. Total RNA (6 μ g) was isolated after treatment for the indicated times (0.5, 1, 3, 6, 9, 12, 24 h), and then subjected to RNA blot hybridization with probes specific for RpARP. The 25S rRNA bands of the corresponding ethidum bromide-stained gels are shown in the right.

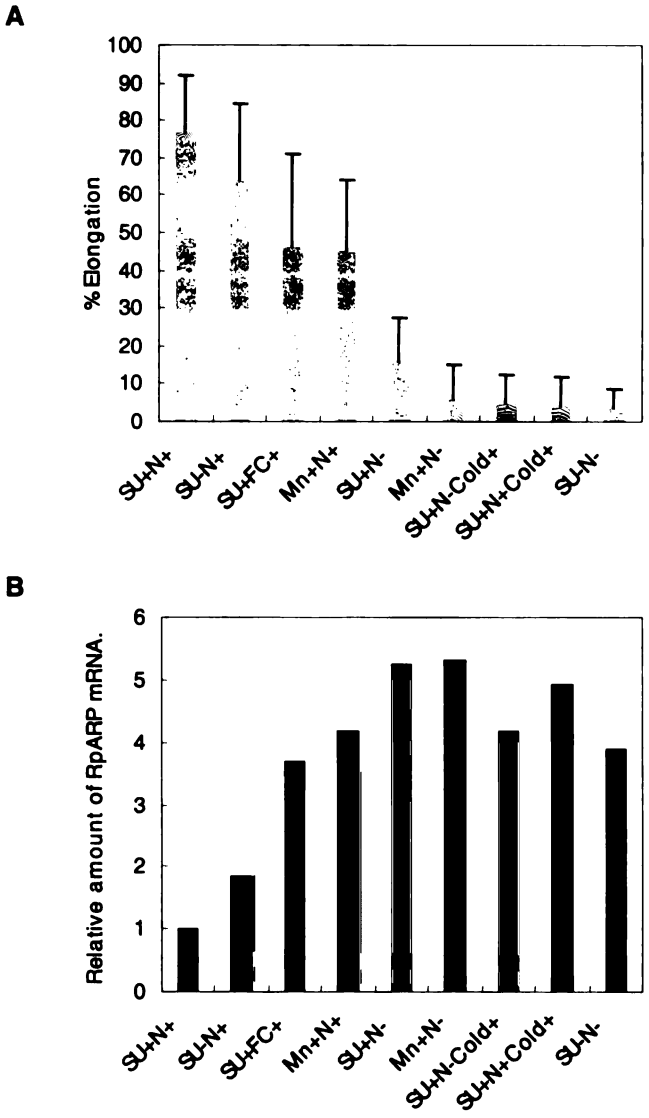


Figure 3-5. Repression of RpARP mRNA is correlated with elongation stimulating phenotype.

The hypocotyl (25 mm length) of black locust seedlings were excised into 0.5 mm segments. The segments were incubated in liquid MS medium in absence or presence of 2 % sucrose (SU), 10 μ M NAA (N), 20 μ M Fusicoccin (FC), 2 % mannose (Mn), or low temperature (Cold, 4oC) for 24 h. (A) For measurement of percent elongation, the length of the segments were measured after 24 h treatment. Data are means + S.D. (%). (B) For measurement of RpARP mRNA expression, total RNA (6 μ g) were isolated from the segments after length measurement, and the autoradiogram signal of the RNA gel blot was quantified using ImageQuant program (Amersham). The values are expressed as a relative value, where the amount of the lowest RpARP mRNA is defined as 1.

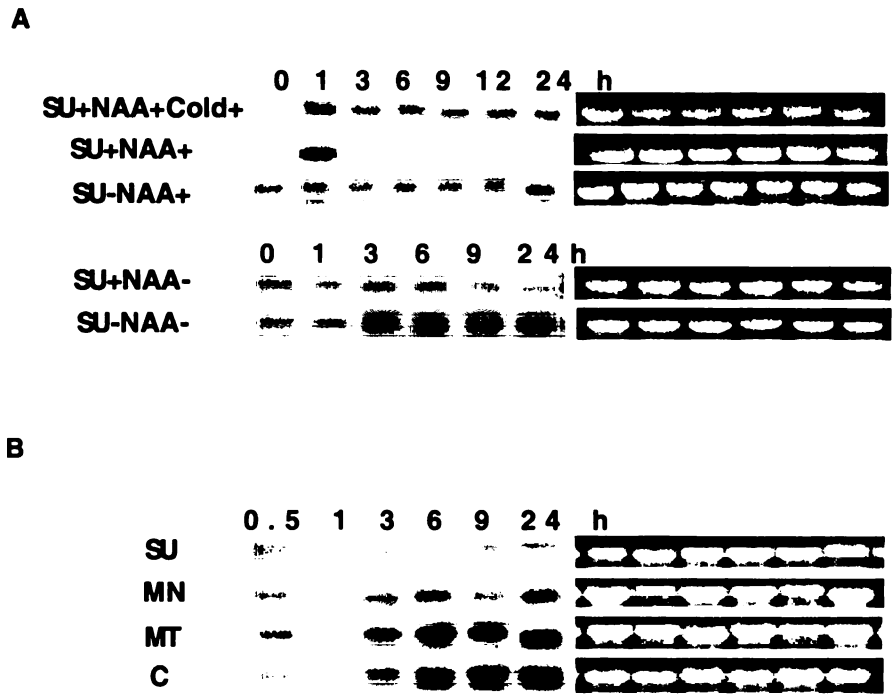


Figure 3-6. Effects of sucrose and low temperature on auxin suppression on RpARP mRNA expression.

(A) The hypocotyl segments were treated with liquid MS medium containing 10 μ M NAA and 2 % sucrose (SU+NAA+), 10 μ M NAA only (SU-NAA+), 2 % sucrose (w/v) only (SU+NAA-), MS medium only (SU-NAA-) or 10 μ M NAA and 2 % sucrose with cold treatments, 4oC (SU+NAA+Cold+). (B) Analysis of RpARP mRNA expression under various sucrose isomers. Hypocotyl segments were incubated for 0.5, 1, 3, 6, 9, 24 h in MS medium containing 2% sucrose (SU), 2% mannose (MN), or 2% mannitol (MT). As a control, MS medium only (C) was used. Total RNA (6 μ g) was isolated after treatment for the indicated times, and then subjected to RNA blot hybridization with probes specific for RpARP. Ethidium bromide stained ribosomal RNA is shown in the right of each panel as a loading control. (SU+NAA+ is a duplication of the left half of Figure 3A

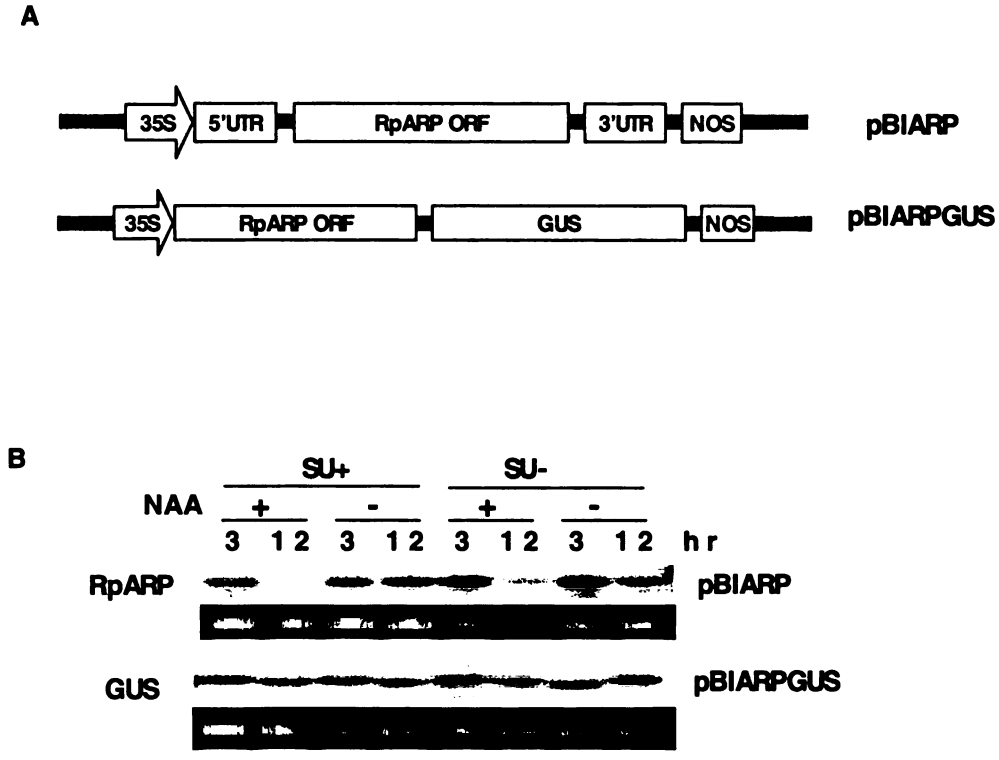


Figure 3-7. The expression of RpARP gene is regulated posttranscriptionally.
(A) Schematic representation of the plasmids used in Arabidopsis transformation and nothern blot. 35S represents CaMV35S;UTR, untranslation region; ORF, open reading frame; NOS, nos terminator. Maps are not to scale. (B) Total RNA was isolated from whole tissues of RpARP overexpressing (pBIARP) transgenic Arabidopsis or RpARP-GUS expressing (pBIARPGUS) after treatment with liquid MS medium in absence (-) or presence (+) of 2 % sucrose (SU) or 10 μ M NAA (NAA) for 3 and 12 h. Each lane was loaded with 6 μ g of total RNA. For hybridization signals, RpARP (Top panel) and GUS (Bottom panel) were used as probe.

Figure 3-8. Nucleotide sequence of RpARP promoter region.
The arrow indicates putative transcriptional start site. The putative TATA box was underlined. The sucrose repressible response element (TATCCAT-motif) is boxed (Huang et al. 1990). The analysis of *cis*-elements has done using the database of PlantCARE (http://oberon.rug.ac.be/PlantCARE)

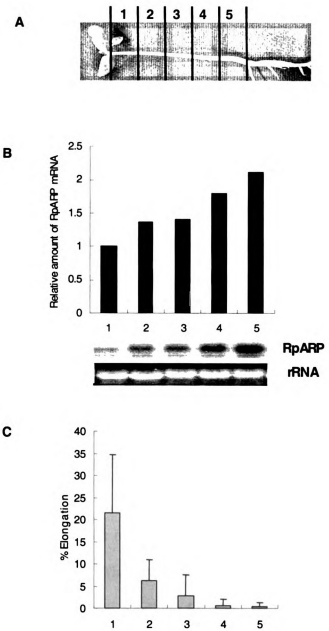


Figure 3-9. Expression pattern of RpARP mRNA is correlated negatively with the extent of elongation of hypocotyl.

(A) The hypocotyl (25 mm length) of black locust seedlings were excised into 0.5 mm segments in order of from top and bottom (1,2,3,4,5). (B) For measurement of RpARP mRNA expression in each segment position, total RNA (6 μ g) were isolated from the segments, and the autoradiogram signal of the RNA gel blot was quantified using ImageQuant program (Amersham). The values is expressed as a relative value, where the amount of the top segments RpARP mRNA is defined as 1. (C) For measurement of percent elongation, the segments were incubated in MS, 2 % Sucrose medium for 24 h and then, the length of the segments were measured. Data are means + S.D. (%)

CONCLUSION

Woody perennials differ in many developmental and physiological respects from herbaceous plants. Recent development of *Populus* genomic resources (e.g., its complete genome sequence and a large number of ESTs) and the advances in genomic and bioinfomatic tools allow me to take an integrated functional and comparative genomics approaches to study the tree specific biological properties.

The large-scale computational analysis of poplar ESTs as part of my thesis works, reveals unique features of expressed genes in the poplar genome. Several multigene families had a higher copy number in poplar than in Arabidopsis suggesting potential lineage-specific proliferation of poplar protein families. Another different feature between *Populus* and *Arabidopsis* transcriptome is that the genes involved in transcriptional regulation are the most divergent while metabolism-related genes are highly conserved.

Another part of my research is about the genetic regulation of seasonal growth cycle in *Populus*. The functional genomic analyses showed that extensive metabolic switch and alteration in cellular functions occurred during transition of rest to quiescent. The genes involved in the pentose phosphate pathway were elevated and defense-related genes were overrepresented during early winter while the genes involved in fermentation and fatty acid β-oxidation were upregulated and signaling-related genes were overrepresented during late winter. Furthermore, clustering of these genes into eight groups based on expression profiles has revealed that plants regulate seasonal growth by integrating environmental factors with developmental stages. The cold related genes were regulated by developmental stages of dormancy. This mechanism could help plants to

control cold hardiness more closely timed with seasonality than unpredictable temperature alone does.

Finally in this thesis work, the auxin repressed gene from black locust (RpARP) has been characterized. As a result, the RpARP gene was repressed by auxin, and the repression was dependent on metabolizable sugar and temperature. In addition, RpARP gene expression is negatively associated with shoot elongation. Interestingly, ARP genes of various plant species characterized from until recently have showed the auxin repression and the close correlation with maturation of organ/tissues. The results may suggest that the ARP gene family perform a similar biological function in various organs.

In conclusion, this study provided the first comprehensive view on genetic regulation of the annual growth cycle and evolutionary adaptation to environment in a woody perennial, *Populus*. This study also showed that the functional and comparative genomic tools recently developed could be used to overcome or compensate for the limitation of the conventional biochemical methods in the study of tree molecular biology.

Ultimately, I believe that the knowledge gained through this study could be leveraged to uncover the evolutionary mechanisms that may account for genetic architechtures underlying the developmental and physiological differences between woody perennials and herbaceous annuals. Within this framework, my thesis study could be further extended to (1) *cis*- and *trans*- transcriptional regulatory networks for the associated signal pathways, (2) species difference and conservation from evolutionary perspective, (3) application of the identified mechanisms to heterologous systems.