8-2006

Expression Analysis of Auxin Regulated Genes in *Populus*

Sara S. Jawdy
University of Tennessee, Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Sara S. Jawdy entitled "Expression Analysis of Auxin Regulated Genes in *Populus*." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Plant Sciences.

Max Cheng, Major Professor

We have read this thesis and recommend its acceptance:

Gerald Tuskan, Robert Trigiano, Stephen DiFazio

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Sara S. Jawdy entitled "Expression Analysis of Auxin Regulated Genes in *Populus*." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Plant Sciences.

Max Cheng, Major Professor

We have read this thesis and recommend its acceptance:

Acceptance for the Council:

Vice Chancellor and Dean
of Graduate Studies

EXPRESSION ANALYSIS OF AUXIN REGULATED GENES IN *POPULUS*

A Thesis Presented for the
Master of Science Degree
The University of Tennessee, Knoxville

Sara S. Jawdy
August, 2006

Dedication

This work is dedicated to my parents, whose example throughout the years has made me ever grateful to have been raised by them.

Acknowledgements

I would first like to thank the members of my committee, Dr. Gerald Tuskan, Dr. Stephen DiFazio, Dr. Max Cheng and Dr. Robert Trigiano. I certainly have learned much from all of them about the scientific process, whether it be through interaction regarding my thesis project or in the classroom. I also owe an infinite amount of gratitude to both Dr. Udaya Kalluri and Lee Gunter for the time and effort they put into teaching a lowly graduate student proper lab technique. I certainly could not have done this without their expertise. Lastly, this acknowledgement would not be complete if it did not include a thank you to my husband, Curt, and my parents, Lyle and Janice, for their loving support and for keeping me sane.

This research was sponsored by the U.S. Department of Energy Office of Science Poplar Genome Based Research for Carbon Sequestration in Terrestrial Ecosystems Program and performed at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

Abstract

Due to its many advantageous characteristics, such as a small sequenced genome, ease of vegetative propagation and availability of genomic tools and databases, *Populus* is widely becoming accepted as the model species among trees. In addition, DOE has chosen hybrid poplar as the model bioenergy feedstock tree. Due to the growing importance of the *Populus* species, genetic and genomic resources (EST and BLAST databases, genetic maps, etc.) are becoming increasingly available and are leading to a greater understanding of the functionality of the *Populus* genome. The goal of this study was to use these resources to further characterize the genetic controls of root growth and development so that these mechanisms may eventually be manipulated to improve carbon sequestration ability in belowground sinks. Because auxin is known to play an important role in lateral root growth as well as many other aspects of plant development, a sequenced subtracted cDNA library from poplar was used to study the expression of genes up and down-regulated in response to exogenous auxin treatments. Results from this study indicate that a daily 9-day exogenous auxin treatment may have induced a stress response as indicated by the high percentage of WRKY transcription factors and stress related proteins that were up-regulated in response to the treatment regimen. A second study was also done using whole-genome oligonucleotide microarray technology to further analyze auxin regulated gene expression including *Populus* homologs of AUX/IAA and ARF genes in Arabidopsis. Results from this study did not appear to correlate well with real-time RT-PCR results indicating that, in the future, more reps need to be used to give the experiment the statistical power necessary to accurately find differentially expressed genes. Results from these gene expression studies can then be used to guide the development of poplar transgenics with increased root growth.

Table of Contents

List of Tables

List of Figures

List of Symbols and/or Abbreviations

1. ABP1 – Auxin Binding Protein
2. ACT – Actin
3. ANOVA – Analysis of Variance
4. ARF – Auxin Response Factor
5. aRNA – Amplified RNA
6. Aux/IAA – Auxin/Indole-3-Acetic Acid
7. AuxRE – Auxin Response Element
8. BAC – Bacterial Artificial Chromosome
9. BAP – Benzyl-Amino-Purine
10. BLAST – Basic Local Alignment Search Tools
11. BLASTn – Basic Local Alignment Search Tools Nucleotide
12. BSA – Bovine Serum Albumin
13. CHL – Chlorophyll a/b Binding Protein
14. C_t – Threshold Cycle
15. CRD – Completely Randomized Design
16. cDNA – Complimentary DNA
17. DMD – Digital Micromirror Device
18. DNA – Deoxyribonucleic Acid
19. DOE – Department of Energy
20. EC number – Enzyme Commission Number
21. EST – Expressed Sequence Tag
22. E-value – Expectancy Value
23. H11-11 – *Populus trichocarpa* x *P. deltoides* (poplar hybrid)
24. IAA – Indole-3-Acetic Acid
25. IBA – Indole-3-Butyric Acid
26. IPA – Indole-3-Pyruvic Acid
27. JGI – Joint Genome Institute
28. KEGG – Kyoto Encyclopedia of Genes and Genomes
29. LB media - Luria-Bertani Media
30. MAS – Maskless Array Synthesizer
31. MIPS – Munich Information Center for Protein Sequences
32. miRNA – micro-RNA
33. mRNA – Messenger RNA
34. NAA – Naphthalene Acetic Acid
35. NPA – Naphthylphthalamic Acid
36. NSWB – Non-Stringent Wash Buffer
37. PAA – Phenylacetic Acid
38. PAT – Polar Auxin Transport
39. PCR – Polymerase Chain Reaction
40. PEG – Polyethylene Glycol
41. PER – Peroxidase
42. PMT – Photomultiplier Tube
43. Q-rich – Glutamine Rich
44. RCA – Rolling Circle Amplification

List of Symbols and/or Abbreviations – Cont.

45. RMA – Robust Multiarray Analysis
46. RNA – Ribonucleic Acid
47. rRNA – Ribosomal RNA
48. RT-PCR – Reverse Transcription PCR
49. SAGE – Serial Analysis of Gene Expression
50. SAUR – Small Auxin-Up RNAs
51. SSH – Suppression Subtractive Hybridization
52. SWB – Stringent Wash Buffer
53. TAIR – The Arabidopsis Information Resource
54. TIR1 – Transport Inhibitor Response
55. Trp – Tryptophan
56. 2,4-D – 2,4-Dichlorophenoxyacetic Acid

Chapter 1 : Goal of the Study

Carbon Sequestration

Levels of atmospheric carbon dioxide have increased steadily from about 280μmol mol^{-1} during the Industrial Revolution to 376μmol mol^{-1} in 2003 (Keeling *et al.* 2004) as a result of increased fossil fuel combustion and land use change (Heath *et al.* 2005). Additionally, levels are expected to continue rising in the future (Rubin *et al.* 1992). As a consequence of the increase in atmospheric CO_2 concentration, climate change in the form of global warming of the atmosphere is occurring, causing what is widely known as a "greenhouse effect". Growing concern about the possible future impact of climate change on the environment and world economy (Houghton *et al.* 2001) has led to a worldwide effort to reduce atmospheric CO_2 levels (Deckmyn *et al.* 2004). The International Climate Change Treaty, also known as the Kyoto Protocol, has accepted the capture, or sequestration, of carbon by plants in soil sinks as one of several valid strategies to reduce the rate of increase of atmospheric CO_2 (Marland and Schlamadinger 1999). The potential use of large-scale hybrid poplar afforestation to sequester carbon in long-lived below-ground sinks has been studied and is reportedly effective (Hansen 1993). Additionally, genetic manipulation of trees to increase their ability to sequester carbon could substantially increase the effectiveness of this particular mitigation strategy (Sedjo *et al.* 1997). Extensive efforts have been made to understand how carbon is allocated from source leaves to sink roots for long term storage within trees (Farrar and Jones 2000, Ericsson *et al.* 1996, Lalonde *et al.* 2004). However, these efforts have stopped short of elucidating the molecular mechanisms that control these source-sink relationships. Before effective genetic manipulation can be carried out, molecular control points of carbon sequestration and partitioning within trees needs to be studied and understood.

Objective

Due to the relatively recent introduction of a wide range of molecular tools that have become available to scientists, it is now possible to directly discover and study

candidate genes controlling carbon sequestration in trees. Those interested in this area of research have turned to a model tree species, *Populus*, as the organism of choice with which to carry out these studies (Wullschleger *et al.* 2002a, Wullschleger *et al.* 2002b). *Populus* has become the model tree species of choice in large part because many members of the genus are important forest crops that have a wide variety of commercial uses (Balatinecz and Kretschmann 2001). Because of this, many characteristics of *Populus*, including its relatively small genome size, abundant variation in natural populations (Brunner *et al.* 2004), ease with which it can be propagated both sexually (Zsuffa *et al.* 1996) and clonally (Bradshaw *et al.* 2000) and its amenability to genetic transformation (Winton 1970, Jouanin *et al.* 1993) have been well studied and used to build a wealth of genomic tools related to the genus. These genomic tools include genetic maps (Yin *et al.* 2002), bacterial artificial chromosome (BAC) libraries, structured pedigrees (Wullschleger *et al.* 2002a) and expressed sequence tag (EST) databases (Sterky *et al.* 1998). Most importantly, efforts to sequence the genome of *Populus trichocarpa* (black cottonwood) were completed in 2005 and work continues to be done in order to improve this crucial resource (Brunner *et al.* 2004).

One of the primary aims of this study is to use H11-11 (*P. trichocarpa x P. deltoides*), a poplar hybrid, and the wealth of genomic tools available to characterize possible molecular controls of root growth in poplar. This aspect of carbon sequestration is important because if mechanisms controlling root growth are understood they can be manipulated to increase root growth, thus increasing the ability of poplar trees to sequester carbon. One way to learn more about molecular controls of root growth is to study the effects of plant hormones that influence root initiation and growth. One such hormone of interest is auxin. The plant hormone auxin has been studied for centuries making it one of the oldest areas of research in plant science (Arteca 1995). Therefore, it is well known that auxin plays an important role in many aspects of plant growth and development including root initiation and elongation (Strivastava 2002).

Many aspects of auxin metabolism, such as its importance as a plant growth regulator, its biosynthesis and molecular transport have been well characterized. However, downstream regulators also need to be characterized. At this time there are 23 known

ARF genes (Guilfoyle and Hagen 2001) and 29 known Aux/IAA genes in Arabidopsis (Liscum *et al*. 2002). The objective of this study is to examine the expression of *Populus* homologs of these genes found using comparative genomics in Arabidopsis. Because auxin induced expression is not limited to just the previously mentioned genes, a *Populus* specific gene discovery technique was also used. The goal is to shed light on the mechanism of auxin response in poplar with the added goal of being able to apply these results to other *Populus* species, including aspen hybrids which are the main organisms used in *Populus* functional genomics research. In turn, these results can later be used to do transformation studies of poplar and aspen genotypes.

In this study, both natural and synthetic auxin treatments were used, first to elucidate possible genes that are either up or down regulated in response to auxin on a whole plant level, and second, to elucidate possible genes that are either up or down regulated in response to auxin specifically in roots. An EST analysis was used to study auxin regulated genes on a whole plant level and whole genome microarray analysis was used to study auxin regulated genes in roots.

Chapter 2 : Literature Review

Introduction

Many members of the genus *Populus* are important forest crops that have a wide variety of commercial uses. Because of this, the characteristics and cultivation of *Populus* have been well studied resulting in the discovery of several aspects of the genus that make it a highly plausible, and much needed, model tree candidate. With a wealth of genomic tools related to the genus now available, it is an ideal subject of research for those who are trying to elucidate molecular characteristics that are unique to woody plants. Once the molecular controls of tree growth and development are known they can be manipulated. The purpose of this study is to use *Populus* and available genomic resources to characterize molecular processes that are up and down regulated in the plant in response to the plant hormone auxin. Of particular long-term interest is how these molecular controls can be manipulated to increase lateral root growth and, consequently, carbon sequestration. In this chapter, characteristics of the *Populus* genus and the plant hormone auxin will be introduced and summarized.

The Genus *Populus*

Taxonomy

The *Populus* genus (known generally as poplar unless a specific subgroup is being discussed) encompasses the deciduous trees aspen, poplar and cottonwood (Bradshaw *et al.* 2000) and its members are a part of the willow family, Salicaceae (Bremer *et al.* 1998). The genus of about 30 species has a wide distribution but its members can be found mainly in the northern hemisphere in temperate climates (Dickmann and Stuart 1983). Both *P. deltoides,* of eastern and central North America, and *P. trichocarpa,* of western North America, are two of eight species that are native to North America (Jones and Luchsinger 1986).

Floral characteristics

Poplars are dioecious and flowering typically occurs before leaves emerge in the spring from floral buds in which preformed inflorescences are contained (Eckenwalder 1996). The male trees produce large amounts of pollen and female trees produce seeds

that are characteristically small and cotton tufted, allowing them to be easily dispersed by wind and rain in the early summer (Braatne *et al.* 1996).

Growth habit

Species of *Populus* are usually single-stemmed and deciduous or, in some cases, semi-evergreen (Eckenwalder 1996). Most are capable of spreading clonally in a field setting, aspen and white poplar in particular, through the production of sucker shoots that form on horizontal roots. Sucker shoots are typically induced after severe stand replacing fires (Burns and Honkala 1990). Poplars can also reproduce clonally by sprouting from the root collar of dead trees or from branches that have become embedded in the soil (Bradshaw *et al.* 2000). Once established, poplars exhibit a rapid growth pattern which is facilitated by the elongation of a preformed shoot from its bud. Expansion of shoots and leaves continue to initiate throughout the growth season leading to a light, diffuse-porous wood structure (Zsuffa *et al.* 1996) and the ability of the trees to rapidly invade disturbed sites (Braatne *et al.* 1996). In addition, these growth characteristics can result in trees that reach heights of forty meters in less than twenty years (Zsuffa *et al.* 1996).

Commercial use

The rapid growth habit and availability of poplar has led to its commercial use in a large variety of North American forest products ranging from paper to chopsticks. One of the most important chief uses of poplar wood is for pulp and paper production to manufacture specialty products (napkins, tissues and roofing felt), building boards (insulation and ceiling tiles) and general purpose pulp. Poplar is not a good candidate for use in the residential construction market due to lower allowable design stresses compared to spruce, pine and fir. However, poplar lumber can be used to make products such as pallets, crates, furniture, interior trim and composite lumber and panels (Balatinecz and Kretschmann 2001). In addition, poplar was suggested as a source of biomass for energy during the energy crisis in the 1970's. During the petroleum shortage, short-term rotation plantations of hybrid poplar trees were established and routinely harvested every 2-5 years. Energy from short-rotation plantations can produce the

equivalent of about twenty-seven barrels of oil per hectare in one year (Isebrands 1979). Other environmental benefits of poplar have been discovered such as their use for windbreaks and shelterbelts, erosion control, phytoremediation and wastewater reuse (Isebrands and Karnosky 2001).

***Populus* as a Model Species**

Why a model tree species is needed

Many characteristics of tree biology are shared by all plants. These common traits have been studied in herbaceous model species, such as Arabidopsis, that are relatively easy to work with and are well characterized. However, there are aspects of biology and physiology that are unique to woody plants (Bradshaw *et al.* 2000) including the formation of secondary xylem and phloem tissue, morphological and physiological phase changes during the aging process, coping mechanisms to combat long-term biotic and abiotic stresses and the capacity to transport water, nutrients and macromolecules long distances. These distinctive characteristics contribute to the ability of wood to be a supportive and conductive structure and, as a result, trees can reach much larger sizes, out-compete for light, water and nutrients, and live much longer than herbaceous plants (Brunner *et al.* 2004). In addition to these differences that are common among most woody plants, poplar species also have roots that form both ecto and endomycorrhizal associations (Martin *et al.* 2004) and are known to support a sizeable number of endophytes within their vascular system (Barac *et al.* 2004), traits that are not characteristic of Arabidopsis.

Due to the differences between woody perennials and herbaceous plants as well as the importance of woody crops and the need for their accelerated domestication (Strauss 2003), the necessity for a model species among trees has become essential. The genus *Populus* (Bradshaw *et al.* 2000) is fast becoming established as the favorite to fill this role because it boasts several advantages above other tree species that will be outlined in the following sections (Wullschleger *et al.* 2002a, Taylor 2002, Wullschleger *et al.* 2002b).

Small genome size

The most advantageous characteristic of *Populus* as a potential model species is related to the fact that the genus has a relatively small genome. The haploid genome size is about 485 million base pairs (Tuskan *et al.* 2006) which are contained on 19 chromosomes (Cervera *et al.* 2001). This is only four times larger than the genome size of *Arabidopsis thaliana* and forty times smaller than a coniferous genome such as loblolly pine (Bradshaw *et al.* 2000). These traits led to a sequencing initiative of *P. trichocarpa* (black cottonwood), the largest native angiosperm in western North America (Fig. 2.1). The sequence was recently completed and a preliminary release of the assembly (7.5X) became available to the public in fall 2004. Additional ongoing work will result in the release of a final draft of the sequence at a later date. With the sequence of poplar available, the power of genomic resources for woody plants has been greatly increased (Brunner *et al.* 2004).

Genetic variation in natural populations

Because the genus has a wide range of distribution, from the tropics to the Arctic Circle, its members have adapted to a diverse set of environmental conditions (Brunner *et al.* 2004). The result is abundant variation, or polymorphism, in natural populations, (Taylor 2002) and much of the variation has been shown to be moderately to strongly genetically controlled (Bradshaw *et al.* 2000). Genetically controlled variation is a valuable tool useful for analyzing physiological mechanisms within any organism. For example, high amounts of polymorphism within a tree population can be used to facilitate association mapping of genetic and phenotypic variation (Yin *et al.* 2004).

Ease of sexual propagation and hybridization

Few trees can be sexually propagated as easily and efficiently as poplar trees. Controlled crosses can be carried out in greenhouses with female branches that have been

Figure 2.1: *Populus trichocarpa* (a) leaves and (b) a natural stand

removed from stock trees. Male pollen can be collected and stored for several years before it is used for breeding and each pollination event yields hundreds of seeds in a relatively short time (4-8 weeks).

Germination generally occurs within twenty-four hours and results in seedlings ranging from one to two meters tall within a year. Hybridization can take place between members of the same section as well as between many members in different sections. In addition, the diploid nature of the genus results in hybridizations that yield fertile progeny (Zsuffa *et al.* 1996). One particular poplar hybrid, known as H11-11 (*P. trichocarpa* x *P. deltoides*), has been used extensively in poplar research and was chosen to do gene expression analysis in this study. The original cross between *P. trichocarpa* and *P. deltoides* was carried out at the University of Washington in 1968 (Stettler 1968 and Subramaniam *et al.* 1993). After the original cross was made, several of the resulting hybrid clones were evaluated and H11-11 was among the few that showed superior performance (Heilman and Stettler 1984).

Ease of clonal propagation

Several members of the *Populus* genus can be easily vegetatively propagated using stem cuttings which can be rooted in water or directly in soil. A few of these include *P. trichocarpa* (black cottonwood) and *P. angustifolia* (narrowleaf cottonwood), both of which are used extensively in genetic studies (Dickmann and Stuart 1983). Cuttings are usually harvested from one-year-old plants during dormancy, which occurs from early winter to early spring before buds flush. Low temperature (just above or just below freezing) storage of cuttings in plastic bags is sufficient to overcome dormancy and serves to protect cuttings from heating and moisture loss (Cram and Lindquist 1982). Clonal propagation is important because it can be used to capture and replicate genetic variation in time and space so that separate experiments can be carried out with the same genetic material. It also allows growth of plants in a controlled field setting that might not otherwise survive in a competitive environment. Lastly, multiple clones of the same tree can be used for destructive sampling for physiological studies, stocks of important

clones can be maintained and materials can easily be shared with collaborators (Bradshaw *et al.* 2000).

Ease of transformation

Genetic transformation of poplars is an important tool used to confer desired or interesting traits that otherwise might not be readily available using conventional breeding methods (Han *et al.* 1996). In addition to commercial applications such as tree improvement, gene function can be studied using transgenic poplars (Sterky *et al.* 1998). Additionally, studies have shown that creating transgenic poplars is relatively easy and efficient when compared to using the technique in other forest tree species. Hybrid aspen (*P. tremula* x *P. alba*) in particular can be used to produce transgenic trees within six to ten months after successful transformation with *Agrobacterium* (Jouanin *et al.* 1993). Successful transformation has been demonstrated using virtually all available plant tissues including callus tissue (Winton 1970), shoot and root tips, nodes, internodes, gametophytic and sporophytic tissues, leaf and petiole fragment and midvein and cambial tissue (Jouanin *et al.* 1993). Techniques utilizing somatic embryogenesis (Michler and Bauer 1991) and protoplasts (Russell and McCown 1988) have also been successful.

Availability of genomic tools and databases

Due to the previously mentioned advantageous characteristics of *Populus*, much work has already been done on the genus. As a consequence, resources such as genetic maps are available and continue to improve (Yin *et al.* 2004, Wu *et al.* 2000), bacterial artificial chromosome (BAC) libraries have been constructed (Wullschleger *et al.* 2002a) and expressed sequence tag (EST) databases are available (Sterky *et al.* 1998). Additionally, structured pedigrees, molecular markers, and gene sequences are accessible (Wullschleger *et al.* 2002a). With all of these resources at hand, the stage is set to learn more about the functionality of the *Populus* genome. Ultimately, the goal among poplar biologists is to gain greater understanding of individual genes so that gene expression and allelic composition can be manipulated to obtain specific desired phenotypes (Bradshaw *et al.* 2000).

DOE model bioenergy feedstock tree

Lastly, hybrid poplar has been chosen as the DOE model bioenergy feedstock tree (Tuskan 1998) because it can be used as a source of renewable energy. Additionally, *Populus* root systems have a remarkable capacity to store and release carbon for use later to fuel growth and maintain tissue biomass (Pregitzer and Friend 1996). For example, a *Populus* hybrid (*P. nigra* x *P. deltoides*) has a tremendous ability to load carbohydrates into roots, with a 20-fold increase in carbohydrate content of roots and a 75-fold increase in starch content in fine roots, when comparing August to November (Nguyen *et al.* 1990). Minor genetic modifications to the poplar genome could potentially result in substantial global carbon sequestration due to the large acreages of poplar trees that could be planted in the near future (Tuskan and Walsh 2001). The carbon sequestration ability of the trees (Farrar and Jones 2000, Dixon *et al.* 1994) could be enhanced using transgenic technology to increase root growth (Bent 2000). Before effective genetic modifications can be made, however, we need to gain a better understanding of the molecular controls of plant root growth.

Auxin

The history of auxin

One way to learn more about genetic control of root growth is to look at the effects of plant hormones that influence root initiation and growth. One such plant hormone of interest is auxin. Much is known about auxin, as auxin biology is one of oldest areas of research in plant science. The phenomenon of gravitropism, or bending of roots toward gravity, was first observed by T. Ciesielski in the late 1800's (Ciesielski 1872). Shortly afterwards Charles Darwin expanded on this research in his book, *The Power of Movement in Plants*, in which he described the effects of light on coleoptile movement. He observed that movement toward light occurred only if the coleoptile tip was not removed from the seedling (Darwin 1880). Subsequent research on the yet unnamed compound controlling the observed gravitropic and phototropic movements was conducted by several scientists including Fitting (1907), Boyson and Jensen (1913), Paal (1918) and Soding (1925) (Arteca 1995). Eventually, the term auxin was coined by a

graduate student named Fritz Went in 1926 who showed that a chemical from excised coleoptile tips had the ability to stimulate growth of decapitated coleoptiles in the dark (Salisbury and Ross 1991). Another source states that the term was first used by Kögl and Haagen Smit in 1931 while studying plant growth modulating substances in human urine they called auxin A and B and heteroauxin (Kögl and Haagen Smit 1931). Heteroauxin was later determined to be indole-3-acetic acid or IAA (Thimann 1977) and the term is still used today, while the term auxin is used to describe a variety of compounds that are structurally different but lead to varying degrees of auxin type responses (Srivastava 2002).

Endogenous auxin

Auxin bioassay studies have revealed many endogenous auxins in addition to indole-3-acetic acid (Fig. 2.2). For example, a lesser known chlorinated form of IAA which exhibits a high amount of auxin activity has been found in several plants (Slovin *et al.* 1999). Phenylacetic acid or PAA, IAA precursors such as indole-3-acetonitril and indole-3-pyruvic acid (Thimann 1977) and indole-3-butyric acid or IBA are also active in bioassays and found in almost all plants. IBA in particular has been studied extensively and is structurally identical to IAA with the exception of two additional methyl groups (Woodward and Bartel 2005).

Synthetic auxins

Auxin can also be created synthetically and are available in the commercial market to be used for a variety of purposes including herbicides and in auxin related experiments. The two main types of synthetic auxins (Fig. 2.3 and 2.4) are 2,4-dichlorophenoxyacetic acid (2,4-D) related compounds and naphthaleneacetic acid (NAA). 2,4-D was first developed during the Vietnam war era when it was used, along with 2,4,5-T, in Agent Orange to defoliate trees (Arteca 1995). 2,4-D is now widely used

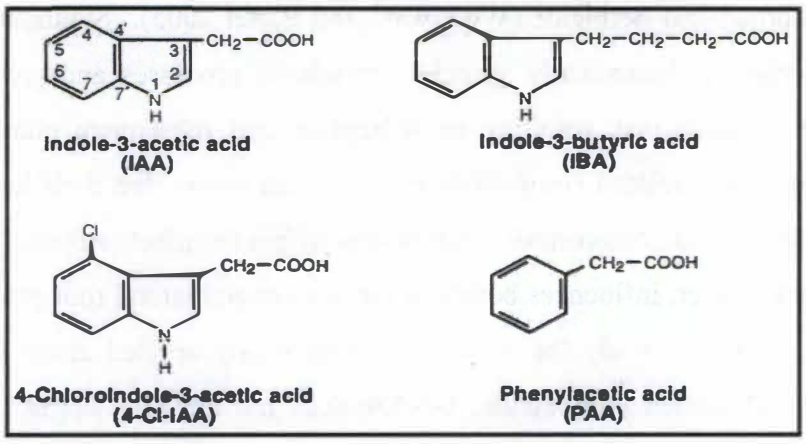

Figure 2.2: Chemical structure of four endogenous auxins IAA, IBA, 4-Cl-IAA and PAA
From Srivastava (2002)

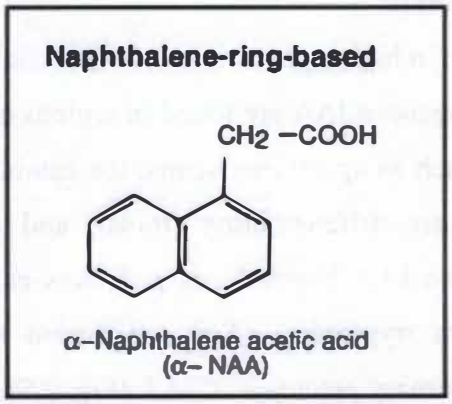

Figure 2.3: Chemical structure of the synthetic auxin NAA
From Srivastava (2002)

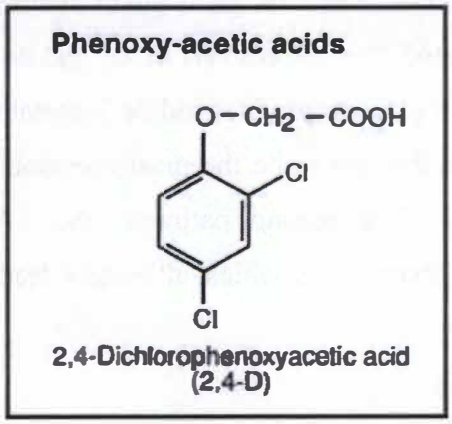

Figure 2.4: Chemical structure of the synthetic auxin 2, 4-D
From Srivastava (2002)

as a selective broad-leaf herbicide (Woodward and Bartel 2005). Synthetic auxin that is used as an herbicide dramatically stimulate metabolic processes and growth beyond a level that can be sustained, resulting in defoliation and subsequent plant death. This occurs because 2,4-D related compounds mimic auxin action, but their levels cannot be regulated by the endogenous control mechanisms within the plant (Arteca 1995). 1-NAA, the active NAA isomer, influences both root elongation and lateral root production and is used in experiments to study the effects of exogenously applied auxin on plants. In contrast to 2,4-D related compounds, 1-NAA does not act as a herbicide (Srivastava 2002).

Endogenous IAA biosynthesis

Trp-dependent IAA biosynthesis

The principle auxin in higher plants is indole-3-acetic acid IAA (Taiz *et al.* 1998). The highest levels of endogenous IAA are found in regions of the plant where active cell division is taking place such as apical meristems, the cambium (Hooykaas *et al.* 1999), organs and tissues that are differentiating (roots), and areas of root proliferation (Salisbury 1991). Multiple IAA biosynthesis pathways exist (IPA, TAM, IAOx, and IAM pathways) in which tryptophan (Trp), an amino acid with and indole ring (Srivastava 2002), is the primary precursor of IAA (Fig. 2.5). Two of these pathways are thought to function in any given plant species (Taiz *et al.* 1998, Arteca 1995). However, there are still essential components in each pathway that have not yet been characterized (Woodward and Bartel 2005). The first plant auxin synthesis pathway (IPA pathway) involves the deamination of Trp to convert it to indole-3-pyryvic acid (IPA). A decarboxylation event takes place to produce indole-3-acetaldehyde, which is oxidized to IAA. The IPA pathway is thought to be the most common of all possible pathways in plants (Srivastava 2002). The second pathway, the TAM pathway, involves the decarboxylation of Trp to tryptomine, which ultimately leads to the production of IAA (Arteca 1995).

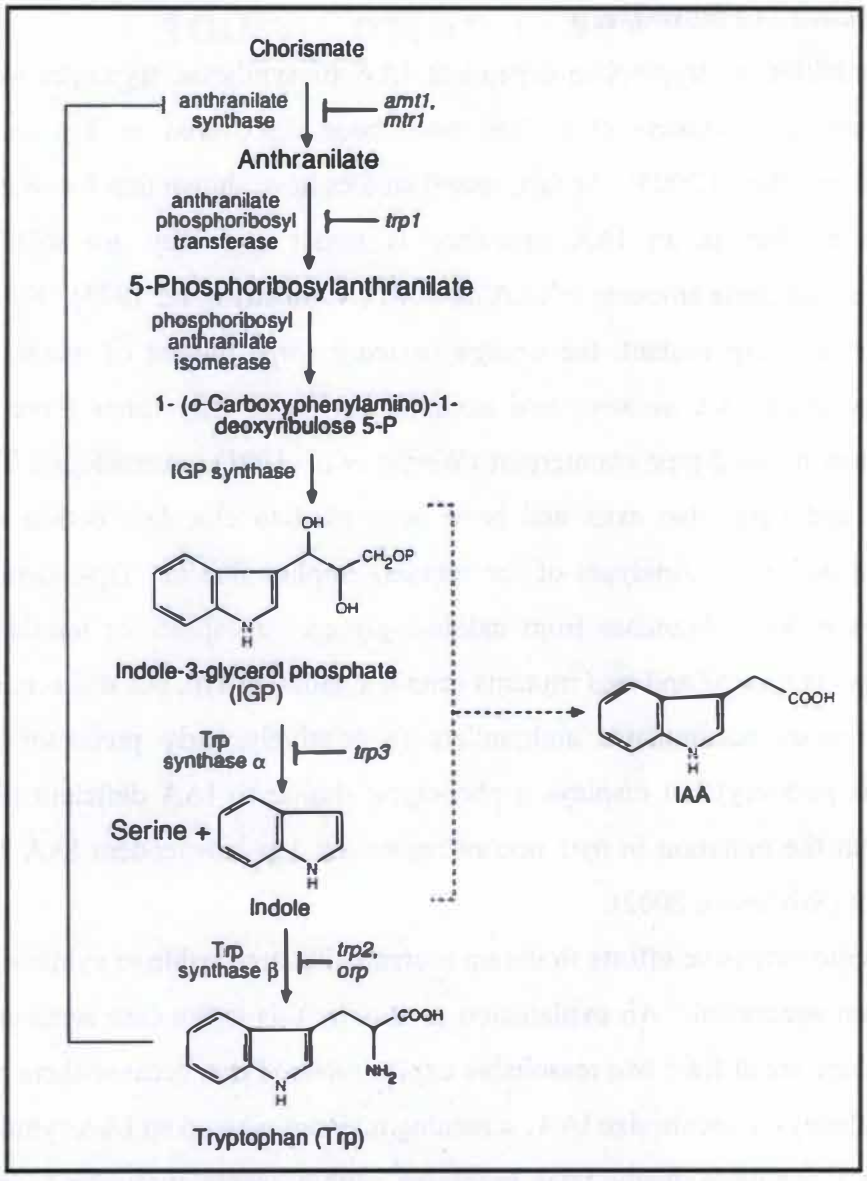

Figure 2.5: Trp dependent IAA biosynthesis pathway
From Srivastava (2002)

Trp-independent IAA biosynthesis

In addition to tryptophan-dependent IAA biosynthesis, tryptophan-independent IAA biosynthesis pathways (Fig. 2.6) have been discovered in Trp mutant plants (Woodward and Bartel 2005). In fact, recent studies have shown that for some plants the importance of Trp as an IAA precursor is minor and they are still capable of accumulating adequate amounts of IAA *de novo* (Normanly *et al.* 1993). For example, a double recessive Trp mutant, the *orange pericarp* (*orp*) mutant of maize, retains the ability to produce IAA *de novo* and accumulates up to fifty times more IAA in its seedlings than its wild-type counterpart (Wright *et al.* 1991). Arabidopsis Trp mutants, *trp1, trp2,* and *trp3,* also exist and have been used to elucidate details of the Trp-independent pathway. Analyses of the mutants implies that the Trp-independent IAA biosynthesis pathway branches from indole-3-glycerol phosphate or indole during Trp biosynthesis. Both *trp2* and *trp3* mutants cannot synthesize Trp, but still accumulate IAA. The *trp1* mutant accumulates anthranilate (a relatively early precursor in the Trp biosynthesis pathway) but displays a phenotype similar to IAA deficient plants. This indicates that the mutation in *trp1* occurs before the Trp-independent IAA biosynthesis branch point (Srivastava 2002).

Despite extensive efforts to obtain mutants that are unable to synthesize IAA, no one has been successful. An explanation as to why this is the case remains uncertain. However, there are at least two reasonable explanations. First, because there are so many possible pathways to synthesize IAA, screening may not pick up an IAA synthesis mutant. Second, IAA regulates many vital processes within plants throughout development. Therefore, plants that are unable to synthesize IAA might not be viable (Srivastava 2002).

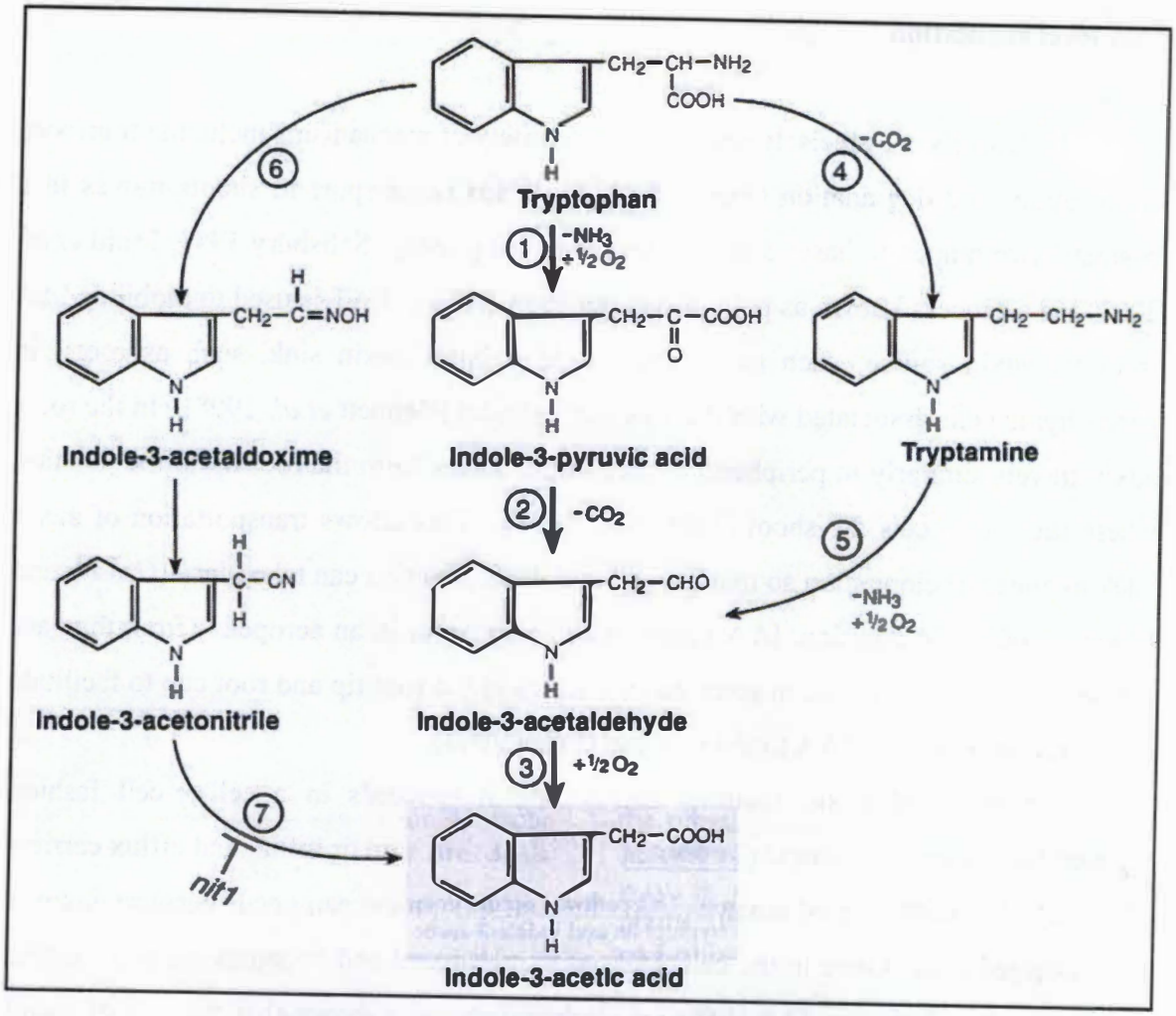

Figure 2.6: Trp independent IAA biosynthesis pathway
From Srivastava (2002)

IAA level regulation

Transport

IAA levels are precisely regulated by a variety of mechanisms including transport, conjugation, and degradation (Fig. 2.7). Active auxin transport in shoots moves in a basipetal (from apex to base) direction regardless of gravity (Salisbury 1991, Friml *et al.* 2002) via a process known as polar auxin transport (PAT). PAT is used to mobilize IAA from an auxin source, such as the shoots, to a basal auxin sink, such as roots, in parenchyma cells associated with the vascular cylinder (Bennett *et al.* 1998). In the roots, auxin travels similarly in peripheral tissues of the cortex from the root tip to the junction where the root meets the shoot (Taiz *et al.* 1998). This allows transportation of auxin back to zones of elongation so that growth and differentiation can take place (Casson and Lindsey 2003). In addition, IAA transport can also occur in an acropetal (from the base towards the apex) direction in some cases such as in the root tip and root cap to facilitate lateral redistribution of IAA to the root cap (Friml 2003).

Transport of auxin requires energy and it proceeds in a cell-to-cell fashion through the plasma membrane (Taiz *et al.* 1998) via diffusion or influx and efflux carriers (Friml 2003). Diffusion of auxin through the cell membrane can occur because auxin is an uncharged acid. Once in the cell, IAA is de-protonated and becomes trapped causing accumulation. However, IAA diffusion experiments have shown that the rate of travel via diffusion is much slower than the rate at which IAA transport actually occurs (5-20 mm/hr). Therefore, it is more likely that PAT is facilitated by influx and efflux carriers. In addition, an efflux carrier must be utilized in PAT because charged IAA could not otherwise move out of the cell (Srivastava 2002).

The influx carriers, or symporters, are thought to be AUX1 proteins. Studies postulate that aux1 proteins are located on all sides of the cell and IAA enters through them in a protonated form. Plants with a mutation in the AUX1 gene are auxin insensitive and show a severely decreased ability to execute gravitropic responses (Bennett *et al.* 1996). Dissociation of IAAH into IAA^- and H^+ inside the cell prevents the negatively charged IAA from leaving the cell except at specific sites where an efflux

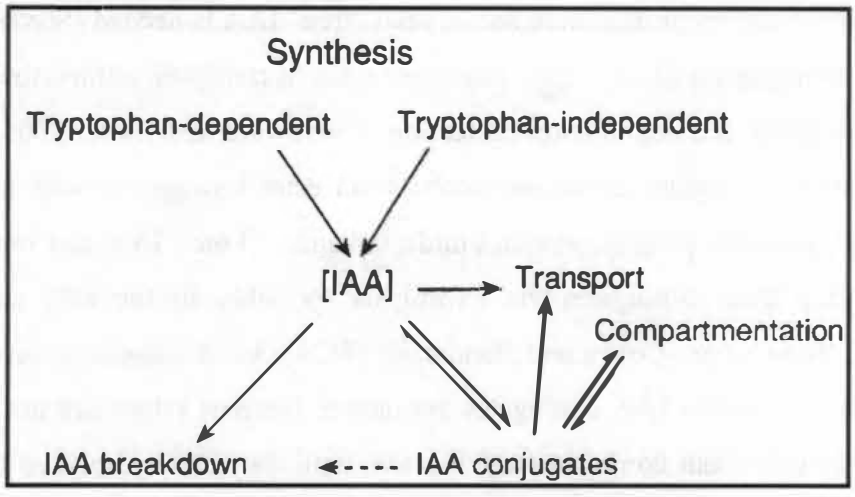

Figure 2.7: IAA level Regulation
From Srivastava (2002)

carrier is located. The efflux carrier is thought to be made up of a family of eight or more proteins known as pin proteins. Pin proteins are located at the morphologically basal end of the cell and ensure that charged IAA will exit out of the cell in a basipetal direction except where acropetal movement occurs in limited portions of roots (Friml 2003). Pin proteins function to establish IAA gradients, which have been shown to be more important than absolute auxin concentration. Auxin gradients are needed to facilitate most auxin responses in roots and shoots, from tropic responses to floral organ formation, because they cause asymmetric cell growth. Plants that are deficient in normal pin protein localization and function cannot shuttle IAA, auxin gradients are not properly established and growth is generally altered in some way. For example, *pin2* mutants have areas of abnormally high IAA concentration in root tips because charged IAA is not shuttled out of cells and distributed properly. As a result, root length and meristem size are diminished. In the case of *pin1* mutants, the lack of properly established auxin gradients results in plants that cannot form floral organs (Friml 2003).

Conjugation

Higher plants keep "free" (active) IAA levels low and store most IAA either in a conjugated form known as "bound" (inactive) IAA or in the form of IBA (Tam *et al. 2000*). Conjugation serves to reversibly inactivate IAA to facilitate storage, mainly in

either vegetative tissues or immature seeds, until "free" IAA is needed (Srivastava 2002). In addition, conjugation of IAA may function to aid in transport, compartmentalization, IAA detoxification, and degradation protection (Woodward and Bartel 2005). IAA can be conjugated with sugars and sugar alcohols via ester linkages or with amino acids, peptides, and possibly proteins via an amide linkage. "Free" IAA can once again be made available from conjugates via hydrolysis (possibly in the ER) and IBA via peroxisomal β-oxidation (Cohen and Bandurski 1982). Auxin bioassays have been used to determine that certain IAA conjugates are active, whereas others are not. However, the general belief is that conjugates are inactive until they are hydrolyzed by the plant tissue and the resulting free IAA is the dominant active form. IAA conjugates that are inactive in auxin bioassays are hypothesized to be intermediates targeted for IAA degradation (Woodward and Bartel 2005).

Degradation

Degradation results in irreversible inactivation of IAA and is thought to occur in two main pathways, nondecarboxylation and decarboxylation. The nondecarboxylation pathway has been well documented in many plants including rice, corn, bean, Brassica, poplar, spruce and pine. In this pathway, IAA is conjugated to aspartate and then oxidized to oxindole-3-acetylaspartate. Further modifications may also occur to the indole ring to facilitate degradation. In an alternative nondecarboxylation pathway which seems to be less common, the IAA indole ring is oxidized while the carboxyl group is left intact. The main products from this alternative pathway are oxindole derivatives (Srivastava 2002).

Evidence for the decarboxylation pathway is not as well supported as the nondecarboxylation pathway, as it has only been documented in *in vitro* assays where tissues are exogenously supplied with IAA. In this pathway it is possible that peroxidases, along with hydrogen peroxide and cations such as Mn^{2+}, degrade IAA to decarboxylated products such as oxindole-3-methanol and 3-methyloxindole to facilitate eventual degradation (Srivastava 2002).

IAA functions in plant growth and development

Endogenous IAA is an essential element controlling many facets of plant growth and development from the embryonic to the reproductive adult stages of growth. (Strivastava 2002). For example, IAA has been shown to play a role in diverse physiological processes such as embryo development (Geldner *et al.* 2000), tropic responses to light (Liscum and Stowe-Evans 2000) and gravity (Parker and Briggs 1990), apical dominance (Gocal *et al.* 1991), cell division and elongation (Campanoni and Nick 2005), delay of leaf abscission (Reid 1985), regulation of floral bud development (Okada *et al.* 1991), promotion of fruit development (Taiz and Zeiger 1998), induction of vascular differentiation (Aloni 1995), and the induction of adventitious and lateral root growth (Celenza *et al.* 1995). Auxin responses in roots and shoots also display a bell shaped dose-response curve, an effect that has become a hallmark of IAA action. However, the curve is shifted to the right in shoots compared to roots, confirming that roots are more sensitive to IAA than shoots and that auxin can have a stimulatory or inhibitory effect depending on its concentration (Knee and Hangarter 1996). In addition, auxin sensitivity can be different within the same organ. For example, primary root growth is inhibited by higher levels of auxin, whereas the same auxin concentration will have a stimulatory effect on lateral root growth (Taiz and Zeiger 1998).

IAA and root development

Because induction of lateral root growth is of particular interest in this study, the role of auxin in this physiological process will be further discussed. Lateral root development occurs when quiescent cells in the pericycle dedifferentiate and proliferate in response to IAA stimulus (Reed *et al.* 1998). This results in the formation of a lateral root primordia in which the cells continue to differentiate and elongate until the newly formed root pushes through the epidermal wall of the primary root. As lateral roots mature they become capable of producing more lateral roots and are structurally very similar to primary roots (Blakely *et al.* 1982).

Proof of the role auxin plays in lateral root development has been provided by studies in which auxin or auxin inhibitors are applied exogenously. Additionally, many

genetic studies have been done using mutant plants with modified endogenous auxin production or sensitivity. For example, Muday and Haworth (1994) demonstrated that exogenous IAA application to growing plants stimulates lateral root initiation and elongation. The same study was also used to demonstrate that exogenous application of naphthylphthalamic acid (NPA), an auxin transport inhibitor, caused a reduction in the number of lateral roots compared to untreated plants. Both Reed *et al.* (1998) and Casimiro *et al.* (2001) have also shown that inhibition of IAA transport from plant shoots inhibits lateral root formation in Arabidopsis. Mutant studies demonstrate the same concepts as exogenous application, but on a genetic level. One such study utilized a *rooty* mutant, a mutant with increased auxin production, to show that lateral root formation is increased when endogenous levels of auxin are increased (Boerjan *et al.* 1995). Other studies have utilized auxin resistant mutants (*axr1, aux1* and *axr4*) to show that a decreased ability to respond to auxin leads to a decrease in lateral root formation (Hobbie and Estelle 1995).

Auxin Signaling

The role of IAA as a signaling mechanism for many diverse and crucial elements of plant development has been the subject of much research and several theories have emerged as to how the pathway is orchestrated. Auxin binding proteins have been postulated as well as auxin response elements (AuxRE), transcription factors, and primary response genes.

Auxin binding protein (ABP1)

Classic hormone response pathways generally begin with a receptor that binds the hormone and tranduces the hormone stimulus into a response. Elucidating auxin receptors has proven difficult. In spite of this, progress has been made in identifying a candidate receptor, ABP1 (Timpte 2001). ABP1 was first purified more than twenty years ago when it was shown to have the ability to bind to auxin (Jones 1994, Klambt 1990). Since then, ABP1 has been identified in many plant species (Jones 1994) and has also been shown to be a crucial element in many of the same responses as IAA, such as

embryogenesis (Chen *et al.* 2001a), cell expansion (Jones *et al.* 1998) and cell division (Chen *et al.* 2001b). Since its discovery, a model of ABP1 localization and action has been postulated. ABP1 is thought to be localized mostly in the endoplasmic reticulum (ER) but it is also detectable in the Golgi and plasma membrane (PM). ABP1 is thought to either associate with a docking protein in the PM or to interact directly with ion channels. Upon auxin binding, a conformational change is probably induced which initiates the cascade of events necessary to transduce the auxin signal to an eventual physiological response (Timpte 2001). In addition to ABP1, another auxin binding protein has recently been studied, (TIR1) and will be discussed briefly.

Transcription factors

ARFs (auxin response factors)

Auxin response factors (ARFs) and auxin/indole-3-acetic acid (Aux/IAA) proteins (Fig. 2.8), two related protein families, have been shown to be key regulators that work together to mediate expression of auxin-induced genes (Guilfoyle *et al.* 1998). The ARF proteins represent an extended family of transcription factors, of which there are 23 in Arabidopsis, that have a DNA binding domain with the ability to bind to the auxin-response element (AuxRE) TGTCTC in promoters of auxin-regulated genes (Ulmasov *et al.* 1997a). The ARF protein family is generally associated with activation of target auxin responsive genes. However, several ARF proteins inhibit target gene expression. Studies with ARF mutants (*arf*5, *arf*6, *arf*7 and *arf*8) have shown that transcription is generally activated by ARF proteins with Q-rich (glutamine-rich) middle regions (Ulmasov *et al.* 1999b), whereas ARF proteins with a P/S/T-rich middle region tend to inhibit transcription, as shown with the *arf*1 mutant. Therefore, whether a given ARF protein activates or inhibits expression is said to be determined by the sequence and

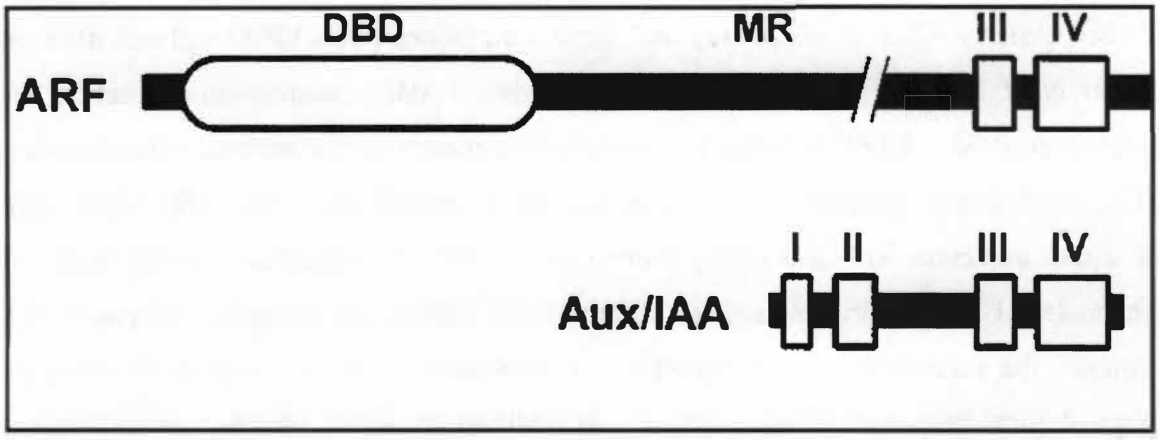

Figure 2.8: Simplified domain properties of ARF and AUX/IAA proteins
From Liscum and Reed (2002)

resulting structure of the protein's middle region (Ulmasov *et al.* 1997c). In addition to a conserved DNA binding domain, ARF proteins also have two conserved regions, III and IV, which are identical to two conserved regions characteristic of Aux/IAA proteins. These identical conserved regions facilitate dimerization between the two proteins (Liscum and Reed 2002).

Aux/IAA

Proteins in the Aux/IAA family, of which there are at least 29 in Arabidopsis (Liscum and Reed 2002), are thought to function as repressors of auxin-regulated genes (Abel *et al.* 1994). Generally, Aux/IAA proteins have four amino-acid-sequence motifs known as domains I-IV (Reed 2001). Domain I is possibly involved in homo-dimerization of Aux/IAA proteins (Hagen and Guilfoyle 2002), while domain II affects stability (Reed 2001). Domains III and IV are thought to mediate hetero-dimerization between Aux/IAA and ARF proteins because the domains are identical between the two protein families (Reed 2001, Guilfoyle *et al.* 1998).

ARF and Aux/IAA interaction

In addition to the information provided previously, it also noteworthy that Aux/IAA proteins are themselves auxin-induced and there is no evidence that the same is

true for ARF proteins (Abel *et al.* 1995). This fact has been useful in piecing together the following model (Fig. 2.9) describing the interaction between the two protein families. It has been proposed that when auxin levels are low, auxin response genes are generally repressed (Hagen and Guilfoyle 2002). Repression takes place when ARF proteins are dimerized, via domains III and IV, with Aux/IAA proteins (Liscum *et al.* 2002). Dimerization between the two proteins probably occurs while ARF proteins are bound to AuxRE (Hagen and Guilfoyle 2002) and may actually involve higher orders of multimers (Morgan *et al.* 1999). Dissociation of Aux/IAA proteins from ARF proteins occurs when auxin levels are increased (Ulmasov *et al.* 1997b) and auxin binds directly to a recently characterized auxin binding protein, TIR1, (Kepinski and Leyser 2005, Dharmasiri *et al.* 2005). Direct auxin and TIR1 binding promotes interaction between TIR1 and Aux/IAA proteins, which functions to target Aux/IAA proteins for degradation in the ubiquitin/proteosome pathway (Kepinski and Leyser 2005). Upon dissociation and degradation of Aux/IAA, ARF proteins are derepressed (Gray and Estelle 2000). This allows transcription of auxin regulated genes, including Aux/IAA genes (Casson and Lindsey 2003). Transcription may be further activated by binding of additional ARF transcriptional activators when Aux/IAA proteins dissociate (Hagen and Guilfoyle 2002) Transcription is repressed again when auxin levels fall and a negative feedback loop is initiated in which Aux/IAA proteins are not degraded and can once again dimerize with ARF proteins (Casson and Lindsey 2003).

Primary response genes

Transcripts from at least three gene families are induced rapidly and transiently in response to auxin. These gene families include the previously described family of Aux/IAA proteins, Small Auxin-Up RNAs (SAUR) and GH3-related transcripts (Woodward and Bartel 2005). Levels of SAUR transcripts are rapidly up-regulated in many plant species including soybean (Walker and Key 1982) and Arabidopsis (Gil *et al* 1994). In some cases, they have been shown to be up-regulated within three to five minutes of auxin treatment (McClure *et. al.* 1989). SAUR transcripts are characteristically rapidly degraded and mutants that produce stable SAUR RNA have no

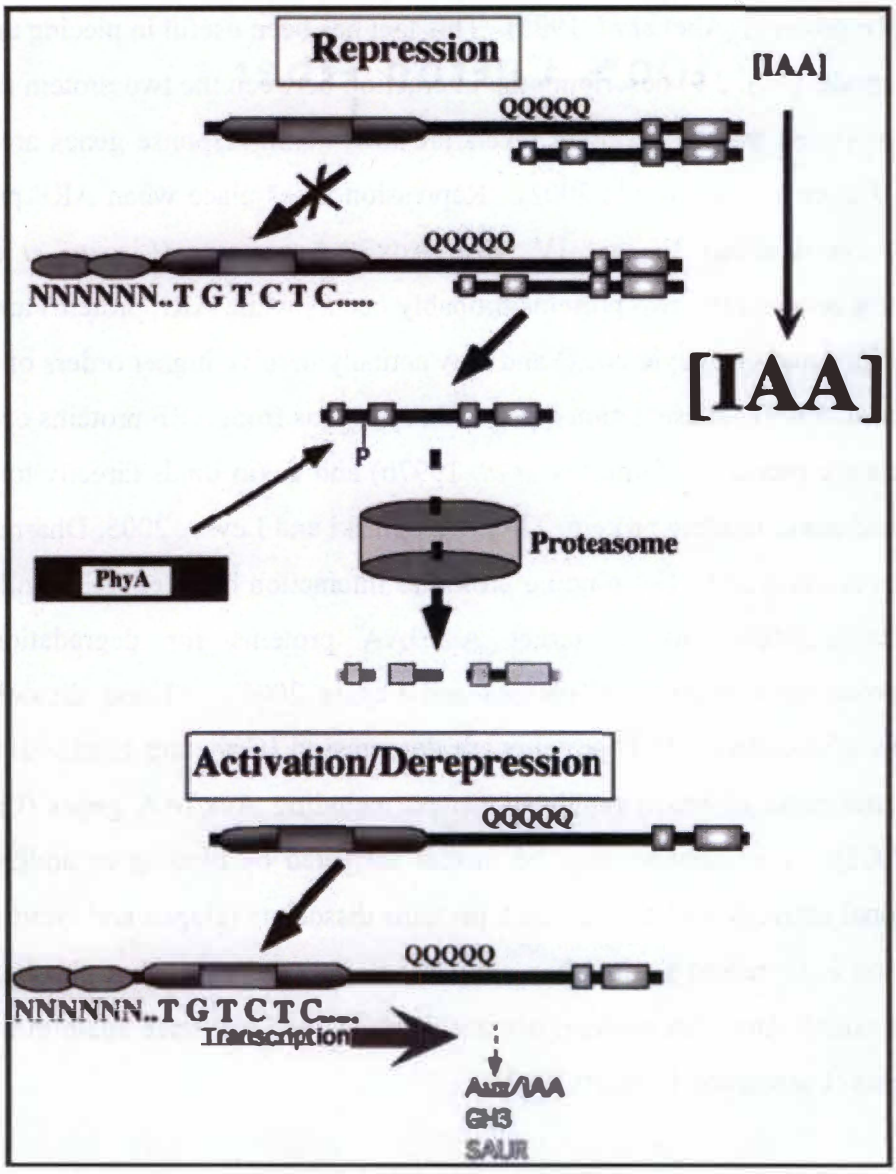

Figure 2.9: Model of Aux/IAA and ARF protein interaction
From Hagen and Guilfoyle (2002)

noticeable morphological phenotype (Johnson *et al.* 2000). Therefore, the function of SAUR small RNAs remains a mystery (Woodward and Bartel 2005). GH3 transcript accumulation in response to auxin has, like SAUR transcripts, also been observed in several plant species. Accumulation can occur as quickly as 5 minutes in *Glycine max* (Hagen and Martin 1991). GH3 genes up-regulated in response to auxin encode IAA-amino acid conjugating enzymes (Staswick *et al.* 2005) and are therefore thought to facilitate dampening of auxin signaling by inactivating IAA via conjugation (Woodward and Bartel 2005).

Chapter 3 : EST Analysis of Gene Expression

Introduction

Populus

Members of the *Populus* genus are included in the family Salicaceae (Bremer *et al.* 1998) and consist of deciduous trees such as aspen and cottonwood (Bradshaw *et al.* 2000). Several members of the genus are commercially important and serve a variety of purposes in the pulpwood and lumber industry (Balatinecz and Kretschmann 2001). In addition, poplars have been used as an alternative energy source (Isebrands 1979) and are known to have a tremendous ability to sequester carbon in their root systems (Pregitzer and Friend 1996). Therefore, they are also potentially important environmental resources as well. On a molecular level, *Populus* is filling the role of model tree species because its numerous attributes including a relatively small sequenced genome (Bradshaw and Stettler 1993), abundant genetic variation (Taylor 2002), ease of clonal (Dickmann and Stuart 1983) and sexual propagation (Zsuffa *et al* 1996), and ability to successfully transform (Jouanin *et* al 1993) have led to the development of a wealth of accessible genomic tools and databases (Yin *et al.* 2002, Wullschleger *et al.* 2002a, Sterky *et al.* 1998). The available tools can be used to understand how individual genes control growth and development, eventually leading to the ability to manipulate these controls.

Objectives

One particular area of interest is the ability of poplar to sequester carbon in lateral roots and, more specifically, how root growth can be manipulated to increase carbon sequestration ability. A well known and essential component of lateral root growth induction is the plant hormone auxin (Muday and Haworth 1994, Reed *et al.* 1998, Casimiro *et al.* 2001). Thus, in order to better understand the molecular controls of lateral root growth, the effects of auxin on gene expression need to be studied. Here, H11-11 (*P. trichocarpa* x *P. deltoides*), a poplar hybrid, was used to analyze differential gene expression between control trees and trees treated with exogenous auxin application. The goal of the study was to use comparative genomics tools to gain a general understanding of downstream regulators of auxin response on a whole-plant level.

Additionally, the results of the study were used to add to the existing poplar EST database, in which approximately 25% of poplar genes are not represented.

There are several methods by which differential gene expression can be studied, including expressed sequence tag (EST) sequencing, differential display, microarrays, and serial analysis of gene expression (SAGE) (Caruli *et al.* 1998). These techniques can be used in high throughput gene expression analysis and have advantages and disadvantages associated with them (Table 3.1).

In this study, a subtraction cloning technique, suppression subtractive hybridization (SSH), was used to create two EST libraries. One library contained ESTs that are up-regulated in response to exogenous auxin treatment and the second contained ESTs that are down-regulated in response to exogenous auxin treatment. The cDNA libraries were amplified with rolling circle amplification (RCA), sequenced and comparative genomic tools were used to characterize the differences between the two libraries. Leaves, stems and root tissues were pooled from the treated plants to create both libraries. This analysis is meant to give a snapshot of whole-plant trends, rather than tissue-specific trends that occur in response to exogenous auxin treatments.

EST Sequencing

ESTs are derived from cDNA libraries generated by reverse transcription of cell-specific mRNA populations that are cloned into plasmids (Bonaldo *et al.* 1996). A specified number of clones are then randomly picked and sequenced. The resulting ESTs represent unique sequence tags for particular transcripts (Caruli *et al.* 1998) that are used to search databases to find information about genes expressed in a given tissue (Yamamoto *et al.* 1997). This technique has been used to study gene expression in several plant species including rice (*Oryza sativa* L.) (Yamamoto and Sasaki 1997), sugarcane (*Saccharum* spp. hybrids) (Carson and Botha 2000), leafy spurge (*Euphorbia esula* L.) (Anderson and Horvath 2001) and *Populus* (Kohler *et al.* 2003, Park *et al.* 2004). However, large-scale EST sequencing can be costly and is hindered by the presence of highly redundant mRNAs produced by "housekeeping genes" that are not necessarily responding to a treatment or tissue of interest. One way to avoid repetitive

Table 3.1: Advantages and disadvantages of high-throughput gene expression analysis techniques

Technique	Advantages	Disadvantages
ESTs	Cost effective (Mayer and Mewes 2001) Large gene discovery tool is available (Wullschleger *et al.* 2002b) Can be used to design microarray experiments (Mayer and Mewes 2001) Technically simple (Caruli *et al.* 1998)	Possible unreliable quality and partial nature of sequence (Mayer and Mewes 2001) Overrepresentation of highly expressed genes (Mayer and Mewes 2001) Scarce genes are not detected (Mayer and Mewes 2001)
Differential Display	Provides quick assessment (Matz *et al.* 1998)	High sensitivity generates false positives (Caruli *et al.* 1998)
Microarrays	Can be used to investigate global transcription patterns quickly (Stein and Liang 2002) Can perform multiple assays on the same array (Caruli *et al.* 1998)	Technically demanding and expensive (Stein and Liang 2002) Limited to genes of known sequence (Caruli *et al.* 1998) Requires a large amount of RNA for probes (Caruli *et al.* 1998)
SAGE	Very high throughput (Caruli *et al.* 1998) Ability to compare SAGE tag data from a variety of samples (Caruli et al. 1998)	Technically difficult to generate good libraries (Caruli *et al.* 1998) Analysis requires highly specialized bioinformatics tools (Caruli *et al.* 1998)

sequencing of these mRNAs is to create an EST library using a subtractive cloning technique (Bonaldo *et al.* 1996).

Subtractive cloning

Subtractive hybridization allows researchers to efficiently obtain clones of genes that are differentially expressed in two populations of mRNA. Subtractive cloning is achieved using a tracer and a driver through a process known as driver excess hybridization to isolate target molecules. A tracer is the nucleic acid from which one wants to isolate differentially expressed sequences known as targets. DNA serves as a good tracer because it is not easily degraded like RNA. A driver is a complimentary nucleic acid that is believed to lack the expressed sequences of interest. RNA serves as a good driver because RNAs that are not removed during the hybrid removal step are easily degraded. The tracer nucleic acid is hybridized to the driver nucleic acid which is present at a much higher concentration. Sequences that are common between the tracer and driver hybridize and are removed along with un-hybridized driver in a subtraction step (Ermolaeva and Sverdlov 1996). The subtraction process is repeated to ensure all common sequences between the tracer and driver have been removed. The remaining nucleic acid is used to make a cDNA library which is sequenced and analyzed as described earlier, but is lacking the repetitive mRNAs of no interest (Sagerstrom *et al.* 1997). Subtractive hybridization can be accomplished using several methods depending upon what has been used as a tracer and driver. However, the basic principle behind each of the methods remains the same (Ermolaeva and Sverdlov 1996).

Subtractive hybridization has been successfully utilized to describe gene expression in many plant species and organisms including gene expression of roots exposed to low temperature in soybean (Liu *et al.* 2000), during abscission in oranges (*Citrus sinensus* L.) (Wu *et al.* 2003) and during cercosporin biosynthesis in *Cercospora zeae-maydis*, the fungus that causes gray leaf spot (Shim and Dunkle 2002). Despite the success that has been achieved using the method, there is at least one disadvantage associated with it. For example, traditional subtraction techniques are not well suited for identifying rare mRNAs because they require several rounds of hybridization. The

particular method used in this study, Clonetech PCR-Select™ cDNA Subtraction, seeks to overcome the limitations of traditional methods using a suppression PCR approach, hence the name suppression subtractive hybridization (SSH). This method is carried out in such a way that enrichment of target molecules, including those that are rare, can proceed while amplification of undesirable molecules is prevented (Clonetech 2002). The process begins by preparing tester and driver double stranded cDNA from the two mRNA samples being compared (Fig. 3.1) which are hybridized and subjected to PCR in a multi-step process.

Rolling Circle Amplification (RCA)

Before cDNA sequencing can proceed, the plasmid templates containing EST inserts must be amplified and purified to prepare them for use in sequencing reactions. This can be a time consuming and expensive procedure that traditionally involves incubating cultures in LB media overnight before lysing and purifying the products with expensive plasmid prep kits. In an attempt to minimize the hands-on time and high costs associated with traditional amplification methodology, rolling circle amplification (RCA) was tested as a viable option for use in the amplification and purification process.

The process of RCA (Fig. 3.2) utilizes Phi29 polymerase, the polymerase used to replicate circular DNA molecules in some organisms, and a random hexamer primer with protected ends to keep it from being degraded by the polymerase (Dean *et al.* 2001). Phi-29 polymerase is an enzyme with DNA polymerase and 3'-5' exonuclease activities that has the ability to catalyze the formation of the initiation complex needed in order for DNA replication to proceed. The enzyme is also very stable and capable of performing strand displacement DNA synthesis at a rate of 50 nucleotides per second for more than 70,000 nt without dissociating from the template. This allows efficient DNA synthesis to occur for many hours (Blanco *et al.* 1989). RCA is initiated when random hexamer primers bind at multiple locations on the denatured circular template and are extended by Phi29 polymerase. The extending hexamers are eventually displaced at their 5' ends and, as polymerization and strand displacement continues, single-stranded tandemly repeated complimentary copies of the template strand are synthesized. In addition to the initiation

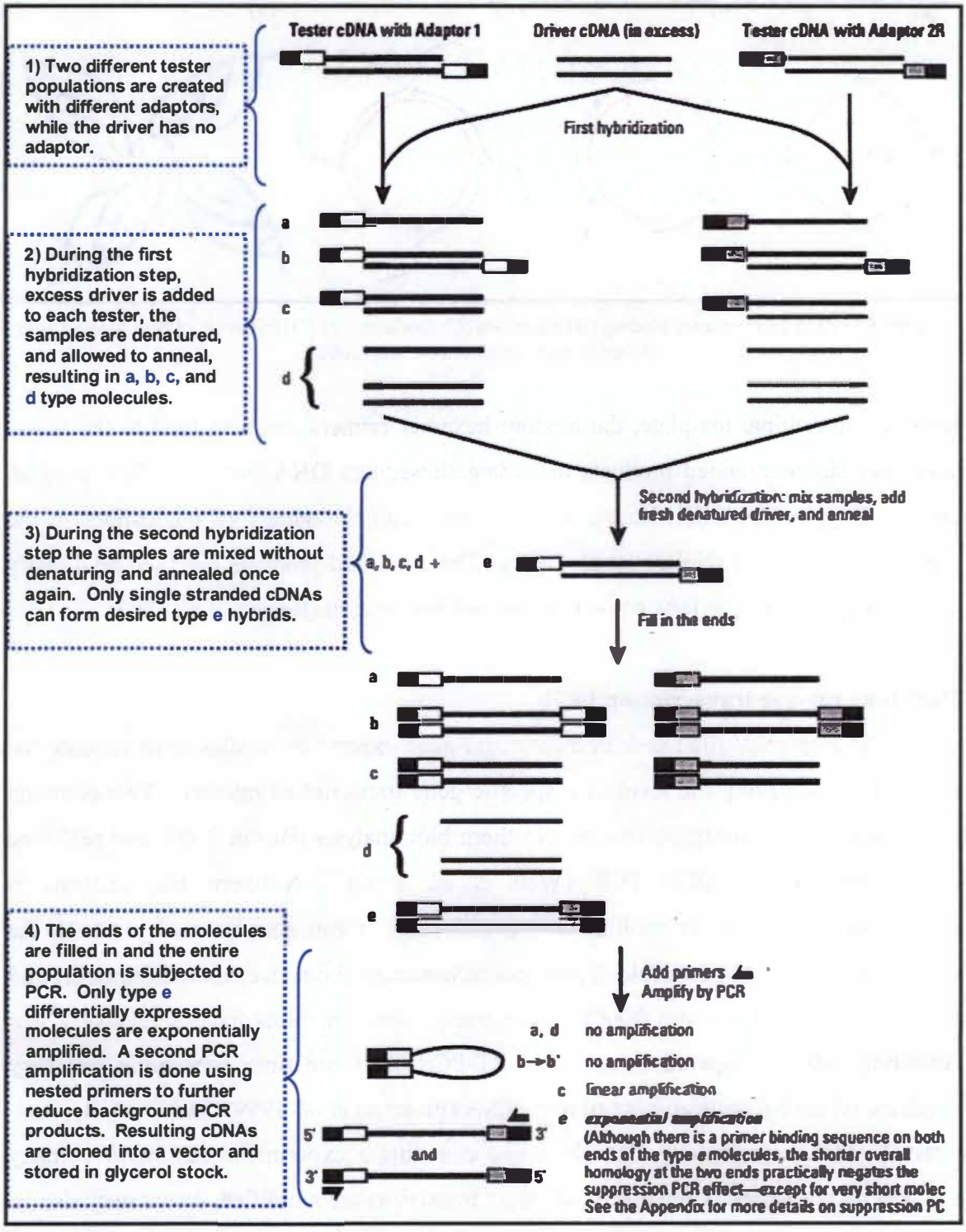

Figure 3.1: Steps in Clonetech PCR™ cDNA Subtraction
Adapted from Clonetech (2002)

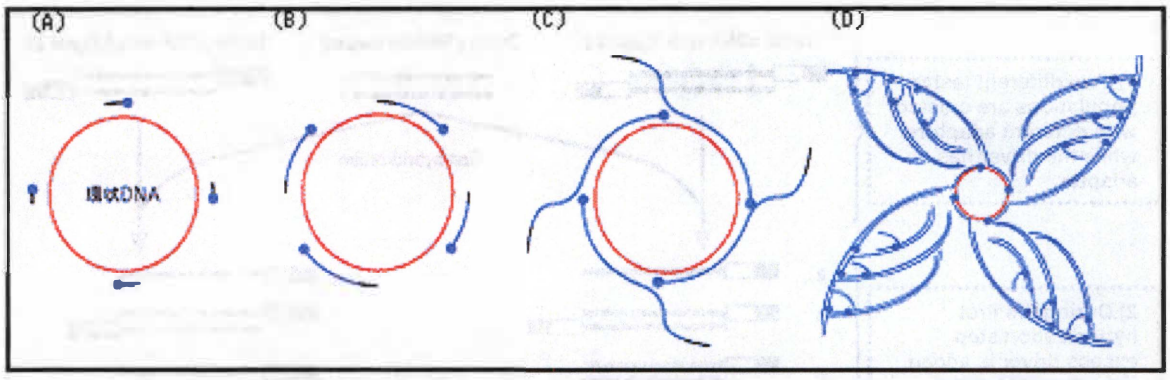

Figure 3.2: RCA (A) hexamer binding (B) extension (C) displacement (D) initiation of new binding sites Adapted from Patki and Nelson (2002)

sites on the original template, the random hexamer primers can now bind to the newly generated single-stranded products to initiate subsequent DNA synthesis (Reagin *et al.* 2003). Exponential amplification will continue until the supply of nucleotides in the reaction is exhausted (Nelson *et al.* 2002). The amplified products can then be directly used in sequencing reactions without additional purification (Reagin *et al.* 2003).

Real-time reverse transcription PCR

An important final task in differential gene expression studies is to validate the results by quantifying the level of a specific gene transcript of interest. Two common methods used to accomplish this are Northern blot analysis (Bustin 2000) and real-time reverse transcription (RT) PCR (Weis *et al.* 1992). Northern blot analysis is advantageous because, in addition to quantification, it can also be used to check the quality of mRNA. For example, it provides information about size, alternate splicing and integrity of mRNA (Bustin 2000). Conversely, Northern blots have a relatively low sensitivity when compared to real-time RT-PCR, they are time consuming and they require a relatively large amount of total RNA (Freeman *et al.* 1999). Additionally, real-time RT-PCR is relatively flexible because a single experiment can involve many samples and/or genes. Conversely, as target transcripts are amplified, errors may also be greatly amplified to a point where they reduce the reliability of quantification results. However, precautions such as careful experimental design, meticulous lab technique and validation of the method for a specific gene of interest minimize errors (Freeman *et al.*

1999). Because real-time RT-PCR can be used to quantify very small amounts of RNA it was used to validate differential expression in this study.

Real-time RT-PCR is a relatively complex assay involving several steps. The first of which is treating the RNA sample of interest with DNase to rid the sample of any residual DNA contamination. This step is important because DNA contamination will lead to inaccurate quantification during the final steps of the assay (Bustin 2002). The second step involves conversion of the RNA template of interest to cDNA using a reverse transcriptase enzyme. This is accomplished through a process known as reverse transcription and it must be done because RNA cannot serve as template for PCR (Bustin 2000). Both random hexamer and oligo-dT primers are commonly used to prime cDNA synthesis because they bind non-specifically, thereby maximizing the number of genes that can be assayed in an RNA pool. However, oligo-dTs are thought to give a more accurate representation of transcript abundance because they anneal to the polyadenylated 3' tails characteristic of mRNA. This ensures that the resulting cDNA represents only mRNA from the total RNA pool (Freeman *et al.* 1999).

During the third step, the newly synthesized cDNA sample is amplified in a real-time PCR reaction. Real-time PCR differs from conventional PCR in that amplified product is measured during each cycle throughout the PCR reaction by a video camera rather than at a final end point. The video camera records light emitted by a fluorochrome that is only detected when it has been incorporated into the newly synthesized PCR product (Gachon *et al.* 2004). The recorded fluorescence values represent the amount of detectable amplified product at a given point in the amplification process. The cycle at which fluorescence becomes detectable is known as the threshold cycle (C_t). The amount of template of interest present at the beginning of the reaction determines the cycle number in which the amplified product eventually reaches a detectable level. For example, the higher the amount of transcript template present at the beginning of the reaction, the lower the number of cycles required to amplify that template to a detectable level. After the reaction is complete, the C_t values for each transcript of interest are normalized and compared to see if differential expression is, in fact, taking place (Bustin 2000).

Along with reagents commonly used in PCR reactions such as a hot-start Taq DNA polymerase, gene specific primers, dNTPs and Mg^{2+}, DNA-binding dye is also needed to facilitate product detection by a video camera. DNA binding dyes bind directly to double stranded cDNA during the extension phase of PCR. While unbound, they display little fluorescence and are not detected. When bound, fluorescence is detected resulting in increasing fluorescence levels as the target product is exponentially amplified. During the denaturing phase of PCR the dye falls off to, once again, become undetectable (Bustin 2000). In the case of this experiment, a DNA binding dye known as SYBR® Green™ was used.

During the fourth step of real-time RT-PCR, the resulting C_t values are normalized and analyzed to see if differential expression has occurred. The general method of data normalization requires amplification of a reference transcript along with the transcript of interest. The reference transcript must be one that is expressed at a constant level among different tissues, developmental stages and treatments. The C_t values for the reference transcript can then be averaged and used to normalize the rest of the C_t values in the experiment (Karge *et al.* 1998). The method used to analyze the data depends upon whether exact or relative transcript quantity is being compared. If a precise transcript quantity is desired, a standard of known quantity needs to be used during the experiment to generate a standard curve. This standard curve is then used to determine the amount of transcript of interest. In contrast, if relative expression level is of interest a standard curve does not need to be generated and normalized values can simply be compared (Freeman *et al.* 1999). As a general rule the equation 2^N, with N being equal to the difference in cycles between two samples, is used to calculate fold-change. For example, if one sample has a C^t value of 20 and another sample has a C^t value of 23, the fold change is 2^3 or 8 between the two samples (Schmittgen *et al.* 2000). For this study, relative expression level comparisons were made rather than exact transcript quantification comparisons.

Materials and Methods

Plant material, experimental design and treatment

Poplar hybrid H11-11 (*P. trichocarpa* x *P. deltoides*) trees were grown, treated and sampled at the University of Florida, Gainesville, FL. Plants were clonally propagated from rooted softwood cuttings and were grown in a controlled greenhouse environment for about seven weeks. A completely randomized design (CRD) was used and plants were treated when the average plant height was 80cm in September 2003. Five plants were foliar sprayed with approximately 50mL of 100μM IAA dissolved in $5x10^{-04}$ N NaOH and five control plants were foliar sprayed with approximately 50mL of $5x10^{-04}$ N NaOH. The plants received the same respective treatment for nine days. At the end of the treatment period, leaves, roots, and stems were harvested and frozen in liquid nitrogen. Phenotypes were not evident on the IAA treated plants at the time of harvest. This may have been the case because it was too early after treatment for auxin induced phenotypes to have occurred. Another possible explanation is that the H11-11 hybrid may not absorb exogenous IAA efficiently.

mRNA isolation, SSH, and cDNA library construction

Isolation of mRNA, SSH, and cDNA library construction tasks were performed at the University of Florida. Tissues of the same type were bulked across plants within the same treatment, pulverized with a mortar and pestle, and total RNA was extracted for each tissue type. Equal amounts of total RNA were bulked across tissue types within treatments resulting in a mixture with equal amounts of leaf, root, and stem total RNA. Poly(A)+RNA (mRNA) was then isolated from the two bulked RNA samples and reverse transcription was used to create cDNA populations.

SSH was performed on the mRNA according to the Clonetech PCR-Select™ cDNA Subtraction protocol manual (Clonetech 2002). Two subtractions were done resulting in two libraries, one with genes up-regulated in response to auxin and one with genes down-regulated in response to auxin. In the library containing differentially expressed genes that are up-regulated in response to auxin, the control sample served as the driver and the auxin treated sample served as the tracer. In the library containing

differentially expressed genes that are down-regulated in response to auxin, the treated sample served as the driver and the control sample served as the tracer. The subtracted products were shotgun cloned into pGEM-T vectors (Fig. 3.3) according to the TOPO TA Cloning® Kit protocol (Invitrogen 2004), transformed into a DH5a *E. coli* cells and plated on X-gal/IPTG. Blue/white colony screens were performed to select successfully transformed colonies. Single colonies were picked and grown for approximately 7 hours. These were used create 500µl overnight cultures of glycerol stocks stored in 96 well plates. Ten 96-well plates were cultured for each library resulting in 960 up-regulated ESTs and 960 down-regulated ESTs These were deep frozen (-80°C) and sent to ORNL on dry ice for amplification, sequencing and subsequent analysis.

cDNA amplification

Two amplification methods were evaluated based on hands-on time, cost and product quality. The first method involved amplification and purification according to the Montage™ Plasmid Miniprep$_{96}$ Kit User Guide (Millipore 2000) with a Beckman Coulter Biomek® 2000 Laboratory Automation Workstation. The only deviation from the kit protocol was that cultures were grown overnight in 2X YT (Sambrook *et al.* 1989) rather than 2X LB. Five 96-well plates were used to test this method.

The second method, RCA, was adapted and standardized from a Fidelity Systems RCA protocol (Fidelity Systems 2004) according to the following specifications. Five µl reactions were prepared consisting of 2.5µl 2x annealing buffer (80mM Tris-HCL, pH 8.0; 20mM $MgCl_2$), 0.25µl Phi29 random hexamer primer, 1µl undiluted glycerol stock (template cDNA) and 1.25µl of sterile deionized water. This was heated for 3 minutes at

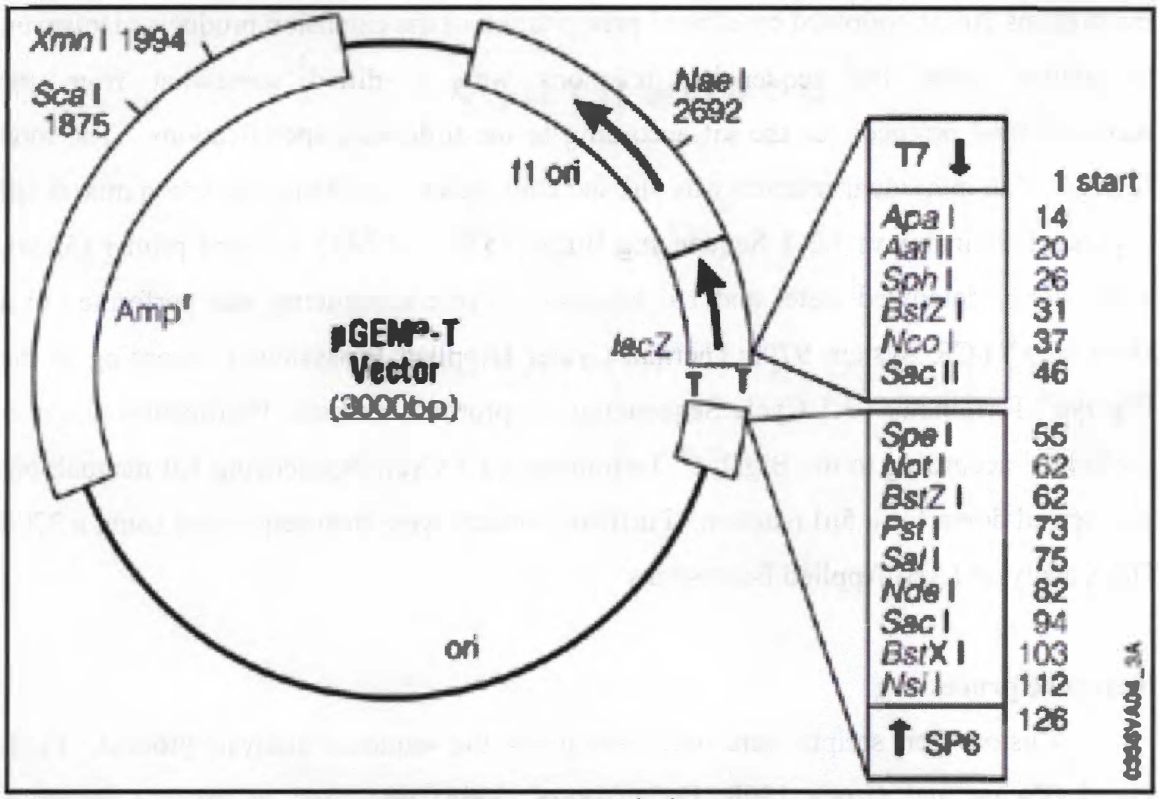

Figure 3.3: pGEM-T cloning vector
From Promega (2005)

94°C, cooled and the following 15µl mixture was added: 2µl of Phi29 10X buffer (500mM Tris-HCL, 100mM $(NH_4)_2SO_4$, 100mM $MgCl_2$, 40mM dithiothreitol, pH 7.5 at 25°C), 0.175µl of Phi29 DNA polymerase, 3µl of dNATPs (4 mM) and 9.825µl of sterile deionized water. The 20µl mixture was vortexed, spun briefly and incubated at 30°C overnight. The polymerase was then inactivated at 65°C for 10 minutes and kept at 4°C until needed for sequencing. The Phi29 random hexamer primer was purchased from Fidelity Systems, Inc. The Phi29 polymerase was purchased from New England BioLabs (New England BioLab 2004). Eleven 96-well plates were used to test this method.

Sequencing

The sequencing reactions were done on template directly from RCA product, as a purification step is not necessary. cDNA sequencing was performed by dye-terminator cycle sequencing using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied

Biosystems 2002), followed by ethanol precipitation of the extension products. However, to reduce costs, the sequencing reactions were modified somewhat from the recommended protocol for the kit according to the following specifications. The total volume of an individual reaction was 5µl and consisted of 1µl Ready Reaction mix, 0.5µl BigDye Terminator v1.1/3.1 Sequencing Buffer (5X), 1µl M13 forward primer (5µM), 1.5µl sterile deionized water and 1µl template. Cycle sequencing was performed in a GeneAmp® PCR System 9700 Thermal Cycler (Applied Biosystems) according to the BigDye® Terminator v3.1 Cycle Sequencing Kit protocol manual. Purification was also performed according to the BigDye® Terminator v3.1 Cycle Sequencing Kit manual, but was scaled down for a 5µl reaction. Purified products were then sequenced using a 3700 DNA analyzer (ABI Applied Biosystems).

Sequence processing

Custom Perl scripts were used throughout the sequence analysis process. First, Phred (Ewing and Green 1998, Ewing *et al.* 1998) was used to process sequence chromatograms to generate two Fasta files per EST. The first Fasta file contained EST sequence with vector sequence identified. The second Fasta file contained quality scores for each base. These Fasta files were modified for use in StackPack (Christoffels *et al.* 2001), which was used to cluster ESTs into non-redundant consensus sequences according to default criteria set by StackPack. Basic local alignment search tools (BLAST), a method used to rapidly search nucleotide and protein databases for regions of similarity, were used to further process the ESTs (Altschul *et al.* 1990). To identify ESTs that were not represented in the current poplar assembly, ESTs were BLASTed against the poplar unigene database. The top hit for each EST (the hit with the lowest e-value) was used for further analysis. A BLASTn was done with consensus sequences and singletons against the poplar gene model database so that subsequent BLASTs could be done using more complete gene model sequences rather than shorter ESTs. The top hit for each EST to the poplar gene model database was then used to perform a BLASTn against the Arabidopsis peptide database. The results were parsed and the top hit for each EST to the Arabidopsis peptide database was used to do a comparative genomic analysis

of the ESTs. Those ESTs that did not have a hit to the gene model database were BLASTed directly against the Arabidopsis peptide database and, once again, the top hit was used for further analysis. LUCY (Chou and Holmes 2001) was used to screen out empty vector sequence and report three lists of ESTs that did not have at least 100bp with an average quality score of 14, 16, or 20. Generally, a single base with a Phred score of 20 or less is considered to be an ambiguous base. These three lists were then used to look at each sequence individually to decide whether to keep it in the analysis based on the LUCY results and BLASTn results. Sequences were retained even if they did not pass the LUCY criteria if they appeared to have a legitimate BLAST hit. This was done to ensure that as many sequences as possible were retained for further analysis. Those that were removed were not included in any further analysis.

In order to further purge the EST libraries of redundant sequence because SSH is not 100% efficient, an electronic hybridization was performed. This was done by removing an EST in one library if it had the same genemodel hit in the other library. For example, if one library had three hits to a gene model and the other library had six hits to the same gene-model, three were removed from each of the libraries. This resulted in an exclusive list of ESTs up-regulated in response to auxin, an exclusive list of ESTs down-regulated in response to auxin and a list shared ESTs that were removed from each library.

Randomization test to determine SSH efficiency

A sampled randomization test was done to determine the efficiency of the SSH method used and, thus, the reliability of the resulting EST analysis. A randomization test consists of three steps. First, an observed sample outcome is considered to be one of many possible and equally likely outcomes to arise by chance. Second, the possible outcomes that could be obtained by randomly rearranging frequencies are enumerated. Third, a distribution of the resulting outcomes is used to determine whether the single observed outcome is deviant enough to reject the hull hypothesis. A sampled randomization test, rather than an exact randomization test, is used when all possible outcomes cannot be determined (Sokal and Rohlf 2000).

To do the sampled randomization test, a series of steps were performed using customized Perl scripts. First, 66,201 non-normalized (unsubtracted) ESTs from Genbank were BLASTed against the Joint Genome Institute (JGI) gene set of 54,000 poplar gene models. This was done to establish the pool from which ESTs were randomly drawn to create hypothetical libraries. 896 ESTs from the pool were randomly sampled from this list and compared to another 839 randomly sampled ESTs from the same list. 896 and 839 were chosen because this represents the same number of ESTs in the two libraries being compared in this study. The number of shared top hits between the two hypothetical libraries was reported. This was repeated 99,999 times to generate a histogram containing the number of sequences that were shared between the two randomly generated libraries each time. A comparison was then made between the generated histogram and the observed number of ESTs that were shared between the up-regulated and down-regulated libraries in this study. The null hypothesis (H_O) is: the number of ESTs shared between the two libraries was not decreased by SSH and is equal to that of randomly generated un-subtracted libraries. The null hypothesis can only be rejected if the actual number of ESTs shared between the two libraries in a one-tailed test is less than 95% of the randomly generated values ($p \leq 0.05$). In other words, if the observed data falls into the lower (left side) 5% range of the histogram, the SSH worked efficiently.

EST analysis

ESTs for each library were put into functional categories using MIPS (Schoof *et al.* 2004). Those that fell into either the "classification not yet clear-cut" or "unclassified protein" categories were used to query TAIR (Rhee *et al.* 2003) to see if a functional category could be assigned using a different database. Three functional categories, transcription factors, metabolism and proteins were then studied more closely to characterize the differences between the two libraries. KEGG (Kanehisa and Goto 2000) was used to characterize and compare metabolic pathway trends in each library.

Real-time reverse transcription PCR

RNA extraction

RNA was extracted from six tissue samples using a Qiagen RNeasy® Plant Mini Kit according to the protocol supplied by the company (Qiagen 2001). Each of the six samples consisted of an evenly homogenized mixture of tissue from the five reps within a particular treatment and tissue type. The six samples included control and treated stem tissue, control and treated leaf tissue and control and treated root tissue. After the initial RNA extraction each sample was concentrated using a Qiagen RNeasy ® MinElute™ Cleanup Kit according to the protocol supplied by the company (Qiagen 2003). RNA quantity was measured using a using a Nano-drop® ND-1000 Spectrophotometer.

cDNA synthesis

One µg of total RNA from each sample was treated with DNase 1 (two units/µg) at 37°C for ten minutes to remove DNA contamination before cDNA synthesis. DNase stop solution was added and samples were incubated according to instructions provided by DNase usage information (Promega 2005). One µg of DNase treated RNA was used for cDNA synthesis (7.7µl). SuperScript ™ III Reverse Transcriptase was used to synthesize cDNA according to the instructions supplied by Invitrogen ™ (Invitrogen 2003). Oligo dT primer was used rather than random hexamer primer so that the resulting cDNA only represented mRNA from the total RNA population. Reverse transcription was performed on an Applied Biosystems Gene Amp® PCR System 9700. After cDNA synthesis each sample was diluted 1:1 with RNase and DNase free H_2O.

Real-time PCR

Real-time PCR 25µl reactions were done using iQ™ SYBR® Green Supermix according to instructions provided by Bio-Rad Laboratories on an iCycler Real Time PCR detection system (Bio-Rad Laboratories 2005). One µl of cDNA sample was used in each reaction. Gene specific forward and reverse primers were designed based upon gene model sequence of two genes of interest. The forward and reverse primers used to detect the transcripts in question were as follows:

peroxidase (F), 5'-GGTGTTGTTATTGCTAGTCAGTGG-3'

peroxidase (R), 5'-CATGGATCATCACACACCATC-3'

chlorophyll a/b binding protein (F), 5'-GAGAACCTCTTGCAGCACATC-3'

chlorophyll a/b binding protein (R), 5'-GTGAATCATAATGCCACTTGTTCC-3'

The first was a putative peroxidase (PER) gene found to be prevalent in the down-regulated library in response to IAA treatment. The second was a gene encoding a light-harvesting chlorophyll a/b binding protein (CHL) up-regulated in response to IAA treatments. The gene used as a control to normalize the data for differences in input RNA and efficiency of reverse transcription between the samples was an actin gene (ACT) expressed at a constant rate in all of the tissue types and treatments. Ten uM each of the appropriate forward and reverse primer were added to each reaction. Three reps were done for each reaction to give a total reaction number of 54 (6 samples x 3 genes x 3 reps = 54).

Data analysis

C_t values of the three reps were averaged within each sample and treatment type. The rep averages of the ACT controls were then averaged *across all* tissue and treatment types to get an overall ACT average. Average ACT control values within tissue types were subtracted from the control average across tissue types, resulting in a cycle correction value for each tissue. The cycle correction values were then added to each rep average for the genes of interest and normalized C_t values were compared. All of the products were stained with ethidium bromide and run on an agarose gel to make sure the primers were gene specific and produced only one band.

Results

Two subtracted cDNA libraries, one containing cDNAs that are up-regulated in response to exogenous auxin treatment and one containing cDNAs that are down-regulated in response to exogenous auxin treatment, were used in this study. A total of 960 cDNAs were sequenced and processed from each library.

Rolling circle amplification

Two methods, RCA and Montage™ Plasmid Miniprep$_{96}$ Kits (Fig. 3.4), were used to amplify cDNAs in both libraries and were compared to determine which process is more efficient and cost effective. After taking into account the cost of a Montage™ kit and additional materials needed as well as the cost of reagents and materials needed for RCA, it was determined that the use of a traditional Montage™ kit is more than 4 times as expensive as the modified RCA protocol used in this study. More specifically, the cost to amplify one 96-well plate of cDNAs using a kit is approximately $120.00, while the cost to amplify the equivalent with RCA is approximately $29.73.

Average quality scores of sequences amplified with the two different methods were also recorded. The average Phred quality score for kit amplified cDNAs was 32.29 +/-1.9 whereas the average Phred quality score for RCA amplified cDNAs was 39.97 +/- 0.66. The quality score averages are significantly different according to a two-tailed t-test ($p \leq 0.001$). Therefore, RCA produces a more consistent and higher quality amplified product than the traditional kit. However, it was observed that when bovine serum albumin (BSA) was used in an RCA reaction, as called for in some protocols, the average sequence quality was reduced to 32.47±1.3. The decreased average was significantly different when compared to the average sequence quality of RCA product without BSA ($p \leq 0.001$). This might occur because there is no purification step when using RCA and it is possible that BSA left in the samples decreases sequencing reaction efficiency. Another possibility is that the method used to purify sequencing reaction products did not adequately rid the samples of excessive amounts of BSA protein, causing the 3700 DNA analyzer capillaries to malfunction.

The total amount of time it took to amplify cDNAs with a Montage™ plasmid prep kit and RCA was roughly the same because both methods require an extensive cDNA incubation period (overnight). However, the amount of time spent on bench work is far less when using RCA. This is due, in part, to the fact that a purification step is not required because the RCA process produces linear amplified product that can be used directly in a sequencing reaction.

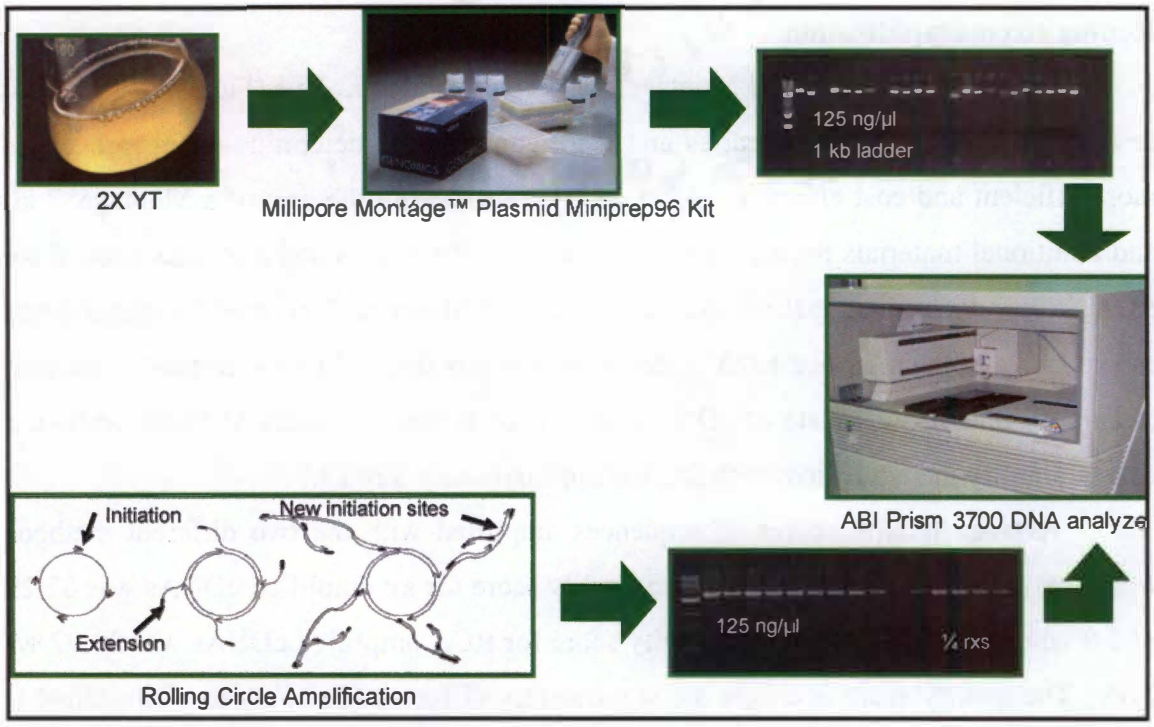

Figure 3.4: Two cDNA amplification techniques, traditional plasmid prep kit (top) and RCA (bottom)

Features of generated ESTs

Up-regulated library features

Of the 960 cDNAs that were sequenced in this library, 78 were either empty vectors or the quality was too low (criteria described in materials and methods) and their BLAST results were removed from the analysis. Removal of the 78 low-quality ESTs and empty vectors resulted in 882 remaining ESTs with an average length of 449bp and an average Phred quality score of 40.5. After clustering the 882 ESTs with StackPack, there were a total of 699 non-redundant ESTs, 224 consensus sequences and 475 singletons, representing a total of 882 ESTs. Of the 699 non-redundant ESTs, 33 (4.7%) were not represented in the poplar assembly database. Electronic hybridization (described in materials and methods) revealed that, of the 882 total number of ESTs in the up-regulated library, 151 (17.1%) were shared with the down-regulated library when comparing hits to the gene-model database. This left 731 ESTs (82.9%) that were exclusive to the up-regulated library. This list of exclusive ESTs was used to characterize the library of up-regulated ESTs.

Down-regulated library features

Of the 960 cDNAs that were sequenced in this library, 53 were either empty vectors or the quality was too low and their BLAST results were removed from the analysis. Removal of the 53 low quality ESTs and empty vectors resulted in 907 remaining ESTs with an average read length of 466bp and an average Phred quality score of 35.9. After clustering with StackPack, there were 622 non-redundant ESTs, 200 consensus sequences and 422 singletons, representing a total of 907 ESTs. Of the 622 non-redundant ESTs, 15 (2.4%) were not represented in the poplar assembly database. Electronic hybridization revealed that, of the 907 ESTs in the down-regulated library, 151 (24.3%) ESTs were shared with the up-regulated library when comparing hits to the gene-model database. This left a total of 756 (83.4%) ESTs that were exclusive to the down-regulated library. This list of exclusive ESTs was used to characterize the library of down-regulated ESTs.

Efficiency of suppression subtractive hybridization

A sampled randomization test was done to generate a histogram (Fig. 3.5) from which the efficiency of the SSH technique could be determined. The y-axis of the resulting histogram is the number of randomly generated library pairs that have a specified number of shared ESTs. The x-axis is the number of ESTs that are shared between each library pair. The observed number of shared ESTs between the two libraries in this study was 151. 99.99% of the randomly generated library pairs had 152 or more shared ESTs. Therefore, the number of observed shared ESTs between the two libraries in this study could not be expected to happen by chance ($p \leq 0.01$) and the H_O is rejected. From these results it can be concluded that the SSH technique efficiently, though not entirely, removed housekeeping genes from the two libraries.

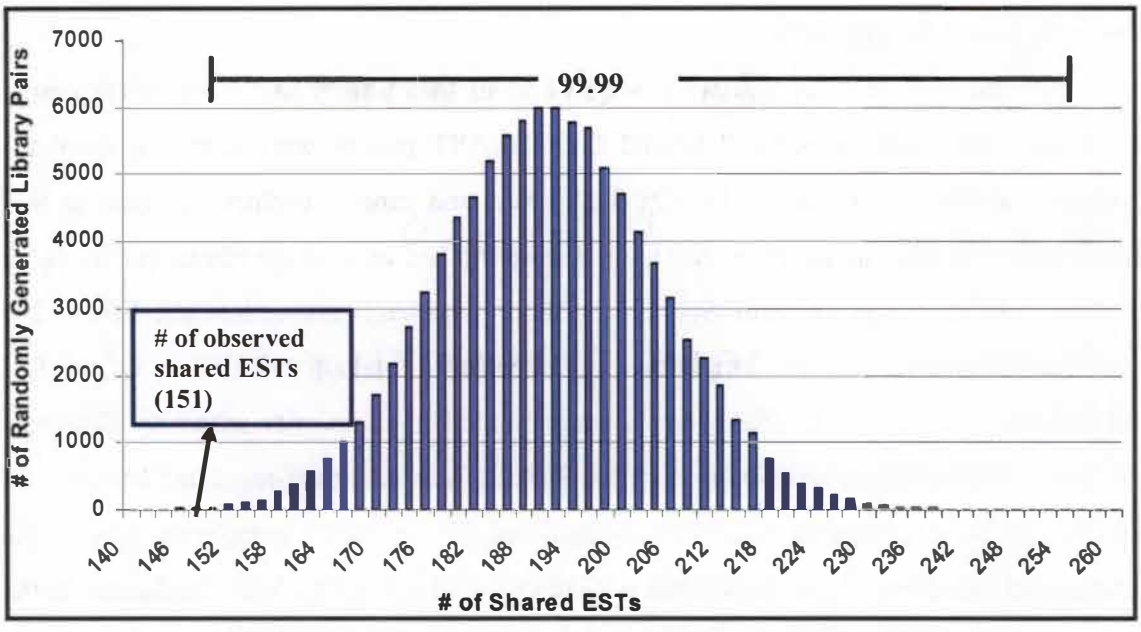

Figure 3.5: Histogram generated with a sampled randomization test

Functional categorization

ESTs in each library were categorized into general functional assignments according to the classification developed by MIPS (Fig. 3.6). These functional categorizations are solely based on inference from BLAST results of the ESTs in each library to full-length poplar gene model hits, which, in turn, were BLASTed to the Arabidopsis peptide database. It should also be noted that, although the initial BLAST comparisons were made using contigs and singletons, percentages in the graphs are calculated from the total number of ESTs in each functional category to reflect relative levels of gene expression. Additionally, multiple ESTs were assigned to more than one functional category. Therefore, percentages are based on the total number of ESTs after functional categorization rather than the original number of exclusive ESTs in each library. The highest percentages of ESTs in each library were found to encode proteins that are, as yet, unclassified or are not in "clear cut" categories. The EST Arabidopsis

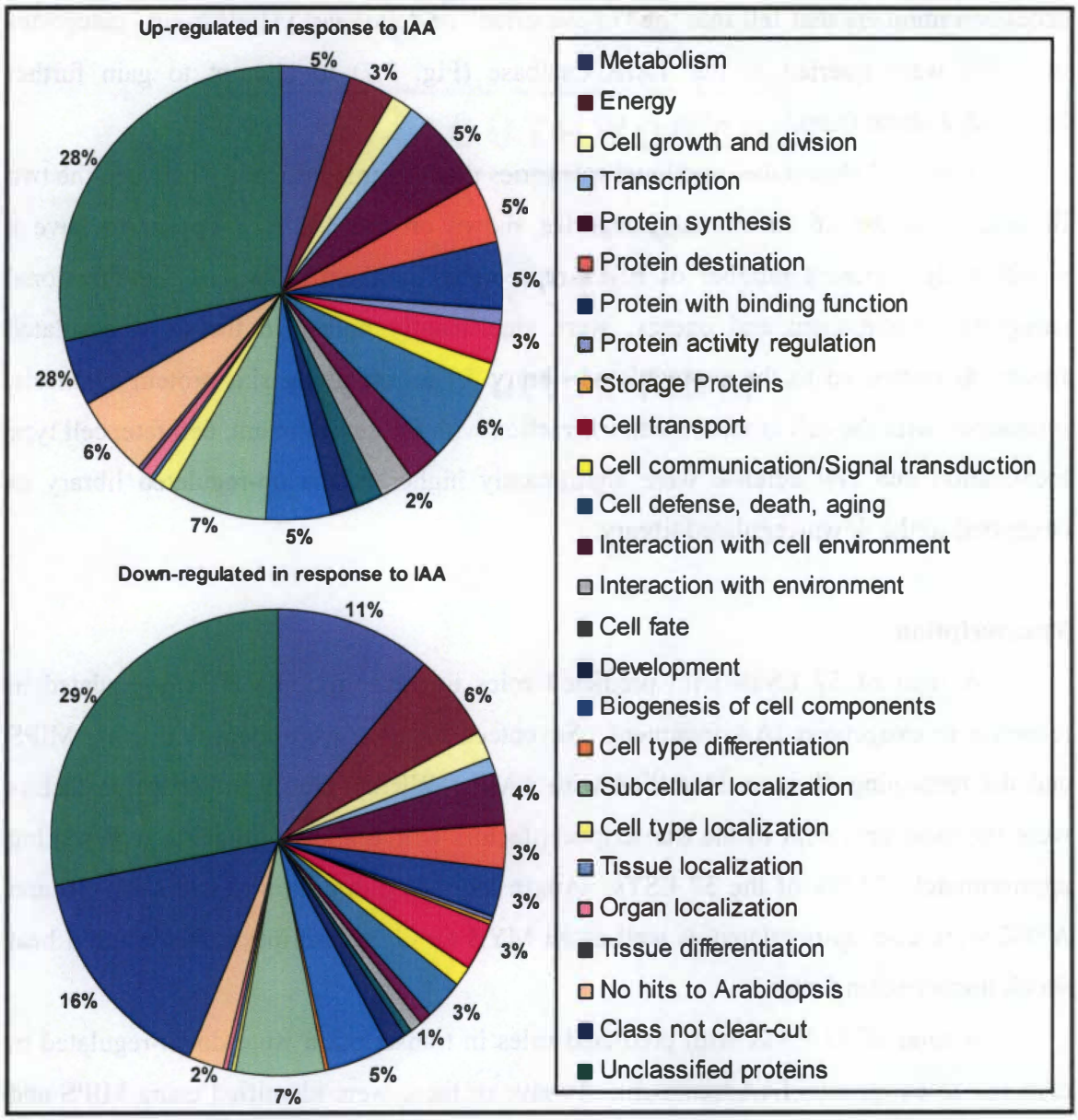

Figure 3.6: Distribution of ESTs into MIPS categories for each of the libraries

accession numbers that fell into the "unclassified" or "class not yet clear-cut" categories in MIPS were queried in the TAIR database (Fig. 3.7) to attempt to gain further knowledge about them.

Table 3.2 shows the functional categories that differ significantly between the two libraries. Of the 26 functional categories shown on the Table, 8 appear to have a significantly different number of ESTs represented in them. Two of the functional categories, metabolism and energy, were significantly higher in the down-regulated library as compared to the up-regulated library. The remaining six, protein synthesis, interaction with the cell environment, interaction with the environment, cell fate, cell type localization and cell defense were significantly higher in the up-regulated library as compared to the down-regulated library.

Transcription

A total of 57 ESTs with predicted roles in transcription were up-regulated in response to exogenous IAA treatment. Seventeen of these were identified using MIPS and the remaining 40 were identified with TAIR. WRKY family transcription factors were the most prevalent of the transcription factors with a known function, representing approximately 22.8% of the 57 ESTs. Auxin induced transcription factors IAA16 and ATB2 were also up-regulated as well as an MYB family transcription factor and a heat shock transcription factor.

A total of 33 ESTs with predicted roles in transcription were down-regulated in response to exogenous IAA treatment. Twelve of these were identified using MIPS and 21 were identified using TAIR. An auxin response factor (ARF9), as well as an IAA7-like protein, were both down-regulated, as well as a scarecrow-like transcription factor, an E2 ubiquitin-conjugating enzyme and an ethylene-insensitive protein (EIN3).

Protein synthesis, destination, regulation and binding

A total of 148 ESTs with predicted roles in protein synthesis, destination, regulation and binding were up-regulated in response to exogenous IAA treatments. The majority of those with a known function were ribosomal proteins (25.7%). In addition, a

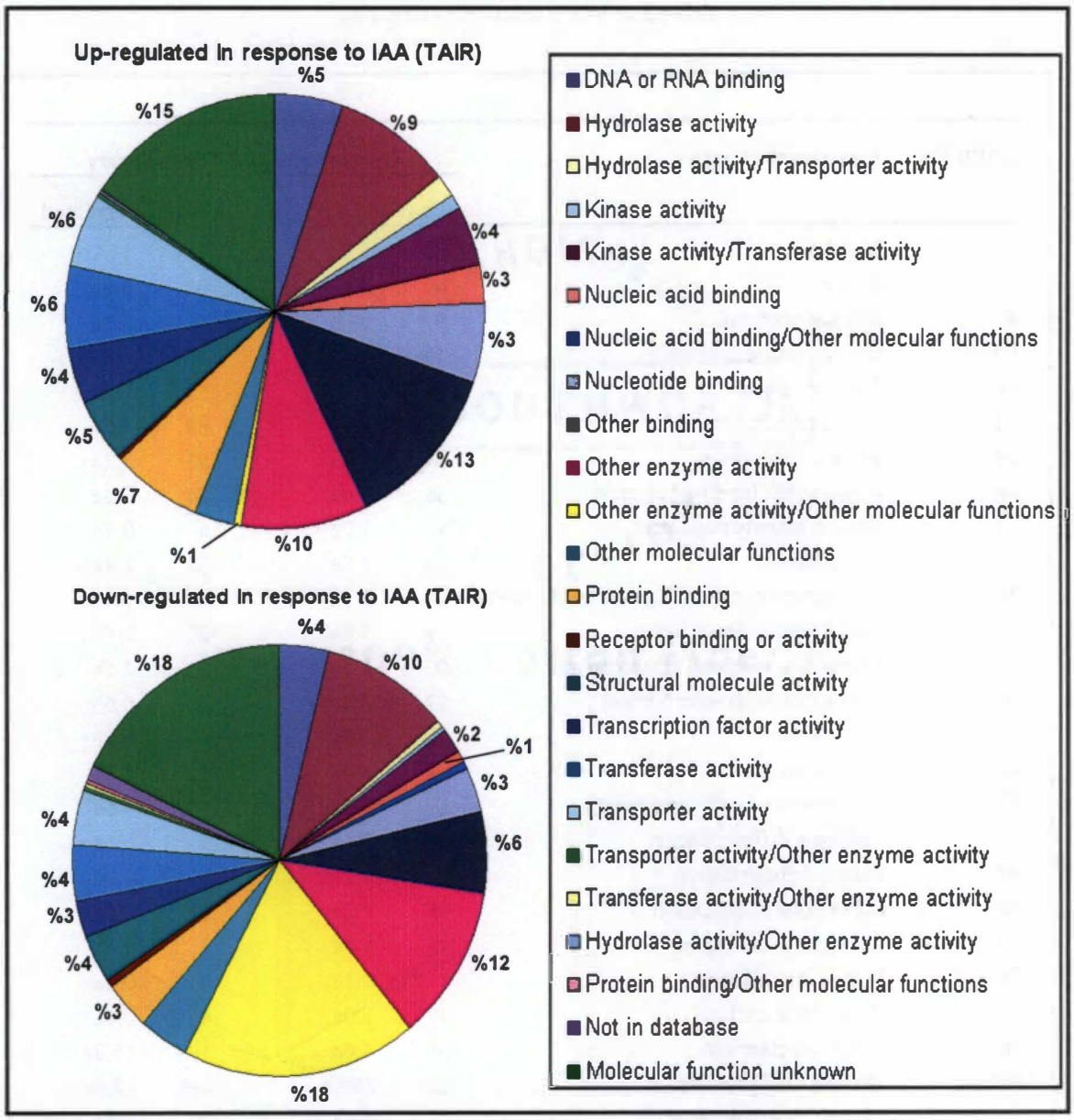

Figure 3.7: Distribution of ESTs into TAIR categories that were "unclassified proteins" or "class not clear-cut" proteins in MIPS

Table 3.2: MIPS functional categories

MIPS No.	Functional category	Number and % ESTs in library			
		Up-regulated		Down-regulated	
1	Metabolism	60	5.1a	114	11.1b
2	Energy	40	3.4a	64	6.2b
4	Storage proteins	0	0.0a	1	0.1a
10	Cell growth and division	19	1.6a	19	1.8a
11	Transcription	17	1.4a	12	1.2a
12	Protein synthesis	65	5.5a	36	3.5b
14	Protein destination	54	4.6a	33	3.2a
16	Protein with binding function	55	4.6a	34	3.3a
18	Protein activity regulation	12	1.0a	4	0.4a
20	Cell transport	33	2.8a	35	3.4a
30	Cell communication/Signal transduction	22	1.9a	14	1.4a
32	Cell defense, death, aging	69	5.8a	28	2.7b
34	Interaction with cell environment	30	2.5a	11	1.1b
36	Interaction with environment	27	2.3a	6	0.6b
40	Cell fate	23	1.9a	8	0.8b
41	Development	24	2.0a	13	1.3a
42	Biogenesis of cell components	55	4.6a	47	4.6a
43	Cell type differentiation	1	0.1a	1	0.1a
45	Tissue differentiation	2	0.2a	0	0.0a
70	Subcellular localization	85	7.2a	67	6.5a
73	Cell type localization	18	1.5a	4	0.4b
75	Tissue localization	1	0.1a	2	0.2a
77	Organ localization	12	1.0a	4	0.4a
98	Class not clear-cut	55	4.6a	163	15.8a
99	Unclassified proteins	337	28.4a	288	27.9a
	No hits to Arabidopsis	70	5.9a	23	2.2a
	Total	1186	100.0	1031	100.0

Percentages within a row followed by a letter in common are not significantly different ($P \leq 0.05$). Calculations were based on the significance tests of Audic and Claverie (1997).

relatively high percentage of the up-regulated ESTs were found to be heat shock proteins (23.0%) and cyclophilin proteins (10.8%).

A total of 119 ESTs with predicted roles in protein synthesis, destination, regulation and binding were down-regulated in response to exogenous IAA treatments. Similar to the up-regulated proteins, the most prevalent proteins in the down-regulated library were ribosomal proteins (31.9%). However, unlike the up-regulated library, only 5.9% of the proteins in this functional category were heat shock proteins and no known cyclophilins were detected.

Metabolism

The resulting Arabidopsis BLAST hits that fell into the MIPS functional category of metabolism were entered into the KEGG database to analyze the different trends in metabolic pathways between the two libraries. Arabidopsis hits that were unique to each library, rather than exclusive poplar gene model hits and their corresponding Arabidopsis hits, were entered into the KEGG database to reduce confusion. This was done because, even though poplar gene model hits were exclusive in each library, the poplar genome contains more genes than the Arabidopsis genome. Therefore it is possible for different poplar gene models to have a hit to the same Arabidopsis accession number.

Three metabolic pathways were focused on including the carbon fixation pathway (Fig. 3.8 and 3.9), the nitrogen reduction and fixation pathway (Fig. 3.10 and 3.11) and the flavonoid biosynthesis pathway (Fig. 3.12 and 3.13). Of the 23 plant specific enzymes that catalyze reactions in the carbon fixation pathway, four were up-regulated in response to IAA treatment and six were down-regulated in response to IAA treatment. The remaining enzymes were not detected. Of the 14 plant specific enzymes that catalyze reactions in the nitrogen reduction and fixation pathway, three were up-regulated in response to IAA and four were down-regulated in response to IAA. The remaining enzymes were not detected. Of the 10 plant specific enzymes that catalyze reactions in the flavonoid biosynthesis pathway, one was up-regulated in response to auxin and one was down-regulated in response to IAA. The remaining enzymes were not detected in either the up or down-regulated libraries.

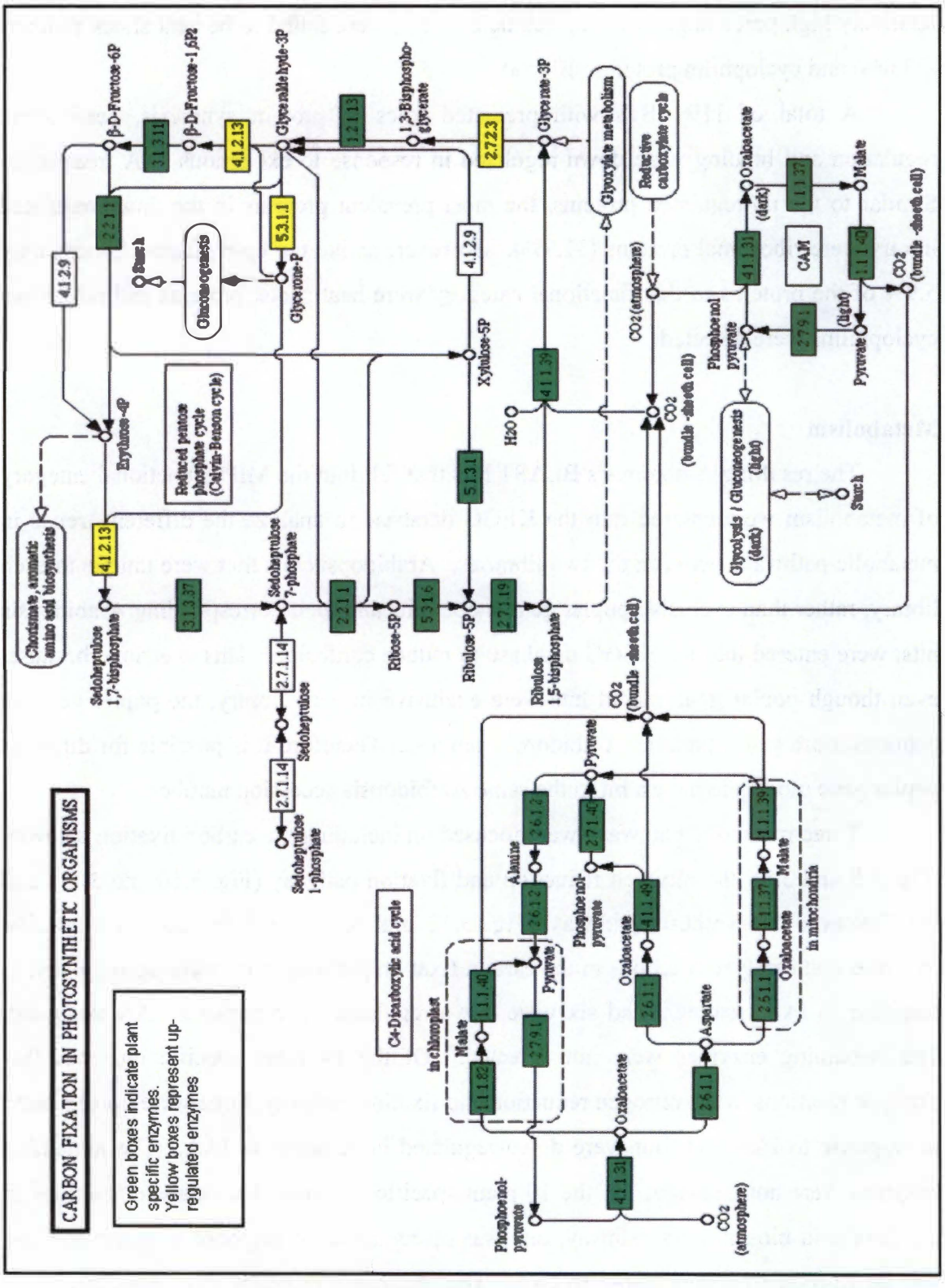

Figure 3.8: Up-regulated carbon fixation pathway
From http://www.genome.jp/kegg/

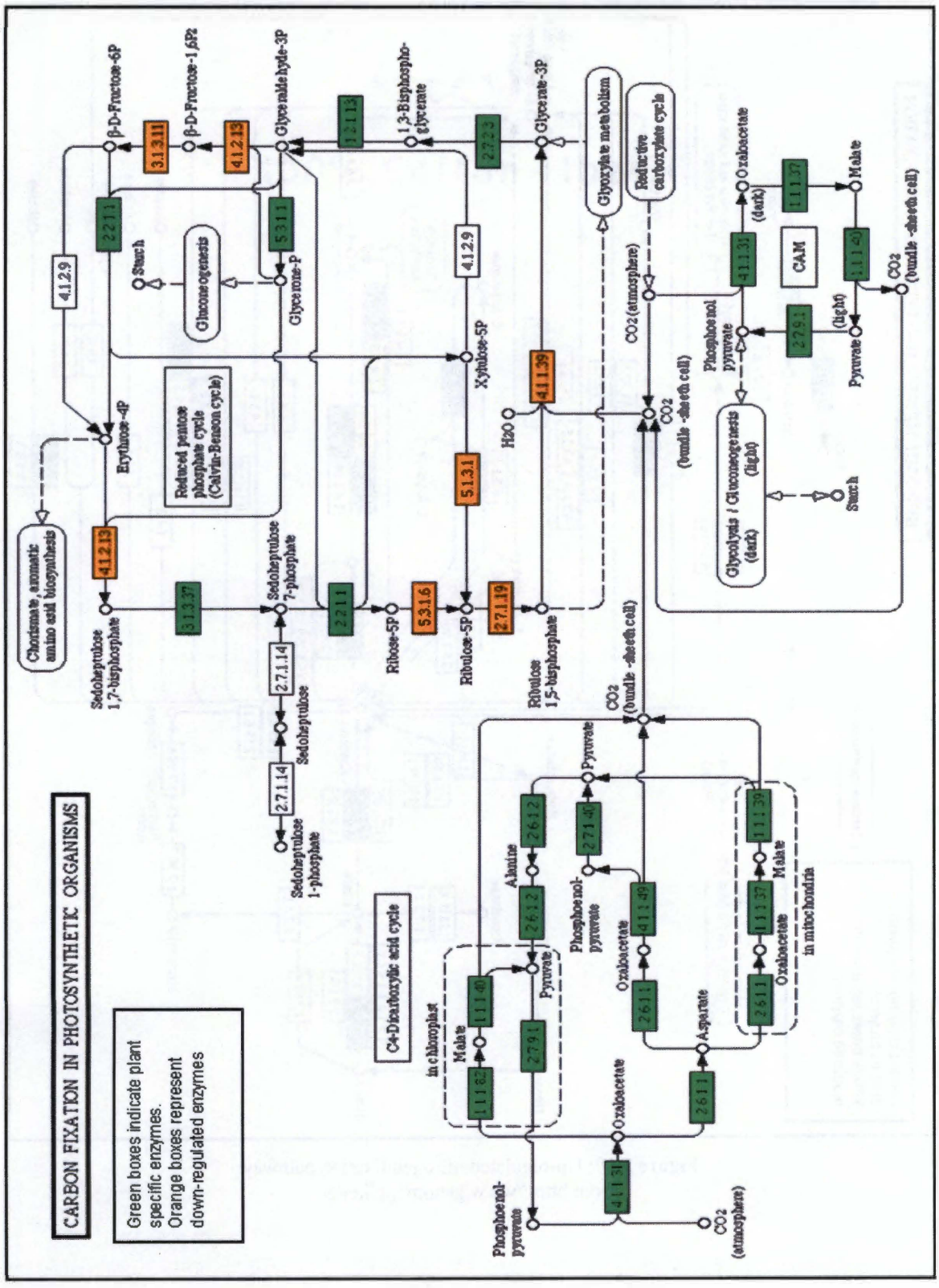

Figure 3.9: Down-regulated carbon fixation pathway
From http://www.genome.jp/kegg/

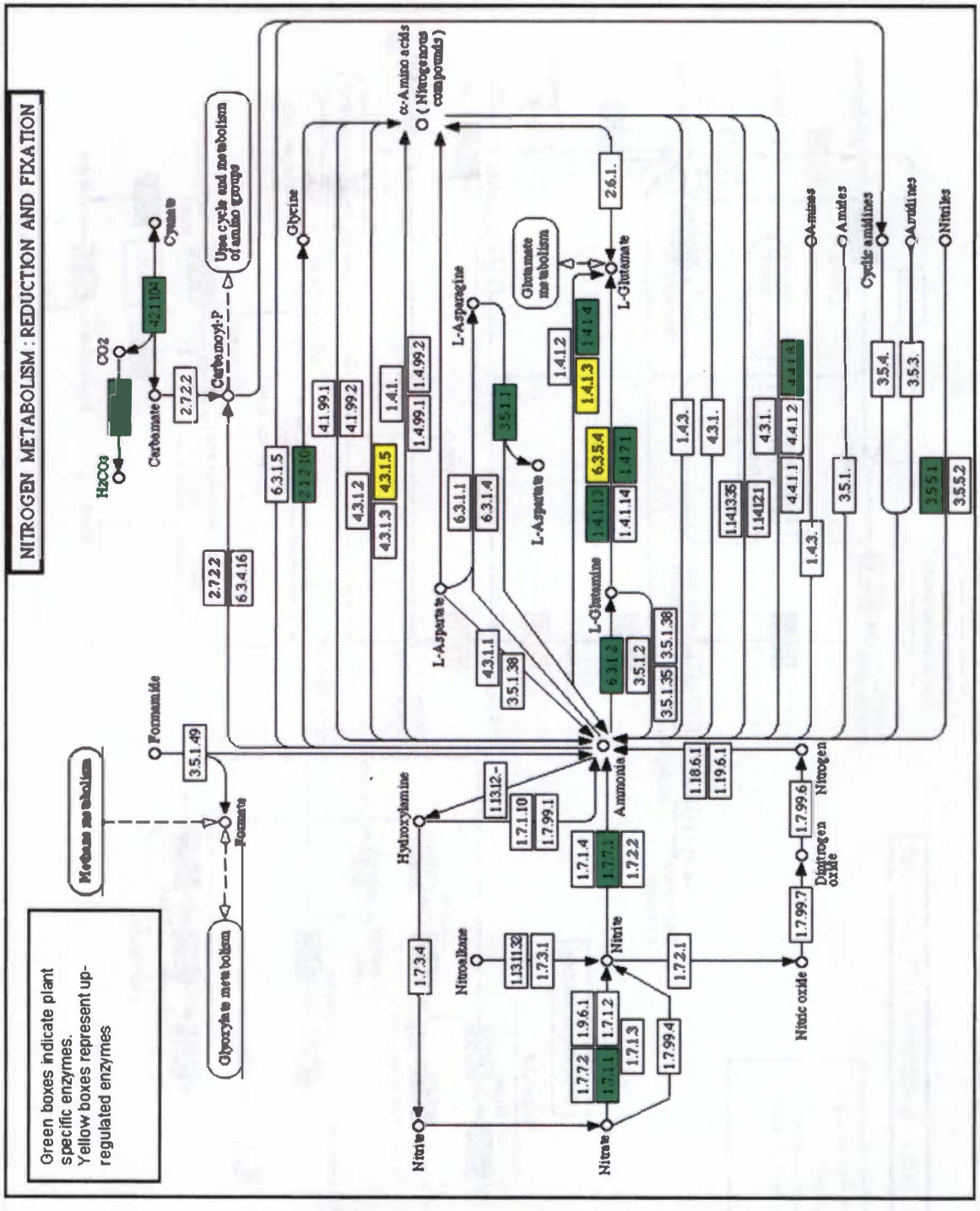

Figure 3.10: Up-regulated nitrogen fixation pathway
From http://www.genome.jp/kegg/

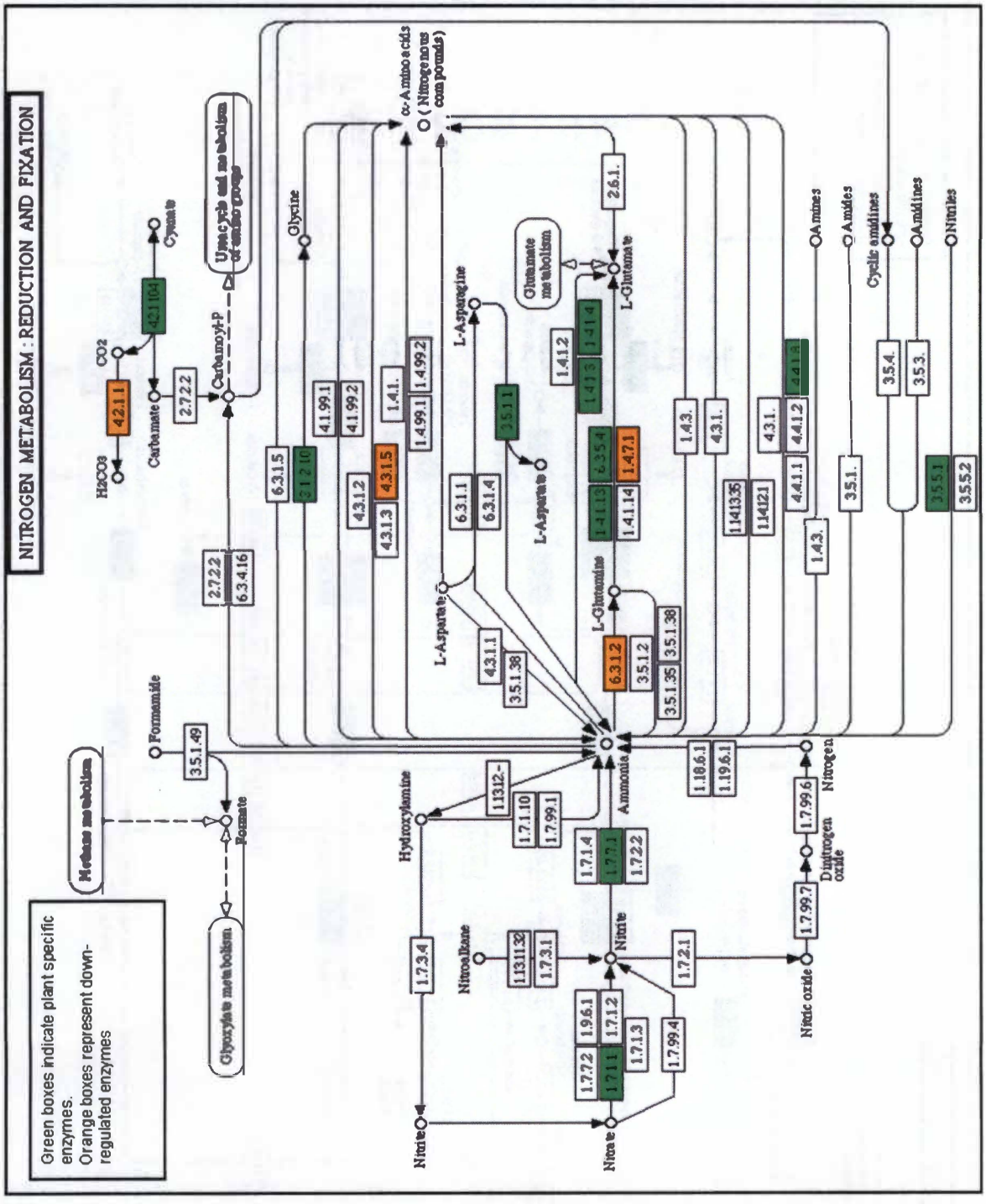

Figure 3.11: Down-regulated nitrogen fixation pathway
From http://www.genome.jp/kegg/

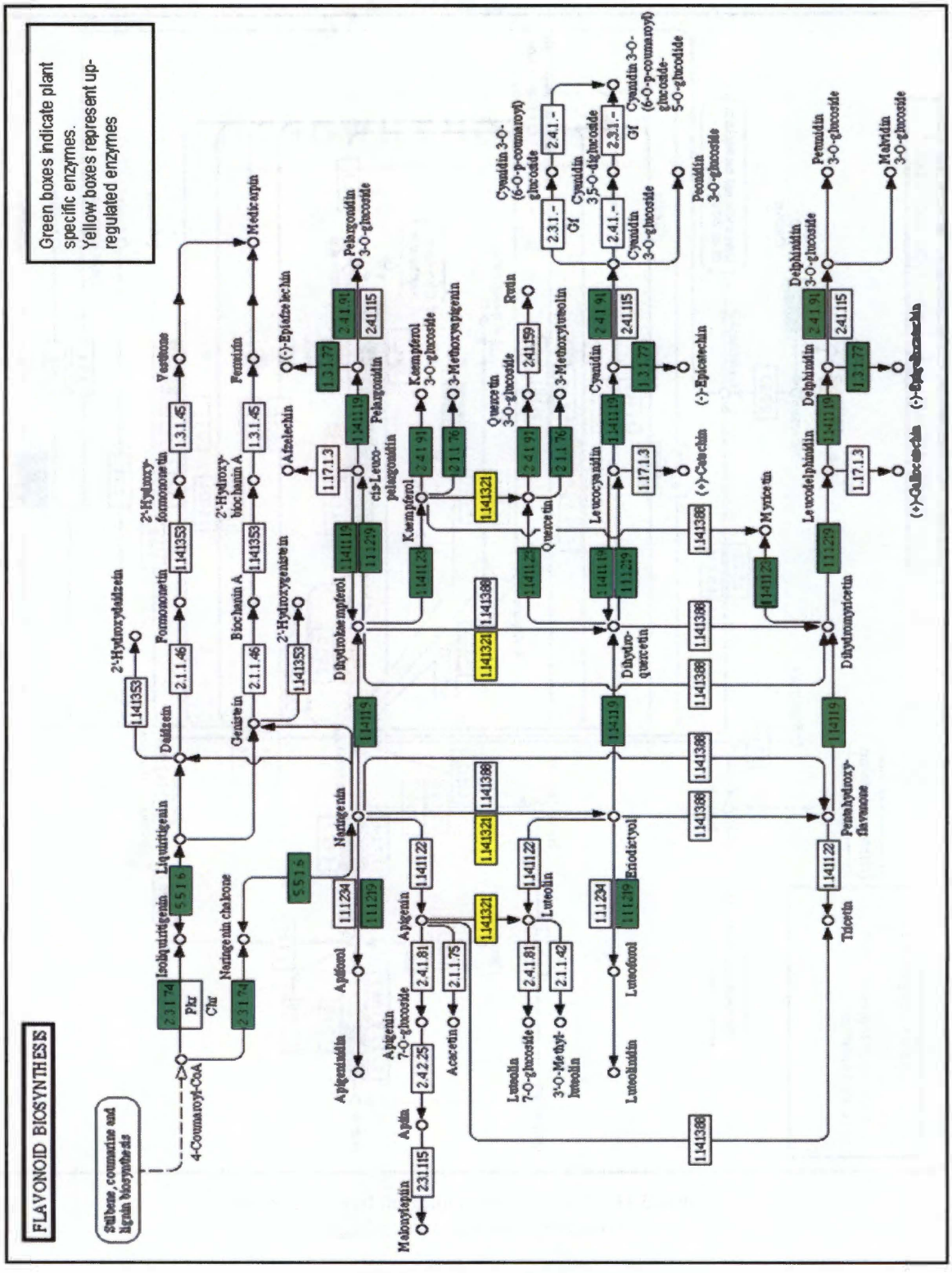

Figure 3.12: Up-regulated flavonoid biosynthesis pathway
From http://www.genome.jp/kegg/

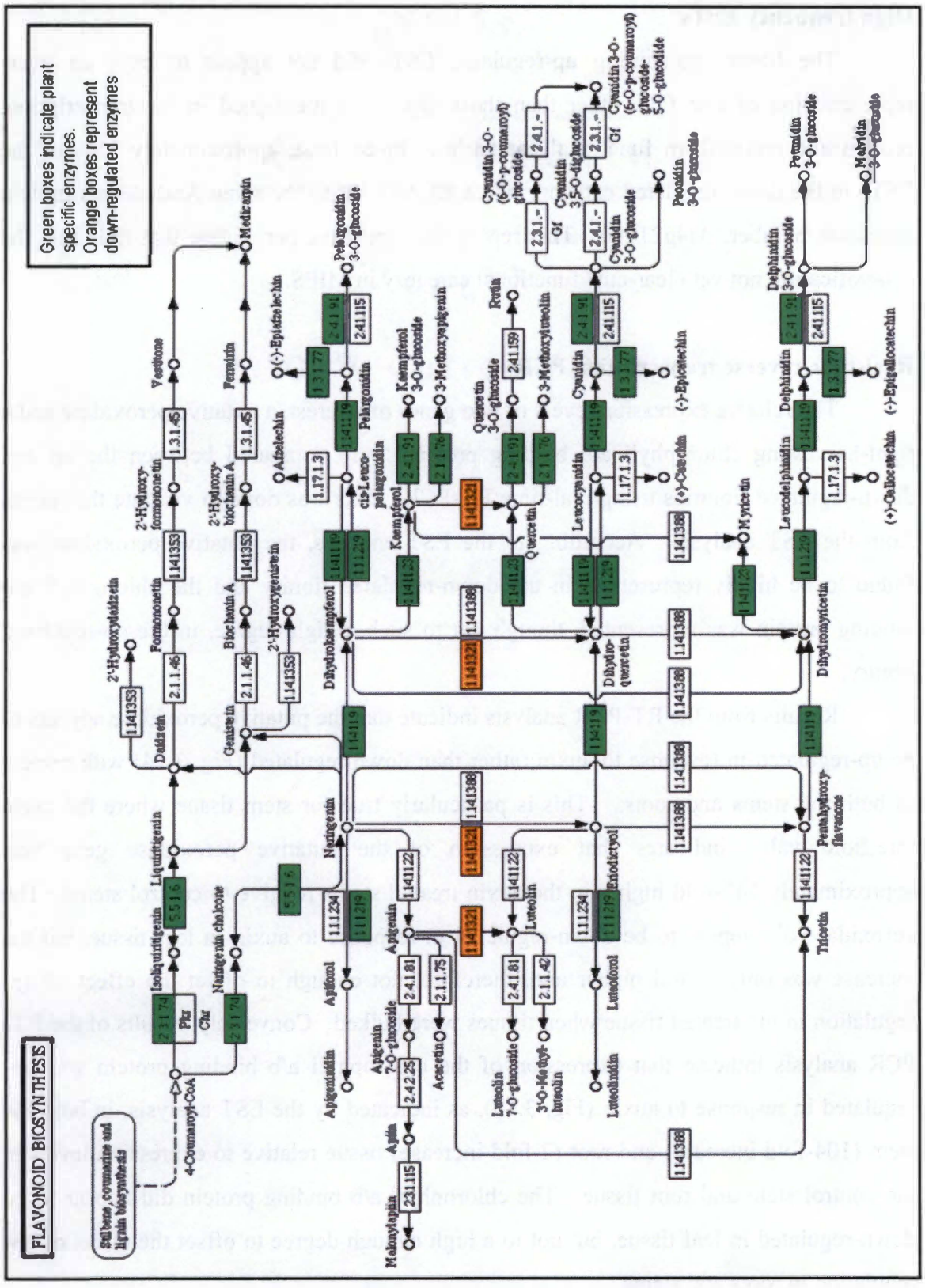

Figure 3.13: Down-regulated biosynthesis pathway
From http://www.genome.jp/kegg/

High frequency ESTs

The library containing up-regulated ESTs did not appear to have an over-representation of one EST other than those that were mentioned in the transcription, protein and metabolism functional categories. In contrast, approximately 15% of the ESTs in the down-regulated category had a BLAST hit to the same Arabidopsis protein accession number, At4g21960. This represents a putative peroxidase that falls into the "classification not yet clear-cut" functional category in MIPS.

Real-time reverse transcription PCR

The relative expression levels of two genes of interest, a putative peroxidase and a light-harvesting chlorophyll a/b binding protein, were compared between the up and down-regulated libraries using real-time RT-PCR. This was done to validate the results from the EST analysis. According to the EST analysis, the putative peroxidase was found to be highly represented in the down-regulated library and the chlorophyll a/b binding protein was represented, though not to such a high degree, in the up-regulated library.

Results from the RT-PCR analysis indicate that the putative peroxidase appears to be up-regulated in response to auxin rather than down-regulated (Fig. 3.14) with respect to both the stems and roots. This is particularly true for stem tissue where the cycle threshold value indicates that expression of the putative peroxidase gene was approximately 147-fold higher in the auxin treated stems relative to control stems. The peroxidase did appear to be down-regulated in response to auxin in leaf tissue, but the increase was only 4-fold higher and, therefore, not enough to offset the effect of up-regulation in the treated tissue when tissues were bulked. Conversely, results of the RT-PCR analysis indicate that expression of the chlorophyll a/b binding protein was up-regulated in response to auxin (Fig. 3.15), as indicated by the EST analysis, in both the stem (104-fold increase) and root (2-fold increase) tissue relative to expression levels in the control stem and root tissue. The chlorophyll a/b binding protein did appear to be down-regulated in leaf tissue, but not to a high enough degree to offset the effect of up-regulation in roots and stems.

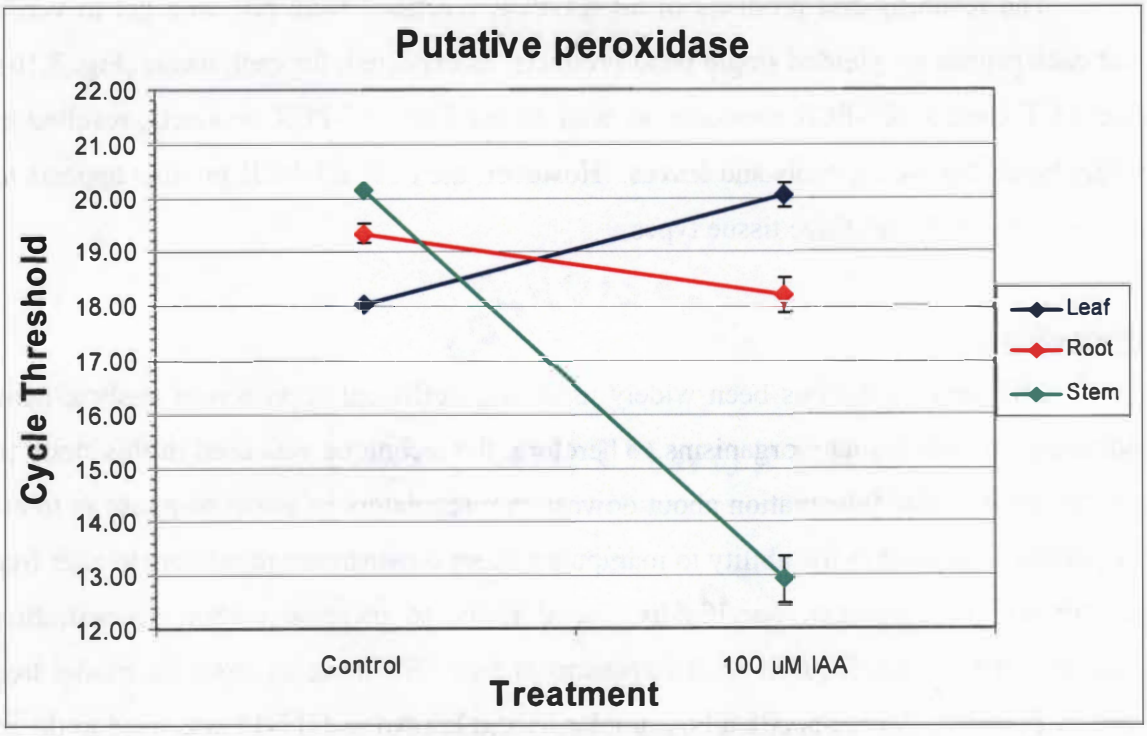

Figure 3.14: Normalized average PER C_t values obtained from triplicate real-time RT-PCR

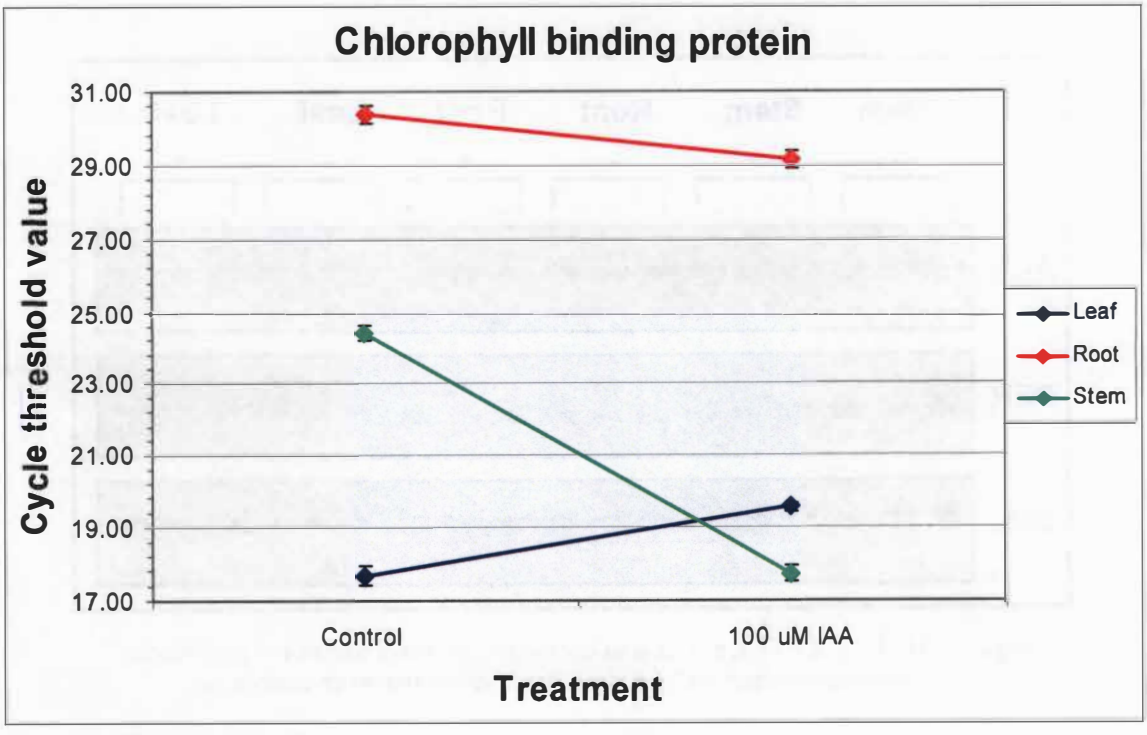

Figure 3.15: Normalized average CHL C_t values obtained from triplicate real-time RT-PCR

The resulting end products of all RT-PCR reactions were run on a gel to verify that each primer set yielded single band products, as expected, for each tissue (Fig. 3.16). The ACT control RT-PCR products, as well as the CHL RT-PCR products, resulted in single bands for stems, roots and leaves. However, the PER RT-PCR product appears to have two bands for all three tissue types.

Discussion

EST sequencing has been widely used as an efficient approach to analyze gene expression trends in many organisms. Therefore, the technique was used in this study to attempt to elucidate information about downstream regulators of auxin response in trees. Of particular interest is the ability to manipulate these downstream regulators to alter tree growth and development, specifically lateral roots, to increase carbon sequestration capacity. This study focused on the creation of two EST libraries from the model tree species *Populus*. More specifically, a poplar hybrid known as H11-11 was used to do an expression analysis of poplar genes that are up and down-regulated in response to exogenous auxin treatment. The two libraries were created using SSH, a technique that

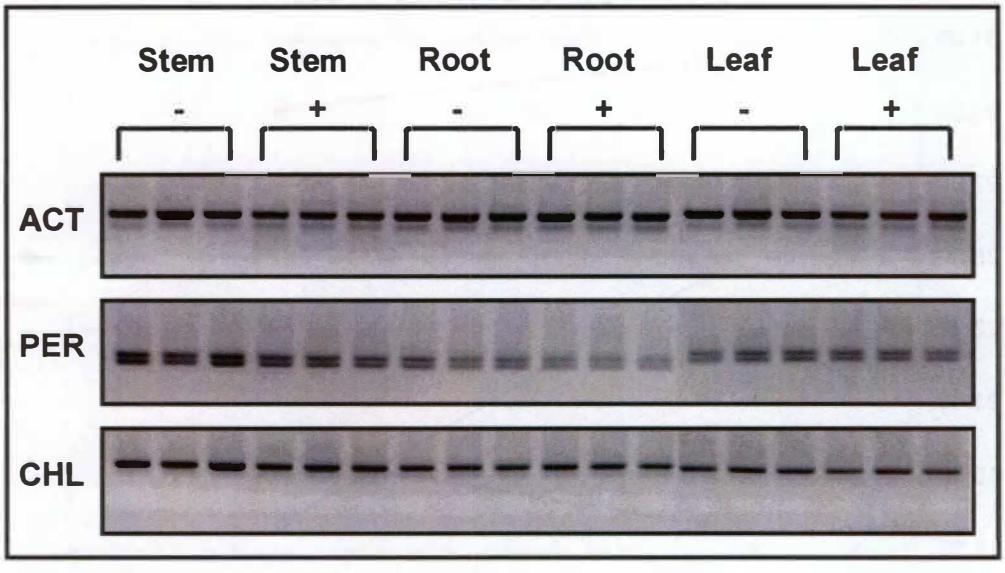

Figure 3.16: Final RT-PCR products of control (-) and auxin treated (+) plant tissues. Products were stained with ethidium bromide and run on an agarose gel.

removes common "housekeeping" gene cDNAs to reveal genes that are differentially expressed between two cDNA populations.

Rolling circle amplification

Traditionally, before cloned cDNAs can be successfully sequenced they need to be amplified in LB media and purified with standard plasmid prep kits. These kits are typically expensive and can be time consuming to use. RCA, an alternative to a traditional plasmid prep kit, was tested to see if it is an efficient and cost effective method to amplify cDNAs in preparation for sequencing. The technique was found to require less benchwork time because no purification step is required, it results in an amplified linear product of higher and more consistent quality and, the most striking attribute, it costs ¼ as much when compared to the Montage™ plasmid prep kit. All of these characteristics make RCA a viable option for cDNA amplification if the protocol is standardized correctly and BSA is not used in the reaction.

A total of 960 ESTs were sequenced in each of the two libraries. BLAST results against the poplar gene model database and clustering revealed that the up-regulated library contained 731 exclusive ESTs and 699 non-redundant ESTs, while the down-regulated library contained 756 exclusive ESTs and 622 non-redundant ESTs. This information was used to functionally categorize ESTs in each library using comparative genomic techniques and to identify ESTs that are not currently in the poplar unigene database.

Transcription factors

Approximately 3.9% of the ESTs in the up-regulated library potentially function to facilitate transcription. Only two of 57 up-regulated ESTs in the transcription category, Aux/IAA16 and ATB2, were identified as previously supposed auxin-induced proteins. In general, Aux/IAA proteins are involved in a feedback loop in which they repress auxin response factors (ARFs) when auxin levels are low (Hagen and Guilfoyle 2002). An increase in auxin facilitates Aux/IAA degradation, ARFs are derepressed and auxin regulated genes, including Aux/IAA, are transcribed. When auxin levels decrease once

again, Aux/IAA protein levels increase and ARFs are repressed (Liscum and Reed 2002). The functional relevance of Aux/IAA16 in particular, however, has not yet been characterized in detail. The second putative auxin induced gene identified in the up-regulated library, ATB2, is known to be translationally, rather than transcriptionally, repressed by elevated levels of sucrose and is expressed in vascular tissues of seedlings and young vegetative tissues. Functionally, it is thought to coordinate metabolism associated processes in newly established sinks (Wiese *et al.* 2004).

WRKY transcription factors were the most abundant type (~22.8%) in the up-regulated library. Members of this family are exclusively found in plants and are characterized by a WRKY domain in their n-terminal end. Most WRKY proteins induce gene expression by binding to DNA W-box domains typically found in defense related genes (Eulgem *et al.* 2000). Three of the four WRKY genes identified in the up-regulated library, WRKY 70 (Li *et al.* 2004, Ulker and Somssich 2004), WRKY 69 and WRKY 39 (Dong *et al.* 2003), are thought to induce expression of defense related genes because of their ability to bind to W-box domains. However, they have not yet been fully characterized. A fourth WRKY family member, WRKY 51, identified in the up-regulated library is thought to play a possible role in leaf senescence (Guo *et al.* 2004).

Approximately 2.4% of the ESTs in the down-regulated library are potentially involved in facilitating transcription and a particular type of transcription factor was not more abundant than any other. Three of the 33 down-regulated genes identified in the transcription category, ARF9, Aux/IAA7-like and scarecrow-like 1(SCR1), were all previously known to be regulated by auxin. ARF9 does not appear to have either of the characteristic Q or P-rich middle regions typical of ARF activators and repressors respectively. Additionally, over-expression of ARF9 has not been shown to activate or repress transcription. Therefore, it is thought that apparently inert ARF9 proteins may function by binding to DNA target sites and serve as scaffolds for ARF and/or Aux/IAA activators and repressors (Ulmasov *et al.* 1999a). Aux/IAA7 has been shown to be a crucial element in several distinct cellular processes at all stages of development. For example, Aux/IAA7 is thought to be involved in tissue patterning in roots, cell enlargement, gravitropism response and seedling shoot development in light (Nagpal *et al.*

2000). Studies have shown that a mutation in Aux/IAA7 leads to auxin insensitivity (Liscum and Reed 2002). In *Brassica*, SCR1 is a transcriptional activator that interacts with histone deacetylase (HDA). It contains domains that are conserved in the GRAS (GAI, RGA, SCR) family of proteins and interacts with HDA19 through a VHIID domain. It is expressed predominantly in the roots, but also in shoots and mature leaves where it might play a role in radial pattering. It is thought to be regulated by auxin, however, it may also be regulated by RNAi through cleavage of mRNA by miRNA (Gao *et al.* 2004).

In addition to the previously mentioned genes, a few other transcription factors were also identified in the down-regulated library including E2 and EIN3 (Ethylene-insensitive 3). E2 is part of a highly conserved pathway that promotes covalent attachment of ubiquitin to a protein to facilitate degradation of the protein by the 26S proteosome (Hellmann and Estelle 2002). EIN3 proteins are essential transcription factors, along with EIL (Ethylene insensitive-like) proteins in the ethylene signaling pathway of higher plants. For example, when ethylene levels are elevated EIN3/EIL proteins are activated resulting in the expression of ethylene response genes.

Due to the high rate of expression of defense and stress related transcription factors in the up-regulated library, it is possible that the intense nine day exogenous auxin treatment induced a stress response that is facilitated by these particular transcription factors. This is supported by the fact that there were no defense related transcription factors in the down-regulated library indicating that defense related genes were not over-expressed in the control plants.

Proteins

Approximately 15.9% of the ESTs in the up-regulated library play a predicted role in protein synthesis, destination, regulation and binding. The majority of the ESTs in this functional category (25.7%) were ribosomal proteins. Heat shock proteins (Hsps) were also highly abundant (23.0%) as were cyclophilin proteins (10.8%). Approximately 10.6% of the ESTs in the down-regulated library play a predicted role in protein synthesis, destination, regulation and binding. As in the up-regulated library, the majority of the

ESTs in this functional category (31.9%) were ribosomal proteins. In contrast to the up-regulated library, there were relatively few down-regulated heat shock proteins (5.9%) and no cyclophilin proteins were identified.

The ribosome is a complex structure that consists of four rRNAs and about 80 ribosomal proteins. Thousands of ribosome complexes are dispersed throughout the cytoplasm or are associated with rough endoplasmic reticulum (ER). They function as an essential piece of machinery responsible for protein synthesis, and as such play a major role in controlling cell growth, division and development (Barakat *et al.* 2001). Therefore, it is not surprising that ribosomal proteins have a relatively high representation in both libraries. On the other hand, it is interesting that there appears to be a shift in the type of ribosomal protein that is expressed between the control plants and the plants treated with exogenous IAA. Unfortunately, the specific ribosomal proteins expressed in the two libraries are not well characterized making it difficult to determine why one set might be expressed under certain conditions while another is not.

Hsps are part of a stress response that is triggered during a sudden elevation in temperature. Hsps function in one of two ways, either by counteracting protein denaturation and aggregation or by facilitating degradation of non-native proteins, to enable cells to survive the harmful effects of heat shock. With the exception of Hsp101 and Hsp60, both present in the down-regulated library, very little specific information is known about most of the hsps identified in this study. Hsp101 is thought to facilitate reactivation of proteins denatured by heat (Gurley 2004) and proteins in the Hsp60 family are thought to play a crucial role in refolding and assembling higher-order proteins. Although hsps are widely known for the role they play in stress response, several hsps are also essential for plant development under normal growth conditions (Becker and Craig 1994). It is possible that the exogenous IAA treatments indirectly caused a heat shock response to occur resulting in the expression of heat shock proteins in the treated plants, while the hsps expressed in the down-regulated library are those that are only expressed under normal growth conditions.

Cyclophilin proteins are generally ubiquitously expressed in all subcellular compartments in response to both biotic and abiotic stresses including heat shock. They are involved in a wide variety of cellular processes such as protein trafficking and maturation, receptor complex stabilization, RNA processing and spliceosome assembly. CYP2 and ROC3, both cytoplasmic cyclophilins, and CYP5, an endoplasmic reticulum cyclophilin, were identified in the up-regulated library (Romano *et al.* 2004). Very little is known about the specific functional relevance of these particular cyclophilins. However, it is possible that they were also expressed in response to heat shock induced by the IAA treatments. Evidence of this is supported by the fact that cyclophilins were not present in the down-regulated library indicating that they were either not expressed in the control plant or were removed during the SSH process. Additionally, CYP-deficient mutants of yeast have shown that this protein is not needed for growth under normal conditions (McLaughlin *et al.* 1992), but is an essential element needed for survival after heat shock (Sykes *et al.* 1993). In higher plants, including Arabidopsis, wounding and other stresses have been shown to induce CYP gene expression and it is hypothesized that CYPs may participate in early signal recognition during stress responses (Chou and Gasser 1997).

Metabolism

When looking at the results from both MIPS and TAIR, proteins potentially involved in metabolism were represented by approximately 4.1% of the ESTs in the up-regulated library. Significantly more, approximately 9.4% of the ESTs, were represented in the down-regulated library. Therefore, the general trend in the amount of proteins dedicated to metabolic processes decreases in response to exogenous IAA treatments.

Three KEGG metabolic pathways, carbon fixation, nitrogen fixation and flavonoid biosynthesis, were generated from MIPS and TAIR data for both the up-regulated and the down-regulated library. This information can be used to observe possible trends in metabolic processes in response to exogenous IAA treatments. However, the presence of a specific enzyme in one of the pathways does not necessarily mean that it is actually metabolically active in that pathway. Additionally, due to the fact

that the electronic hybridization was done with EST top hits to poplar gene models rather than EST top hits to the Arabidopsis accession numbers used in the KEGG database, some enzymes appear to be represented in both the up and down-regulated libraries when, in fact, they are not. This can occur because, in some cases, the same Arabidopsis accession number represents two different, but highly similar, poplar genes.

A general trend seen within the up-regulated carbon fixation pathway indicates that there may be a shift towards increased starch production in response to exogenous IAA application (Fig. 3.7). Triosephosphate isomerase, EC number 5.3.1.1, appears to be present in the up-regulated library and is needed to facilitate the production of glycerone phosphate. Glycerone phosphate is then used during glucogenesis and may result in the production of starch. IAA application has been shown to increase starch content in sorghum possibly by facilitating the transport of sugars into grains and their transformation into polysaccharides (Bhatia and Singh 2002) and has also been shown to inhibit starch degradation during banana ripening (Purgatto *et al.* 2001). The presence of both ribose 5-phosphate isomerase A, EC number 5.3.1.6, and ribulose-phosphate 3-epimerase, EC number 5.1.3.1, in the down-regulated library indicates that the production of ribulose-5 phosphate and ribulose 1,5-biphosphate may be down-regulated in response to IAA thereby decreasing glyoxylate metabolism.

The resulting KEGG metabolic pathways for nitrogen fixation (Fig. 3.9 and 3.10) indicate a few possible differences in nitrogen metabolism in response to exogenous IAA treatment. First, it appears that the production of carbonic acid is down-regulated in response to IAA treatments because the enzyme that catalyzes its production, carbonic anhydrase 1 (E.C. number 4.2.1.1), is present in the down-regulated library. A study done with *Medicago* (alfalfa) roots indicated that, while treatment with cytokinin (BAP) induced carbonic anhydrase production, treatment with auxin (2,4-D) did not (Coba de la Peña *et al.* 1997). Therefore, it is possible that the enzyme is down-regulated in response to exogenous IAA treatment. Second, it appears that an alternative route in glutamate metabolism may be up-regulated in response to exogenous IAA treatment when compared to the down-regulated library. For example, the pathway leading to glutamate metabolism from ammonia is down-regulated in response to IAA (Fig. 3.10), as indicated

by a decrease in glutamine synthetase (E.C. number 6.3.1.2). At the same time, a theoretical increase in glutamate dehydrogenase (E.C. number 1.4.1.3) in response to exogenous IAA treatment (Fig. 3.9) indicates that glutamate metabolism from L-asparagine may be the favored over glutamate metabolism from ammonia.

The resulting KEGG metabolic pathways for flavonoid biosynthesis (Fig. 3.11 and Fig. 3.12), a secondary metabolic compound, did not reveal any differences between the two libraries. This is due to the fact that the same enzyme, flavonoid 3'-monooxygenase (E.C. 1.14.13.21), appeared to be present in both libraries. This was somewhat unexpected as flavonoids are thought to play a role in auxin transport. For example, plants grown in media containing, but not limited to, flavonoids had reduced auxin transport (Brown *et al.* 2001) and flavonoids have been shown to reduce polar auxin transport in zucchini hypocotyls (Jacobs and Rubery 1988). Additionally, analysis of Arabidopsis plants defective in flavonoid biosynthesis resulted in phenotypes that were consistent with elevated auxin transport (Brown *et al.* 2001). Flavonoids are thought to inhibit auxin transport by modulating vesicular cycling of auxin efflux carrier proteins known as PIN proteins (Peer *et al.* 2004). Therefore, a difference in flavonoid biosynthesis between the two libraries was expected.

Conclusion

The study of genes that are regulated by the plant hormone auxin is a crucial step in understanding how *Populus* root growth can be manipulated to increase carbon sequestration. Additionally, any information that can be added to the existing poplar EST database contributes to the tools already available to those studying the functionality of the *Populus* genome. The EST analysis carried out in this study resulted in the discovery of 33 ESTs from the up-regulated library and 15 ESTs from the down-regulated library that were not currently in the *Populus* EST database. However, further analysis, which included characterization of the ESTs in each of the libraries using comparative genomics tools, could only be used to make broad generalizations about trends that ***might*** be occurring in poplar in response to exogenous applications of IAA. This is due in large part to several limitations on the experiment. First, the exogenous IAA application and

subsequent suppression subtractive hybridization were carried out in a different lab before being sent to ORNL for sequencing and analysis. This left little control as to how these procedures were done. Second, plant roots, stems and leaves were bulked in each library. This may have caused many important genes differentially expressed relative to tissue type, but present in both control and treated bulk samples, to be removed during the SSH and not included in further analysis. Third, though equal amounts of total RNA were pooled before cDNA synthesis, this does not ensure that each tissue type will still be equally represented in the resulting EST library. Real-time RT-PCR verified that unequal tissue representation may have occurred by showing that, though the putative peroxidase appeared to be highly down-regulated in response to exogenous auxin treatment, it was only moderately down-regulated in the leaves and was actually greatly up-regulated in stems. Fourth, late responses to auxin (9 days in this case) have not been well studied, especially in tree species, making it difficult to interpret the results of the EST characterization. Fifth, the initial sample size in this study, 960 ESTs per library, is relatively small compared to the number of genes that are probably up or down regulated in response to IAA treatment. To obtain a more thorough idea of the extent of auxin response in poplar, a larger number of ESTs should be studied. Sixth and last, the exogenous IAA application appeared to induce a stress response. This may have masked a legitimate late auxin response. In the future, a one time application will administered rather than an intense daily application.

Of particular interest in the future is the use of whole genome microarray analysis to study the expression of *Populus* in response to auxin treatment. This would eliminate the problem of having to select only a limited number of clones in an EST analysis and would give a much more thorough indication of the genes that are up or down regulated in response to auxin. Additionally, future work should also include tissue or organ specific analysis. This would decrease the complexity of interpretation because responses could be localized to a particular plant tissue or organ type.

Chapter 4 : Microarray Analysis of Gene Expression

Introduction

Populus

Members of the *Populus* genus are included in the family Salicaceae (Bremer *et al.* 1998) and consist of deciduous trees such as aspen, poplar and cottonwood (Bradshaw *et al.* 2000). Several members of the genus are commercially important and serve a variety of purposes in the pulpwood and lumber industry (Balatinecz and Kretschmann 2001). In addition, poplars have been used as an alternative energy source (Isebrands 1979) and are known to have tremendous ability to sequester carbon in their root systems (Pregitzer and Friend 1996). Therefore, they are also potentially important environmental resources as well. On a molecular level, *Populus* is filling the role of model tree species because its numerous attributes including a relatively small genome that has been sequenced (Bradshaw and Stettler 1993), abundant genetic variation (Taylor 2002), ease of clonal (Dickmann and Stuart 1983) and sexual propagation (Zsuffa *et al* 1996), and ability to successfully transform (Jouanin *et* al 1993) have lead to the development of a wealth of accessible genomic tools and databases (Yin *et al.* 2002, Wullschleger *et al.* 2002a, Sterky *et al.* 1998). The available tools can be used to understand how individual genes control growth and development, eventually leading to the ability to manipulate these controls.

Objectives

One particular area of interest is the ability of poplar trees to sequester carbon in lateral roots and, more specifically, how root growth can be manipulated to increase carbon sequestration ability. A well known and essential component of lateral root growth induction is the plant hormone auxin (Muday and Haworth 1994, Reed *et al.* 1998, Casimiro *et al.* 2001). Thus, in order to better understand the molecular controls of lateral root growth, the effects of auxin on gene expression in roots needs to be studied. Here, H11-11 (*P. trichocarpa* x *P. deltoides*), a poplar hybrid, was used to analyze differential gene expression between control trees and trees treated one hour and twenty-

four hours prior to harvest. Treatment consisted of a one time foliar application of 100 μM NAA, a synthetic auxin. The goal of the study was to use whole genome microarray analysis to identify genes differentially expressed between the three treatment groups. The resulting information could be used to gain a general understanding of downstream regulators of auxin response, specifically in the roots, with the idea being that genes differentially expressed at one hour represent early response genes and genes differentially expressed at twenty-four hours represent late response genes.

Microarray technology

Microarray analysis is a hybridization based technology that can be used to simultaneously analyze the mRNA levels of hundreds or hundreds of thousands of different genes, resulting in an indication of gene activity levels within an organism (Wullschleger and DiFazio 2003). Thus far, one of the most important applications of array technology is considered to be monitoring gene expression patterns in an organism of interest. Additionally, the high throughput nature of arrays makes them an effective substitute for traditional more labor intensive procedures used to measure gene expression (Lockhart and Winzeler 2000). The use of microarray technology in plant research has lead to extensive databases of gene expression information. The databases are used to assign functional information to genes of otherwise unknown function based on the theory that genes with similar expression patterns are involved in the same biological processes (Somerville and Somerville 1999)

In general, DNA probes, or relatively short sequences representing genes in an organism of interest, are immobilized in a pattern on a solid surface and exposed to a population of mobile DNA or RNA target molecules. Labeled target molecules represent genes that are being expressed in an organism at a given point in time in response to a treatment or developmental stage. When exposed to a complimentary probe sequence, the target hybridizes to the probe and can be visualized by fluorescence excitation during the scanning process (Zhu 2003). Microarrays are categorized into two general types depending upon whether the immobilized probes are amplified cDNA fragments, also

known as cDNA microarrays, or synthetically synthesized oligonucleotides, also known as oligonucleotide microarrays (Schulze and Downward 2001).

cDNA microarrays

cDNA arrays are usually prepared using probes derived from PCR products or expressed sequence tags (ESTs) representing specific genes (Fig. 4.1). The probes vary in length from several hundred to a few thousand base pairs (Duggan *et al.* 1999) and are robotically deposited onto glass slides at 20-30 per square millimeter. An immobilization step is also required to ensure the probes are securely anchored to the glass surface (Zhu 2003). Target molecules are derived from mRNA pools extracted from treated and controls tissues of interest. Each pool is labeled with a different fluorescent label, either Cye3-dUTP (Cy3) or Cye5-dUTP (Cy5), and hybridized simultaneously to the arrayed probes (Brown and Bolstein 1999). After hybridization, a laser scanner is used to produce separate images capturing monochromatic signal from each fluorescent label depending upon its excitation wavelength. Specialized software applies a different pseudo-color to the two signals and merges the Cy3 and Cy5 images on top of one another. Changes in gene expression between the control and treated tissues are analyzed based upon the strength and type of fluorescent signal that is generated when hybridization occurs between probe sequence and target molecules. The resulting signal intensities are used to compare the relative amount of specific transcripts present in each target pool. From this information, conclusions can be drawn about genes that are differentially expressed between the target pools after the data are normalized (Duggan *et al.* 1999).

cDNA microarrays have been used extensively to study gene expression in a wide variety of different organisms and can be manufactured in-house or commercially (van Bakel and Holstege 2004). For example, the medical research community has benefited greatly from the technology and have used cDNA microarrays to study gene expression during prostate cancer progression (Nalbandian *et al.* 2005), do intestinal and diffuse gastric cancer gene expression comparisons (Wu *et al.* 2006), understand the molecular

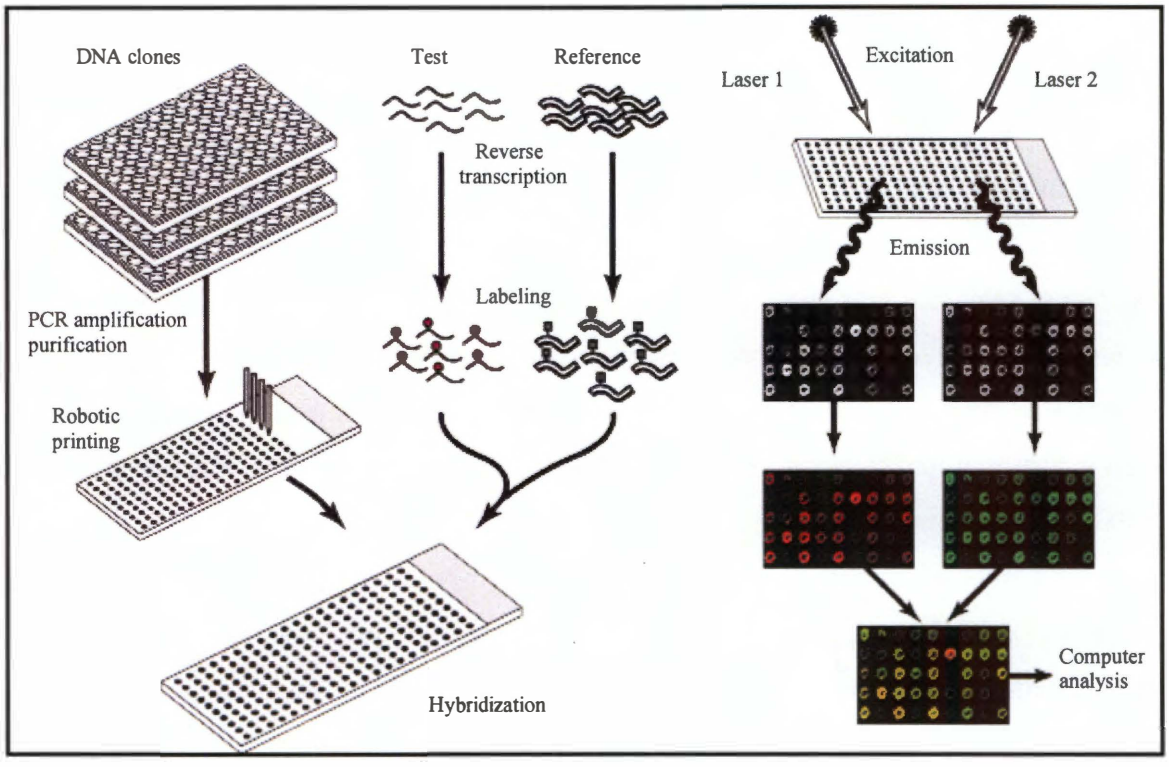

Figure 4.1: Summary of cDNA microarray steps
Adapted from Duggan *et al. 1999*

basis of kidney aging (Melk *et al.* 2005) and better understand the hepatitis C replicon system (Abe *et al.* 2005). cDNA microarrays have also been used to do extensive research in the plant community on important crop plants such as rice (Rabbani *et al.* 2003) and soybeans (Thibbaud-Nissen *et al.* 2003), fruits such as strawberry (Aharoni *et al.* 2002) and tomatoes (Eriksson *et al.* 2004) and woody plants such as pine (Brinker *et al.* 2004), hybrid aspen (Aspeborg *et al.* 2005) and eucalyptus (Kirst *et al.* 2004).

cDNA microarrays are ideally suited for expression studies on non-model species because a fully sequenced genome is not required to produce the probe matrix. Rather, cDNA microarrays only require a large cDNA library as a source of clones to be arrayed (Gibson 2002). Additionally, another advantage is that cDNA probes are long (500-5000bp) compared to traditional oligonucleotide arrays and therefore tend to be more sensitive. In contrast, there are several disadvantages associated with cDNA microarrays. These include the fact that a variable amount of DNA is spotted onto each spot because PCR efficiency is not uniform. There is also a relatively high rate (10-20%) of failed PCR reactions resulting in missing probes that are wrongly presumed present (Hoffman *et al.* 2004). Additionally, there is no control over the actual sequence or length of the cDNA probes. Though long probe sequences can be advantageous, they can also lead to cross-hybridization of related genes, causing decreased specificity when trying to distinguish genes in a gene family (Li *et al.* 2002). Most importantly, it is often costly and challenging to validate, track, and maintain the cDNA clones, resulting in misidentification 10-30% of the time (Watson *et al.* 1998). In light of the disadvantages associated with cDNA technology and because the poplar genome has been sequenced, a cDNA platform was not used for this study.

Oligonucleotide microarrays

Oligonucleotide microarrays (Fig. 4.2) use synthetically manufactured probes that are synthesized directly or deposited mechanically onto a microarray chip or slide (Zhu 2003). Probes are designed based upon prior genome sequence knowledge so oligonucleotide microarrays are ideally suited for model species with a genome that has been sequenced. There are several companies currently marketing oligonucleotide

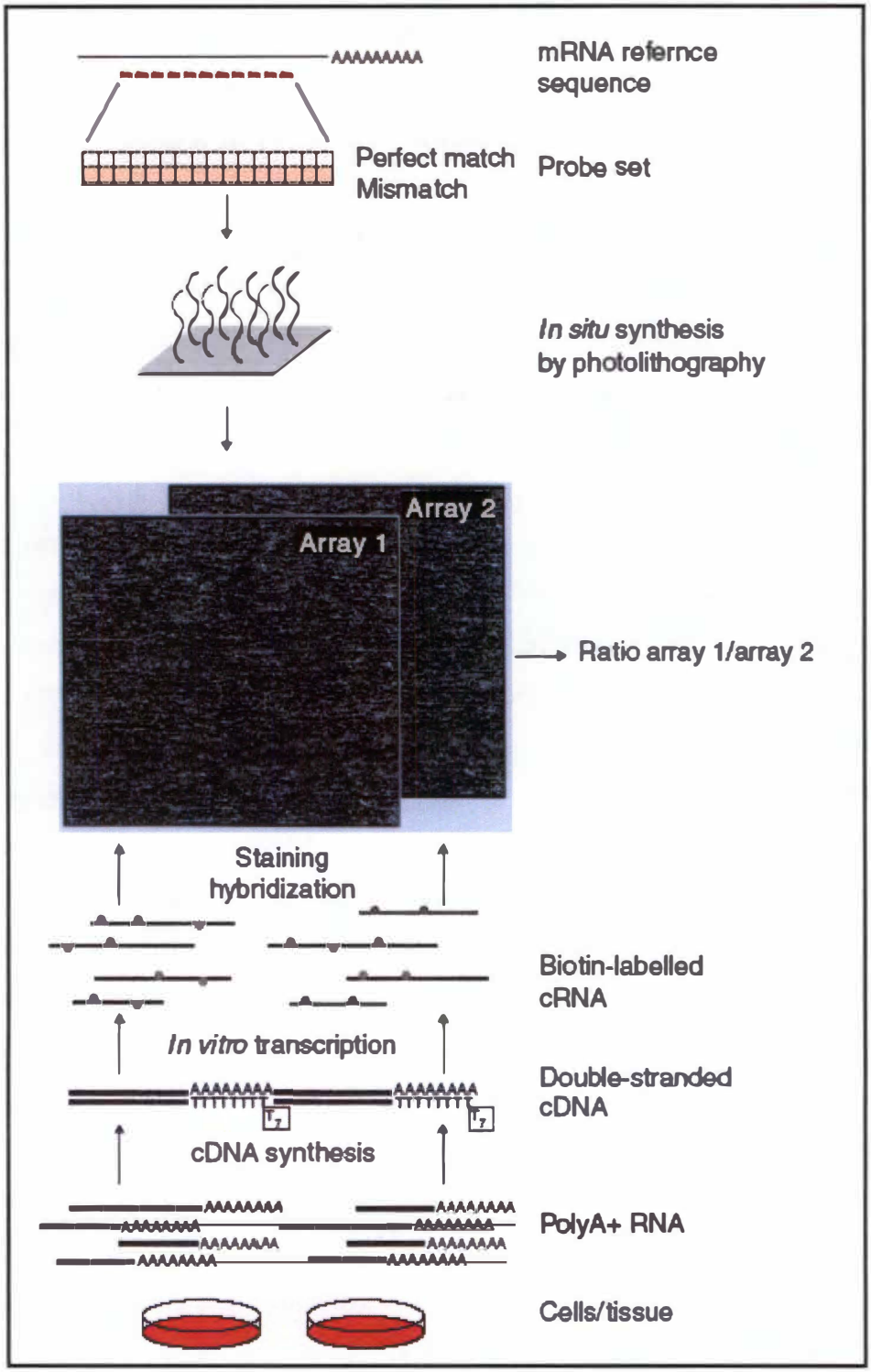

Figure 4.2: Example of oligonucleotide microarray
Adapted from Schulze and Downward 2001

microarrays including Agilent, Invitrogen, NimbleGen Systems and Affymetrix, all of which manufacture microarrays by directly synthesizing probes on the microarray slide surface. Another company, Applied Biosystems, mechanically deposits pre-synthesized oligonucleotides onto microarray slides. Though, the previously mentioned companies all have the ability to manufacture whole genome microarrays, they have all developed slightly different technologies to achieve this goal (Gershon 2004). A few examples of the different technologies, including both direct synthesis and mechanical deposition, will be discussed further.

Affymetrix GeneChip®technology

The most widely used oligonucleotide platform (Gershon 2004), commercially known as Affymetrix GeneChip® technology, utilizes photolithographic masks to synthesize 25-mer probes directly on a chip such that there are 11 probes that match the target perfectly and 11 probes that have a one base-pair mismatch relative to the target sequence. The mismatch probes serve as a form of control for hybridization specificity (Hoffman *et al.* 2004). The probes are spatially separated to allow for individual quantification and are hybridized with anti-sense copy RNA (cRNA) labeled with biotinylated ribonucleotides (Aharoni *et al.* 2002).

One Affymetrix GeneChip® chip can generally accommodate representative genes for an entire genome. Information about each of the genes is obtained using an algorithm that compares the differences between the match and mismatch averages for each oligonucleotide. Differences in the match and mismatch intensity scores are used to draw inferences about gene expression (Gibson 2002).

Affymetrix GeneChip® platforms have been used to study gene expression in a broad range of different organisms from humans to bacteria. For example, the technology has been used to gain a better understanding of the genetic changes that take place during prostate cancer (Magee *et al.* 2001) and breast cancer (Wilson *et al.* 2005) progression in the human body. The technology has also been used to elucidate the molecular mechanisms by which the bacterium *Caulobacter crescentus* and related species can live in low nutrient environments (Hu *et al.* 2005). Currently, an Affymetrix

platform is being used to design a high-density canine microarray (Holzwarth *et al.* 2005). Affymetrix GeneChips® have also been used quite extensively to study the molecular controls of plant growth and development. For example, gravitropic responses (Moseyko *et al.* 2002), pollen development (Pina *et al.* 2005 and Sze *et al.* 2004), plant hormone regulated gene expression (Müssig *et al.* 2002), ability to withstand salinity stress (Walia *et al.* 2005) and disease susceptibility (Ramonell *et al.* 2005) have all been studied using high-density Affymetrix GeneChips®.

An Affymetrix GeneChip® with 25-mer probes has several advantages associated with it. For example, probes are consistent in size across a chip and are designed specifically to discriminate between genes with closely related sequences. These two characteristics increase the chances that probes will bind to target molecules with greater specificity because optimum hybridization conditions can be calculated more accurately. As probes become longer and more variable in length across a slide, as is the case with cDNA microarray probes, it becomes increasingly difficult to optimize hybridization conditions leading to decreased specificity (Li *et al.* 2002). In addition, oligonucleotide arrays are more applicable to model species, often have greater genome coverage, and are more replicable and comparable across research groups (Gibson 2002). In contrast, there are also disadvantages associated with this type of microarray technology. Studies have indicated that relatively short probes (25-mer) tend to have poor hybridization properties when bound to a surface, making them less sensitive than cDNA probes (Lockhart *et al.* 1996, Shchepinov *et al.* 1997). This occurs because the closer an immobilized molecule is to a solid support the less likely it is that a given probe will freely come into contact with the target molecules to which the probe can potentially hybridize. Low sensitivity is especially problematic for studies involving low abundance targets because there is an even greater chance that the target will not come into contact with its corresponding probe (Shchepinov *et al.* 1997). Also, the inflexibility of the technology is becoming increasingly problematic because, while the demand is increasing, it is expensive and difficult to design new arrays (Hughes *et al.* 2001). This is especially true because a large number of photolithographic chromium masks are needed to control the addition of nucleotides as probes are being synthesized on a GeneChip (Singh-Gasson *et al.* 1999).

In light of the disadvantages associated with this technology, especially the characteristic decreased sensitivity of 25-mer probes, an Affymetrix platform was not used for this study.

Spotted oligonucleotide microarray technology

The second type of oligonucleotide microarray technology utilizes spotted/long oligonucleotide probes (Hughes *et al.* 2001). Long oligonucleotide probes are generally 50 to 75 bp in length and are synthetically designed based on knowledge of the target sequence of interest (Kane *et al.* 2000). After synthesis the probes can be modified by the addition of a 5' amino linker and covalently attached to pre-activated glass slides or they can be left unmodified and attached to glass slides ionically (Call *et al.* 2001). The spotting process is generally done with a liquid handling system (Hoffman *et al.* 2004). The immobilized probes are hybridized to labeled control and treated sample cDNA and analyzed with methods identical (or very similar) to those used for cDNA arrays (Barczak *et al.* 2003). Due to the length of long oligonucleotide probes they are more sensitive than short (25-mer) oligonucleotide probes (Hughes *et al.* 2001) because they can interact freely with the solution containing the cDNA molecules (Shchepinov *et al.* 1997). Specificity is also increased compared to cDNA array probes because oligonucleotide probes are designed to hybridize to a specific sequence of interest (Wang *et al.* 2003).

Spotted oligonucleotide microarray platforms have recently been used to study gene expression in several different types of organisms. For example, the technology has been used to develop a platform to study human gene expression (Barczak *et al.* 2003). More specifically, spotted oligonucleotide microarrays have been used to study novel pathways controlling the transformation of normal melanocytes to melanomas (Hoek *et al.* 2004), regulation of the p53 tumor suppressor gene (Zhao *et al.* 2000), detection and measurement of eukaryotic alternative splicing (Srinivasan *et al.* 2005) and expression profiling of cardiovascular disease development (Buermans *et al.* 2005). The platform has also been used to study zebrafish (Meijer *et al.* 2005), mice (Turk *et al.* 2004), bacteria (Denef *et al.* 2003), microbial populations (Tiquia *et al.* 2004) and plant

pathogens such as *Pseudomonas syringae* pv. *syringae* which is known to cause plant cankers, blights and leaf spots (Lu *et al.* 2005).

In addition to its increased sensitivity relative to Affymetrix GeneChips® and increased specificity relative to cDNA microarrays, spotted oligonucleotide technology has several other advantages associated with it. For example, the cost of the technology is lower than cDNA arrays when labor costs associated with obtaining and maintaining cDNA libraries are considered (Barczak *et al.* 2003). Also, probes are synthetically manufactured so spot concentrations are known and probe identification is easier to track, resulting in far fewer misidentification events. The technology is also highly flexible because the researcher is free to choose only the genes of highest interest to be represented on a microarray slide (Hughes *et al.* 2001). However, there are also disadvantages associated with spotted oligonucleotide technology. For example, the probe spotting process is automated, but spot positioning and deposition are not as accurate and reproducible as photolithography (Schulze and Downward 2001). It is also relatively costly, in terms of time and money, to synthesize and spot a large number of oligonucleotides onto glass slides, making the technology ill-suited for whole genome expression analysis experiments unless the genome of the organism is small (Hoffman *et al.* 2004). In light of the disadvantages associated with this microarray platform, especially because the objective was to do a whole genome expression analysis, a spotted oligonucleotide platform was not used.

NimbleGen™ microarray technology

The previously described microarray platforms are well known and all of them have been used extensively to do expression analysis studies in a wide variety of different organisms. However, they all have characteristic flaws associated with them. Additionally, scientists are becoming more interested in whole genome expression analysis as model organism sequence information becomes increasingly available (McCormick 2004). This has lead to the emergence of new high density platforms that are not currently as widely used, but are vastly improved when compared to existing high density microarray platforms (Singh-Gasson *et al.* 1999). One such platform is

manufactured by NimbleGen™ systems. NimbleGen™ microarray probes are synthetically manufactured and therefore fit into the oligonucleotide microarray category. NimbleGen™ arrays are similar to Affymetrix GeneChips® in that they are both high density platforms ideally suited for whole genome expression analysis. Probe synthesis is also carried out directly on a solid surface for both systems. However, the method used to do this is different on a NimbleGen™ slide. NimbleGen™ microarrays are built using Maskless Array Synthesizer (MAS) technology. The MAS system depends upon a Digital Micromirror Device (DMD) containing more than 700,000 tiny aluminum mirrors. Each mirror can be controlled and collectively the mirrors are used to create precise patterns of light which drive light-directed probe synthesis (Nuwaysir *et al.* 2002). To do this, the computer controlled mirrors create "virtual masks" by reflecting UV light onto precise slide locations. Each probe being synthesized has a UV-labile protecting group. When a particular nucleotide needs to be added, the mirrors reflect UV light onto the extending probe and the UV-labile protecting group is cleaved, allowing the next nucleotide to be added. This continues until all probes reach 60 nucleotides in length. The technology is such that 390,000 relatively long 60-mer probes can fit on a single array (Singh-Gasson *et al.* 1999).

Much like the previously mentioned technologies, NimbleGen whole genome microarrays have been used to study a variety of organisms. However, they have not been used as extensively as established technologies due to their relatively recent introduction. A few examples of studies done with NimbleGen microarrays include phenotypic screening of *Escherichia coli* (Winterberg *et al.* 2005) and the effect of histone H3 Lys 27 methylation on human polycomb target genes (Kirmizis *et al.* 2005). The technology has also been used on plants to study the effects of a chimeric AtMYB23 repressor on the growth and development of Arabidopsis (Matsui *et al.* 2005).

NimbleGen™ microarray chips have several advantages associated with them. In contrast to cDNA microarrays, it is easier to optimize NimbleGen™ chip hybridization conditions because the probes are consistent in length across the chip and sequences are known. As a result, hybridization specificity is increased compared to cDNA microarrays. Additionally, probes are synthesized directly on the chip surface at a

relatively high density making them well suited for high throughput whole genome analysis studies. Thus, NimbleGen™ microarrays were a better choice than spotted oligonucleotide microarrays for this particular study. Most importantly, the relatively long 60-mer probe length results in increased sensitivity relative to Affymetrix GeneChips® (Albert *et al.* 2002). Therefore, with the exception of the fact that NimbleGen™ microarrays arrays are relatively new and more expensive, there are very few disadvantages associated with the technology. As a result, NimbleGen™ microarrays were chosen to do whole genome analysis in this study. It should also be noted that, in June 2004, Affymetrix and NimbleGen became partners. As a result of the partnership, NimbleGen gained access to the established customer base of Affymetrix and Affymetrix gained access to NimbleGen's NimbleExpress array technology (Gershon 2004).

Experimental design

Although deciding what microarray platform to use for an experiment is important, there are several additional aspects of the technology that must be considered in order to ensure meaningful results. This is true because microarray experiments are complex multi-step processes. If these steps are performed accurately and consistently, the reliability and significance of experimental results will be greatly increased (Forster *et al.* 2003).

Experimental design

One of the most important steps in a microarray study is experimental design. If a study is poorly thought out in its beginning stages, it is unlikely that the results will be useful or reliable. To begin, the objective of the experiment should be well defined (Yang and Speed 2002). Once the objective of the experiment is clear, the statistical requirements to achieve the desired objective need to be considered. For example, the experiment needs to be designed in such a way that it has a sufficient statistical power and confidence level to produce meaningful results. This is accomplished by controlling variation throughout the steps of the experimental process. One way of controlling variation is by including the proper number of both biological and technical replicates so

that formal statistical methods can be used for analysis (Imbeaud and Auffray 2005). Equally as important as replication is remaining meticulously consistent while extracting RNA, labeling target samples and hybridizing the arrays. A preliminary study to determine the amount of variance that can normally be expected in a particular microarray experiment may also be necessary (White and Salamonsen 2005).

Gene probes and control probes

The probe sets representing a specific gene that will be included on an array can be one of two types. The first type of probe set is one in which probes have different nucleotide sequences but represent the same gene. These are known as probe variants and, because signal strength is generally sequence dependent, analysis becomes more complicated. Ideally, probe variants should be analyzed separately until the final data comparison if possible. The second type of probe set is one in which all probes representing the same gene have the same nucleotide sequence. These are known as probe replicates and, because they should all have the same signal strength, signal intensities can be averaged or analyzed using analysis of variance (ANOVA).

Microarrays should also include control probes, also known as external spikes, and fiducials. External spikes are synthetically produced RNA molecules that are added in known amounts to the biological target samples of interest. The data resulting from hybridization of the spikes to corresponding probes on the microarray can be used to determine the accuracy of the array results (van Bakel and Holstege 2004). Control probes should include an adequate number of both negative and positive controls. Negative controls can be used to indicate what a probe signal intensity should be if it does not hybridize with a target. The positive controls are generally representative of known housekeeping genes or are spikes. Positive controls are used to demonstrate the signal intensity of a range of expression levels by including a known concentration of spike targets in the target sample mixture (Forster *et al.* 2003). Fiducials are probes that are used as reference spots to mark the perimeter and center of the printed area on a microarray slide. These are used during image quantification to insure that probe grids are lined up correctly.

Microarray scanning and quantification

Microarray scanning, or image acquisition, is the process used to detect the fluorescent signals given off by hybridized probes on an array. This is usually done with a confocal laser scanner. However, there are several different types of scanners available and array design, slide type and spot morphology need to be taken into account when deciding which one is appropriate for a particular study. In addition, specific scanner settings are important and they need to be optimized to obtain the best possible representation of signal intensity and distribution on each slide (Forster *et al.* 2003). After a microarray image is scanned, it is converted to numerical data using a process called image quantification. Image quantification is done by transforming the signal intensity within a defined area of pixels into a defined unit. Transformation is usually done with software packages that are associated with the scanner used for image acquisition (Imbeaud and Auffray 2005). While microarray scanning can be optimized from slide-to-slide, image quantification should be done in exactly the same manner for each slide (Forster *et al.* 2003).

Normalization

Technical variation and measurement error between arrays can lead to non-biological effects, or noise, within a microarray data set (White and Salamonsen 2005). Normalization is a series of processes that are used to alleviate this problem by adjusting data means or variances. As a result, noise is reduced and significantly differentially expressed genes can be detected with greater accuracy (Forster *et al.* 2003). There are several ways to normalize microarray data and there is no universally accepted method of doing so (White and Salamonsen 2005). However, there are two general methods that are used, depending upon the original experimental design. These include ratiometric methods and absolute value methods. Ratiometric methods are used in two-dye experiments and involve the comparison of each array to a common reference. The comparison results in a ratio of expression for each gene that is $\log_2$ (ratio) transformed and used to represent relative changes in gene expression. In contrast, absolute value methods are used to normalize data from single-dye experiments because there is no

suitable common reference with which to calculate ratios. One absolute value normalization method is to calculate the 75th percentile value for each array and, from these, calculate a global mean known as a reference value. To get a final normalized intensity for each probe, the reference value is divided by the 75th percentile value for each individual array and multiplied by each probe intensity value on that particular array (Forster *et al.* 2003). Absolute value normalizations can also be done using a z-score normalization technique. Here, z-scores are calculated by subtracting the global or local mean intensity value from individual probe intensity values and dividing the result by the local or global standard deviation. The resulting z-scores for a chip or across chips, depending upon whether a global or local normalization has been done, have a mean of zero and a standard deviation of one. One does need to be careful when choosing whether to do a global (across all arrays in an experiment) or local (within arrays in an experiment) normalization. For example, a global normalization should not be done when signal intensities across arrays in an experiment have large differences because resulting z-scores are skewed by the presence of outliers (Cheadle *et al.* 2003).

Data analysis

After data normalization is complete, the next step is to identify differentially expressed genes based upon a measure of statistical significance (White and Salamonsen 2005). As is the case with the microarray study described in this paper, more involved analysis methods are necessary if more than one experimental factor is being considered. When this is the case, analysis of variance (ANOVA) is recommended rather than paired t-tests (Forster *et al.* 2003). This can be carried with statistical software such as SAS/STAT (Wolfinger *et al.* 2001). Once ANOVA is complete, clustering techniques can be used to identify patterns of gene expression (Quackenbush 2001) with the assumption being that genes with similar expression profiles are co-regulated (Forster *et al.* 2003). A few examples of the clustering methods that can be used are hierarchical clustering, k-means clustering and self-organizing maps. Hierarchical clustering is one of the most widely used techniques and it begins by putting each gene in its own cluster. A pairwise distance is calculated for all of the genes and the two with the most similar

distance matrix are clustered together. Distances are then calculated for the new clusters and the process is repeated until no new clusters are formed. K-means clustering can be used if there is a predetermined number of clusters that should be represented in the data. Initially, genes are randomly assigned to a cluster and then systematically moved to more appropriate clusters until moving anymore genes would make the clusters more variable. Self-organizing maps are similar to k-means clustering except for the fact that genes are assigned to a cluster based upon the similarity of their expression vector to a reference vector within each cluster (Quackenbush 2001). Software such as GeneSpring (Silicon Genetics) can be used to carry out each of the clustering methods mentioned above. Once genes with similar expression patterns are grouped together, functional information about the genes in each group can be used to elucidate the biological relevance of the data. Functional information can be gleaned from previously published journal articles, gene ontology databases and other similar sources (Imbeaud and Auffray 2005).

Microarray Validation

One of the most important steps when working with microarray data is verification of results to ensure that conclusions drawn from a study are accurate. This is especially important when working with microarrays because the introduction of artifacts is possible during every step of the experimental process. In addition, microarray results can be further skewed by cross-hybridization of probes with non-target sequences. Though validation can be by comparing results with information that is available in previously published literature or databases, it is best to validate results using either real-time PCR or northern blots (Chuaqui *et al.* 2002).

Materials and Methods

Plant material, experimental design and treatment

Hybrid poplar H11-11 (*P. trichocarpa x P. deltoides)* trees were grown, treated and sampled at the University of Florida, Gainesville, FL. Plants were clonally propagated from rooted softwood cuttings and were grown in a controlled greenhouse environment for about seven weeks. A completely randomized design (CRD) was used

and the experiment took place in January 2005 when the average height of the plants was 60 cm. Six plants in total were used in this particular experiment. Two were foliar sprayed once with about 50ml of 0.5 mM NaOH and roots were harvested after one hour. Two were foliar sprayed once with 100 μM of the synthetic auxin NAA dissolved in 0.5 mM NaOH and roots were harvested after one hour. The last two were foliar sprayed once with 100 μM of NAA dissolved in 0.5 mM NaOH and roots were harvested after 24 hours. A synthetic auxin was used rather than an endogenous auxin because synthetic auxins are more stable compounds making them more appropriate for experiments requiring exogenous application. Additionally, NAA acts similarly to IAA upon uptake (Ribnicky *et al.* 1996). All plants were treated at the same time and, in all cases, harvested roots were frozen in liquid nitrogen immediately after harvest. Samples were transferred to a -80°C freezer until they were processed further.

RNA isolation, target synthesis and labeling

Total RNA was extracted from the six root samples using an RNeasy® Plant Mini Kit according to the protocol provided by Qiagen, with one exception. To increase the RNA yield, 22μl of 20X polyethylene glycol (PEG) 8000 was added to the lysis buffer for each sample. PEG has been shown to increase RNA yield by binding polyphenols and polysaccharides which otherwise bind and inhibit RNA precipitation (Gehrig *et al.* 2000). After the initial RNA extraction each sample was concentrated using a Qiagen Rneasy ® MinElute™ Cleanup Kit according to the protocol supplied by the company (Qiagen 2003). RNA quantity was measured using a Nano-drop® ND-1000 Spectrophotometer and stored at -80°C until they were needed for target synthesis. Target amplification, labeling, purification and fragmentation were carried out according to the instructions provided by Ambion using a MessageAmp™ II-Biotin *Enhanced* Kit. A uniform amount, 1 μg, of total RNA was used from each root sample. During the synthesis process for the particular system used, antisense amplified RNA (aRNA) is synthesized using biotin-modified UTP resulting in biotin-labeled target aRNA. The biotin label used in a later step to facilitate conjugation to alexa fluor 555 dye. This process was used to synthesize labeled target aRNA for each of the six root samples. In

addition, another three target samples were synthesized after pooling equal amounts of total RNA (0.5 ug of each) from the two control root samples, the two samples from roots harvested one hour and the two samples from roots harvested after 24 hours.

Microarray construction

Populus whole-genome microarrays were used for this study. Probes on the microarray were designed to represent all putative *Populus* nuclear and organellar genes. Additionally, probes representing aspen transcripts and micro-RNA (miRNA) were also present but are not relevant to this particular study. All genes on the microarray slides were represented, generally, by a total of three 60-mer probes, each having a different nucleotide sequence. The microarrays were manufactured by NimbleGen™ on glass slides after the completion of probe design based upon predetermined uniqueness criteria. The total number of probes on each slide, including genes, positive and negative controls and reference fiducials, equaled 190,000 (Tuskan *et al.* 2006). However, the probes relevant to this study included only the poplar genes (55,969) and a negative control.

Probe hybridization, dye conjugation and slide washing

Due to the cost of the technology, only six arrays were hybridized in total (Fig. 4.3). The arrays were hybridized according to a protocol from NimbleGen™ that was optimized at ORNL. A 40µl pre-hybridization solution made up of a final concentration of 0.1µg/µl herring sperm DNA, 0.5µg/µl acetylated bovine serum albumin (BSA) and 1X MES hybridization buffer was denatured using a GeneAmp® PCR System 9700 (Applied Biosystems) at 65°C for 5 minutes, 95°C for 5 minutes and then held at 45°C until it was needed. The microarray slide was placed in the bottom portion of a hybridization chamber and two small wells on either end of the chamber were filled with 14ul sterile RNase free water. A cover slip was placed over the printed area of the slide and the pre-hybridization solution was dispensed under the cover slip. Assembly of the air and watertight hybridization chamber was completed, placed in a 42°C water bath for 15 minutes, washed briefly in a falcon tube containing 50 ml of sterile type I water to

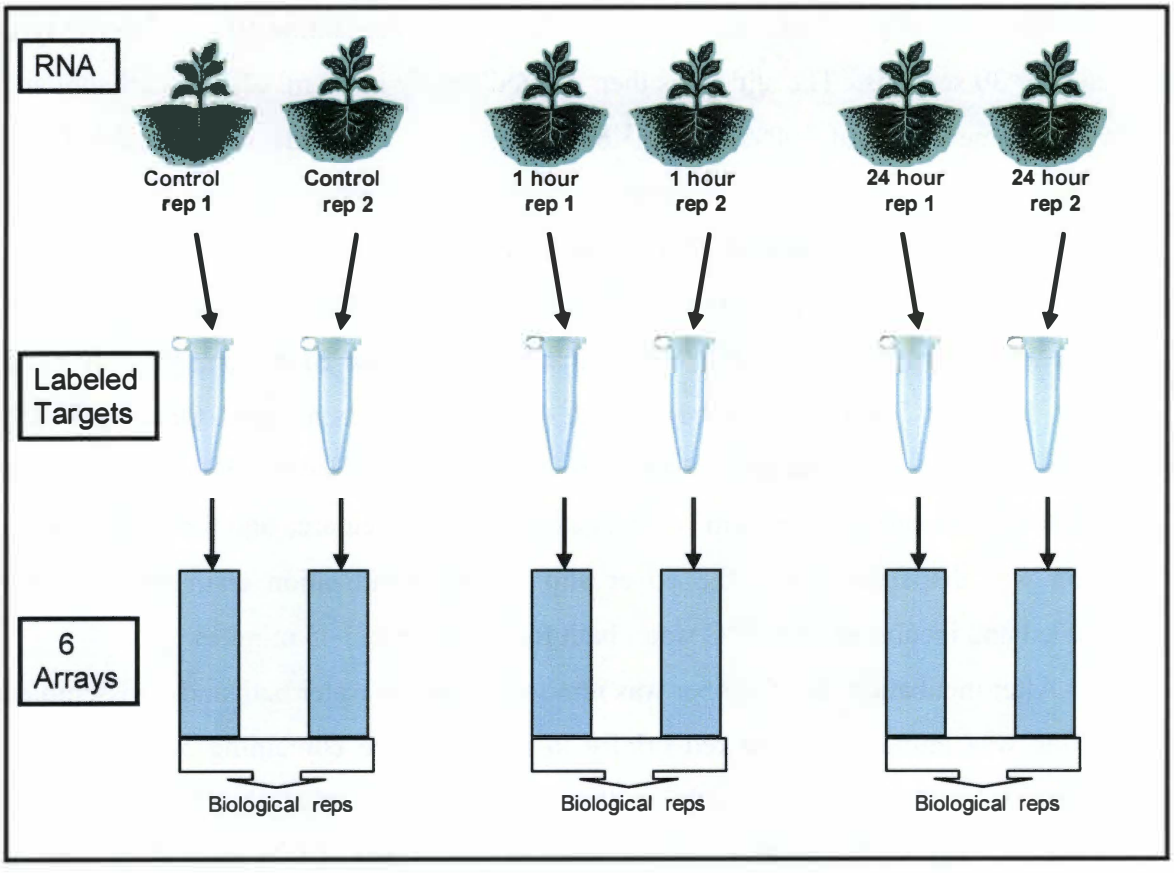

Figure 4.3: General experimental process
Plants were treated, harvested at the appropriate time and RNA was extracted from the roots. Biotin labeled targets were synthesized, amplified and hybridized to microarray slides.

remove the cover slip and then moved to new falcon tube containing 50 ml of sterile type I water for 30 seconds. The slide was then dunked briefly in 50 ml of 100% ethanol and moved to a new tube of 100% ethanol for 30 seconds. The slide was then dried in a centrifuge at 1500 rpm for about 2 minutes. A 40µl hybridization solution containing 10 µg of fragmented biotin labeled aRNA, 1nM CPK6 oligonucleotide (for the reference fiducials), 1µg/µl herring sperm DNA, 0.5 µg/µl BSA and 1X MES hybridization buffer was denatured using a GeneAmp® PCR System 9700 (Applied Biosystems) at 65°C for 5 minutes, 95°C for 5 minutes and then held at 45°C until it was needed. Once again, the microarray slide was placed in the bottom portion of a hybridization chamber, the wells were filled with water, a cover slip was placed over the printed area and the hybridization solution was dispensed under the cover slip. The hybridization chamber was then assembled and incubated in a 42°C water bath for 16 hours and 45 minutes.

After incubation, the chamber was removed from the water bath and disassembled. The slide was immediately dunked briefly in a falcon tube containing 50 ml of non-stringent wash buffer (NSWB) made up of a final concentration of 6X SSPE and 0.01% Tween-20 to remove the cover slip. The slide was then moved to a new 50 ml tube of NSWB, the tube was capped and the slide was gently agitated for about one minute. The slide was then moved to a 50 ml tube of 42°C stringent wash buffer (SWB) made up of a final concentration of 1X MES buffer, 0.026M NaCl and 0.01% Tween-20 and incubated at 42°C. The slide was gently agitated every five minutes for 15 minutes, moved to new SWB and gently agitated every five minutes for 15 minutes. After incubating in the SWB buffer the slide was removed and placed in fresh NS buffer for about one minute. The slide was taken out and placed printed side up in a sterile Petri dish. One ml of stain solution (1X stain buffer, 2µg/µl BSA and 0.01µg/µl Alexa Fluor 555 dye) was dispensed over the printed area of the slide. The slide was agitated every few minutes for 15 minutes at room temperature and placed in fresh NS buffer for about one minute. The slide was then placed in 50 ml of chilled final was buffer (50% NS buffer and 50% sterile type I water) for exactly 30 seconds. The slide was immediately dried by centrifuging for two minutes at 1500 rpm.

Slide scanning and quantification

Slides were scanned using a ScanArray Express manufactured by Perkin Elmer at a 5μm resolution with PMT gain of 65% and a laser power of 90%. A wavelength of 543nm was used to detect Alexa Fluor 555 dye fluorescence. Images were saved as TIFF files, imported into NimbleScan version 2.1 (NimbleGen™ systems, Inc.) and cropped to include only the printed area. The manual alignment option was used to align the imported images with the *Populus* whole-genome slide design file provided by NimbleGen. Alignment was accomplished using four corner fiducials located at each corner of the printed area. NimbleScan was then used to generate a feature report and a pair report from each of the aligned images. The feature report contains a description of every non-empty feature in the design, while the pair report contains raw signal intensities for all experimental and control probes within the array. The pair reports were then used for further statistical analysis.

Statistical analysis

The purpose of the differential expression experiment was to elucidate genes that are differentially expressed 1 hour and 24 hours after a one-time exogenous auxin treatment as compared to genes expressed in untreated control plants and each other. Six arrays were used in the differential expression experiment including two biological reps for each of the three treatments. More specifically, two arrays were hybridized to labeled target aRNA from two different control plants, two arrays were hybridized to labeled target aRNA from two different plans harvested 1 hour after exogenous auxin treatment and two arrays were hybridized to labeled target aRNA from two different plants harvested 24 hours after exogenous auxin treatment.

Pair reports from each of the six arrays generated with the NimbleScan software were imported into JMP version 6 software (SAS Institute Inc., Cary, NC, 1989-2005), a statistical analysis package. Non-relevant probe data was removed from each pair report, leaving 164,531 raw intensity values per pair report. Normalization was carried out within each array by converting each of the raw intensity values to a z-score to adjust the

mean intensity value of each array to approximately zero and the standard deviation of each array to approximately one.

In order to identify lower limit outliers within the experimental dataset, a separate analysis was done on a negative control contained on all of the arrays (12 arrays in total). The negative control probe was spotted 12 times on each array resulting in 144 data points in the analysis. The mean and standard deviation of the negative control were calculated using SAS/STAT® software, version 9.1 of the SAS System for Windows XP-Professional (SAS Institute Inc., Cary, NC, 2002-2003), and were -0.32 and 0.67, respectively. Lower limit outliers within the experimental dataset were determined to be those values that were two standard deviations below the mean of the negative control. As a result, 5298 values within the dataset less than or equal to -1.7 were excluded from further analyses. In order to identify upper limit outliers within the experimental dataset the mean and standard deviation of all experimental probes, 987,042 in total, were calculated using SAS and were -0.06 and 0.95, respectively. Upper limit outliers were set at approximately 10 standard deviations higher than the overall mean. In addition, if a probe representing the same gene exceeded this threshold value on more than one array, it was kept in the analysis. As a result, 156 values greater than or equal to 10 standard deviations occurring only once among probes representing the same gene were excluded from further analysis.

After lower and upper limit outliers were removed from the experimental dataset, per-chip normalizations were done again to remove the effect of outlier values on individual z-scores. The decision to do per-chip normalizations rather than a global normalization across all chips was based upon the fact that positive control means were more consistent across chips when per-chip normalizations were done as compared to a global normalization. The data from all six files was then merged into one file for further analysis in SAS/STAT, a statistical analysis software package. SAS/STAT was used to do a mixed model analysis of variance (ANOVA). A completely randomized design was used for the microarray experiment and is represented in the following mathematical model:

Initial model: $Y_{ijk} = \mu + T_i + G_j + T*G_{ij} + R_k + T*R_{ik} + G*R_{jk} + T*G*R_{ijk}$

Pooled into E_{ijk}

Final model: $Y_{ijk} = \mu + T_i + G_j + T*G_{ij} + E_{ijk}$

T = time (fixed effect)
G = gene (random effect)
R = plant/rep (random effect)
T*G = interaction between time and gene (random effect)

SAS/STAT was then used to do an LSD mean separation to test the following null hypotheses:

H_o: $\mu_c = \mu_1$
H_o: $\mu_c = \mu_{24}$
H_o: $\mu_1 = \mu_{24}$

μ_c is the mean probe intensity of gene x for the control
μ_1 is the mean probe intensity of gene x for the 1 hour treatment
μ_{24} is the mean probe intensity of gene x for the 24 hour treatment

The null hypothesis for a comparison was rejected and the difference between the two means was considered to be significantly different if the probability value (p-value) was ≤ 0.01. A p-value measures the false positive rate, or rate at which the null hypothesis is rejected when it is actually true, that occurs when a particular comparison is called significant. As can be imagined due to the large number of comparisons being made, even if the significance cut-off is set relatively low ($p \leq 0.01$), a large number of potential false positives may be generated. To reduce the number of false positives among the comparisons thought to be significantly different, a second statistical analysis was done using GeneSpring 7.2 (Silicon Genetics) software. Raw data (with upper and lower limit outliers removed) was imported into GeneSpring and per-chip normalizations were done by normalizing each chip to the 50^{th} percentile of the measurements taken from that chip. Statistical analysis was done in GeneSpring using the ANOVA option where variances are assumed equal and the cut-off was set to $p \leq 0.01$. The same three hypothesis described above were tested using GeneSpring. After obtaining a list of differentially expressed genes using both SAS/STAT and GeneSpring, the two lists were compared and genes that were significantly differentially expressed according to *both* analyses were retained and used for further analysis. Based upon the mean signal

intensity for the probes representing each gene in a comparison, lists were further divided so that they contained genes that were up-regulated or down-regulated in response to auxin for that particular comparison.

Cluster analysis

Significantly up and down-regulated genes were clustered according to expression behavior using GeneSpring 7.2 (Silicon Genetics) software. In order to identify the most significant patterns within the set of significantly differentially expressed genes, a principle component analysis (PCA) was done. Based upon the results of the PCA, genes were clustered via k-means clustering into three groups using a standard correlation. Hierarchical clustering was also carried out on the dataset using a standard correlation.

Real-time reverse transcription PCR

RNA extraction

The same RNA that was extracted to synthesize target pools for microarray analysis was also used to synthesize cDNA to do real-time RT-PCR. Total RNA was extracted from the six root samples using an RNeasy® Plant Mini Kit according to the protocol provided by Qiagen, with one exception. To increase the RNA yield, 22µl of 20X polyethylene glycol (PEG) 8000 was added to the lysis buffer for each sample. PEG has been shown to increase RNA yield by binding polyphenols and polysaccharides which otherwise bind and inhibit RNA precipitation (Gehrig *et al.* 2000). After the initial RNA extraction each sample was concentrated using a Qiagen RNeasy ® MinElute™ Cleanup Kit according to the protocol supplied by the company (Qiagen 2003). RNA quantity was measured using a using a Nano-Drop® ND-1000 Spectrophotometer and stored at -80° C until they were needed for real-time RT-PCR.

cDNA synthesis

One µg of total RNA from each sample was treated with DNase 1 (two units/µg) at 37°C for 10 minutes to remove DNA contamination before cDNA synthesis. DNase stop solution was added and samples were incubated according to instructions provided

by DNase usage information (Promega 2005). One µg of DNase treated RNA was used for cDNA synthesis (7.7µl). SuperScript ™ III Reverse Transcriptase was used to synthesize cDNA according to the instructions supplied by Invitrogen ™ (Invitrogen 2003) using random hexamer primers. Reverse transcription was performed on an Applied Biosystems Gene Amp® PCR System 9700 for 2 hours at 50 °C. The reaction was terminated by incubating at 85 °C for 5 minutes. After cDNA synthesis each sample was diluted 1:1 with RNase and DNase free H_2O.

Real-time PCR

Real-time PCR 25µl reactions were done using iQ™ SYBR® Green Supermix according to instructions provided by Bio-Rad Laboratories on an iCycler Real Time PCR detection system (Bio-Rad Laboratories 2005). Prior to mixing reaction reagents, cDNA concentrations were quantified using a Nano-Drop® ND-1000 Spectrophotometer and each sample was diluted to the same concentration (500ng/µl) to minimize variation that can result from inputting different starting amounts of cDNA. One µl of cDNA sample was used in each reaction. Gene specific forward and reverse primers were designed based upon gene model sequence of two genes of interest. Primers were designed for a total of 18 significantly differentially expressed genes (Table 4.1) and were screened for uniqueness by blasting each primer sequence against the JGI poplar gene model database. The gene used as a control to normalize the data for differences in input RNA and efficiency of reverse transcription between the samples was an 18s housekeeping gene. Ten uM each of the appropriate forward and reverse primer were added to each reaction. Three reps were done for each reaction to give a total reaction number of 54 (6 samples x 18 genes x 3 reps = 324).

Data analysis

C_t values of the three reps were averaged within each sample and treatment type. The rep averages of the 18s controls were then averaged *across all* treatment types to get an overall 18s average based on the fact that 18s is expressed at a constant level regardless of treatment. Average 18s control values within tissue types were subtracted

Table 4.1: List of genes used for microarray analysis validation

Category	Description	Primers F/R
1 hr up-regulated vs. control	FID	5'-GGATACCTATCGATACTGGACTCG-3'
		5'-TTGTGCTTCCTAGTGACACCTC-3'
	PROT	5'-ATTGGTGTCATTCGGCACTC-3'
		5'-TGGAATGATGAGAAGAACGAG-3'
	E2	5'-ATGGTTGCTTCACCATGTCAG-3'
		5'-CATGCTGCTTCGCCATAG-3'
1 hr down-regulated vs. control	ZIP	5'- CCGCACCTTTCCTATTTCC-3'
		5'-CGTAAATGTTGTTGTGTGAGACTG-3'
	PS2	5'-CAAGACGTCATCAACTCCTACAATC-3'
		5'-TGGCTCGGCTATCATTCAC-3'
	HYP	5'-GCGAACAGCTTGTGGGTTC-3'
		5'-AGGTTTATGACCCTTCCCAAC-3'
24 hr up-regulated vs. control	ELON	5'-AAGCAACTTGGTTACCTTGAGAG-3'
		5'-AGCATACCAAGTTCATCACAACAC-3'
	DIOX	5'-GCAGCCTGAAGAACTGCTG-3'
		5'-CTGTTCAAACACAGGAACATTGC-3'
	PUT	5'-AAGTCGCCAAGGTTGGAAG-3'
		5'-CTTCTAACTGTCAACACCACACC-3'
24 hr down-regulated vs. control	WRKY	5'-AACGTATTGCATCGATCTTGG-3'
		5'-TAAGAACGCATCCTCATAGCAC-3'
	FAS	5'-TTCGTTGGTGGTGTGAACTTG-3'
		5'-GGATTGACCACGTACCTAACAG-3'
	FERR	5'-TTGCAGTCAGAATTGAATAGGG-3'
		5'-TGGGAACATTTCCACATCG-3'
24 hr up-regulated vs. 1 hour	EAUX	5'-CAACATATTGAACTCGGGAGAA-3'
		5'-GCAGGATGGATTTAATCGTGA-3'
	SER	5'-CGCATGTAAATGTCATCGGTAG-3'
		5'-GTTGCTATGGCCAAAGTGAAAT-3'
	COP	5'-GCAGCCTTAACACCCTCTTTC-3'
		5'-AACGAAAGCCCAGACAGATAG-3'
1 hr up-regulated vs. 24 hour	RING	5'-TTTGGGTTGTACTGTGAAGCTG-3'
		5'-GCAATAACTGGCCCGAAG-3'
	UNK	5'-GCACGGCTGACTGCTAAAG-3'
		5'-CACATGAAACCGTGGAACAG-3'
	PUT2	5'-CCAATCCCACACAGAACAATC-3'
		5'-GAGAATTGAAGAACTCAAGGGAAAG-3'

from the control average across tissue types, resulting in a cycle correction value for each tissue. The cycle correction values were then added to each rep average for the genes of interest and normalized C_t values were compared (Brunner *et al.* 2004). All of the products were stained with ethidium bromide and run on an agarose gel to make sure the primers were gene specific and produced only one band.

Auxin binding domains

In order to see if the differentially expressed genes contained a higher number of auxin transcription factor binding domains than would be expected by chance, a chi-square analysis was done. Genes within a subset of the differentially expressed genes, the 18 genes used in real-time PCR analysis, were queried in the JGI *P. trichocarpa* v1.1 genome browser. Sequence 2000 base pairs upstream of the UTR or coding region of each gene was queried in PLACE (Higo *et al* 1999), a database of plant cis-acting regulatory DNA elements. The number of predicted auxin binding domains was recorded for each gene and included ARFAT (TGTCTC), an ARF binding site found in the promoters of primary response genes (Ulmasov *et al.* 1999), AUXRETGA1GMGH3 (TGACGTAA), a putative auxin transcription factor binding site (Liu *et al.* 1997) and AUXREPSIAA4 (KGTCCCAT), an auxin responsive element found in pea root tip meristems (Ballas *et al.* 1993). The same procedure was repeated for a random sampling of genes that were not differentially expressed according to SAS/STAT and GeneSpring analysis. Two Chi-square analyses were done. The first was a goodness-of-fit Chi-square where a gene was counted as having an auxin transcription factor binding site if it had at least one of the three binding sites included in the analysis. The expected ratio was calculated from the randomly chosen non-significant genes. A second calculation was done where the number of transcription factor binding sites was taken into account. For this calculation, the expected value was the average number of auxin transcription factor binding sites amongst the randomly selected non-significant genes.

Probe non-target sequences

In an analysis carried out by Dr. S. DiFazio (comm. 2006), it was found that a large number of probes on the microarray are not unique and as a result have a high probability of cross hybridizing to one or more non-target sequences. It can be expected that those probes with a sequence highly similar to a non-target sequence will consistently have a higher signal on the microarray. We therefore examined the relationship between the potential number of non-target sequences of each probe and the corresponding normalized intensity on the microarray.

Results

Statistical analysis

Statistical analysis using SAS/STAT indicated that the number of significantly differently expressed genes according to a $p \leq 0.01$ was 7184 for the 1 hour versus control comparison, 897 for the 24 hour versus control comparison and 1913 for the 1 hour versus 24 hour comparison. Statistical analysis using GeneSpring indicated that the number of significantly differentially expressed genes at a $p \leq 0.01$ was 1025 for the 1 hour versus control comparison, 475 for the 24 hour versus control comparison and 662 for the 1 hour versus 24 hour comparison. Subsequently, as only the comparisons found to be significantly differential using both SAS/STAT and GeneSpring were retained, the number of significantly differentially expressed genes in the final dataset was 381 for the 1 hour versus control comparison, 77 for the 24 hour versus control comparison and 211 for the 1 hour versus 24 hour comparison. Based upon the mean signal intensity for the probes representing each gene in a comparison, lists were further divided so that they contained genes that were up-regulated or down-regulated in response to auxin for that particular comparison. This resulted in 108 genes up-regulated and 273 down-regulated genes when comparing the 1 hour treatment to the control, 36 genes up-regulated and 41 genes down-regulated when comparing the 24 hour treatment to the control and 71 genes up-regulated and 140 genes down-regulated when compared to the 24 hour treatment.

Cluster analysis

PCA carried out using GeneSpring software revealed three expression profile trends contributing to a combined total of 100% of the variability within the data (Fig. 4.4). Subsequent k-means clustering of the significantly differentially expressed genes into three clusters was used to categorize 183 genes into one cluster, 299 genes into a second cluster and 134 genes into a third cluster (Fig. 4.5). Hierarchical clustering was also done and did not appear to correlate well with the k-means clustering results despite the fact that a standard correlation algorithm was used for both clustering methods (Fig. 4.5).

Real-time reverse transcription PCR

Real-time RT-PCR was done on a sampling of significantly differentially expressed genes to determine whether or not results from the microarray analysis were accurate. A total of 18 genes were chosen including three up-regulated genes in the 1 hour versus control treatment comparison, three down-regulated genes in the 1 hour versus control treatment comparison, three up-regulated genes in the 24 hour versus control treatment comparison, three down-regulated genes in the 24 hour versus control treatment comparison, three up-regulated 24 hour treatment genes from the 24 hour versus 1 hour treatment comparison and three up-regulated 1 hour treatment genes from the 1 hour versus 24 hour treatment comparison.

Results from real-time RT-PCR analysis indicate that only 5 of the 18 differentially expressed genes tested behaved similarly to the expected trend demonstrated by microarray analysis. These genes included one gene (ZIP) down-regulated in the 1-hour versus control comparison (Fig. 4.6), one gene (DIOX) up-regulated in the 24-hour versus control comparison (Fig. 4.7), two genes (WRKY and Ferr) down-regulated in the 24-hour versus control comparison (Fig. 4.7) and one 24-hour gene (COP) up-regulated in the 24-hour versus 1-hour comparison (Fig. 4.8). It

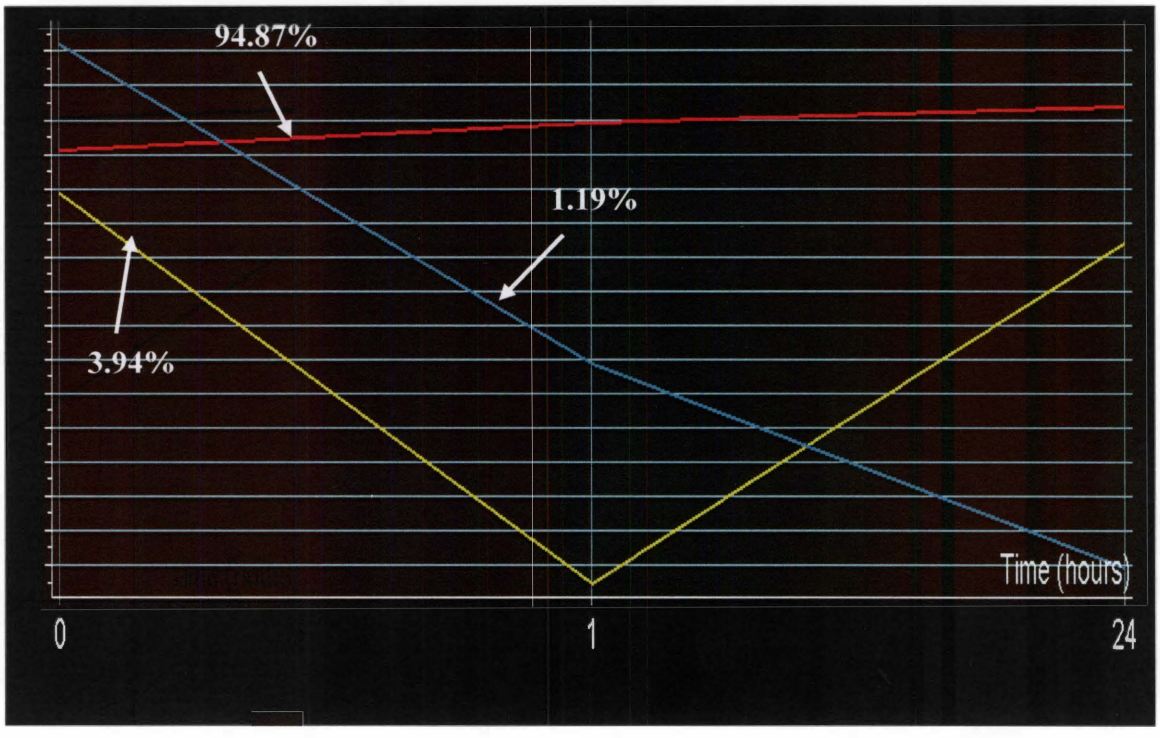

Figure 4.4: PCA of significantly differentially expressed genes

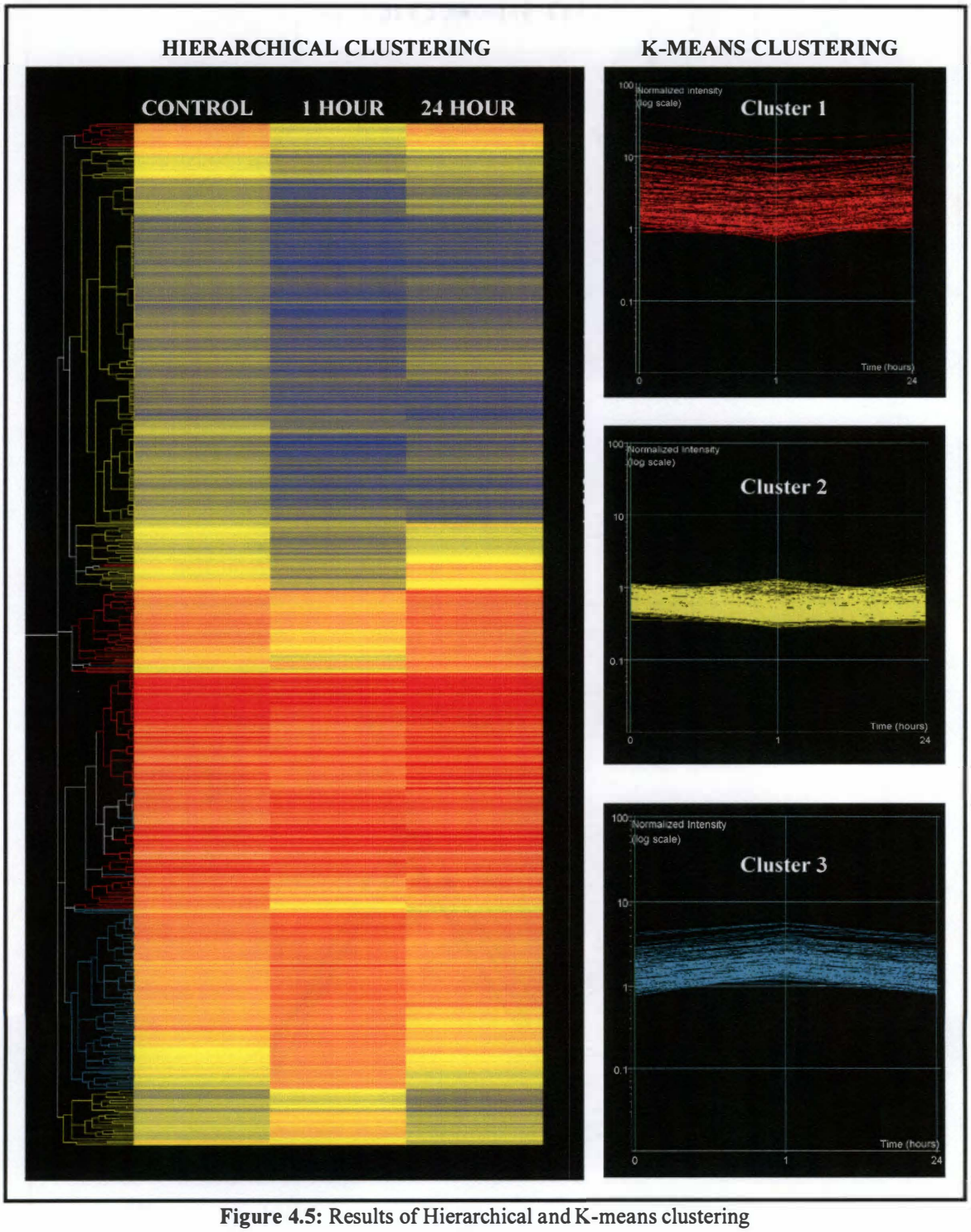

Figure 4.5: Results of Hierarchical and K-means clustering

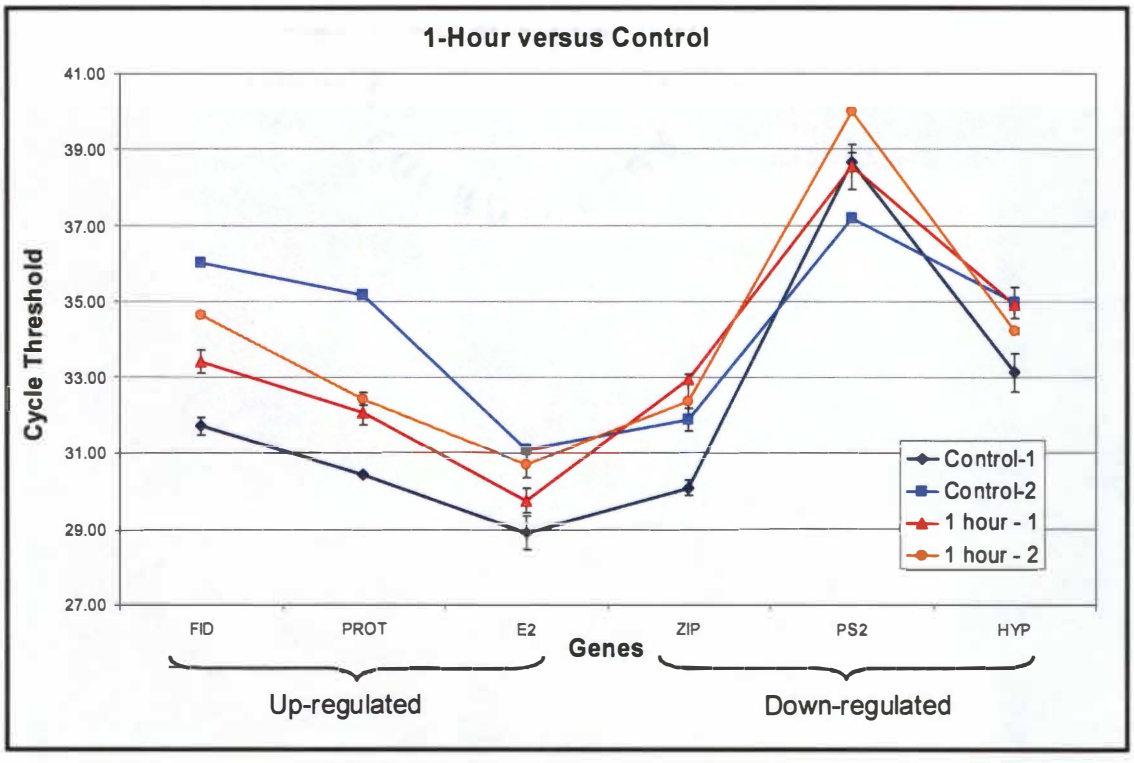

Figure 4.6: Real-time RT-PCR 1-hour versus control comparison

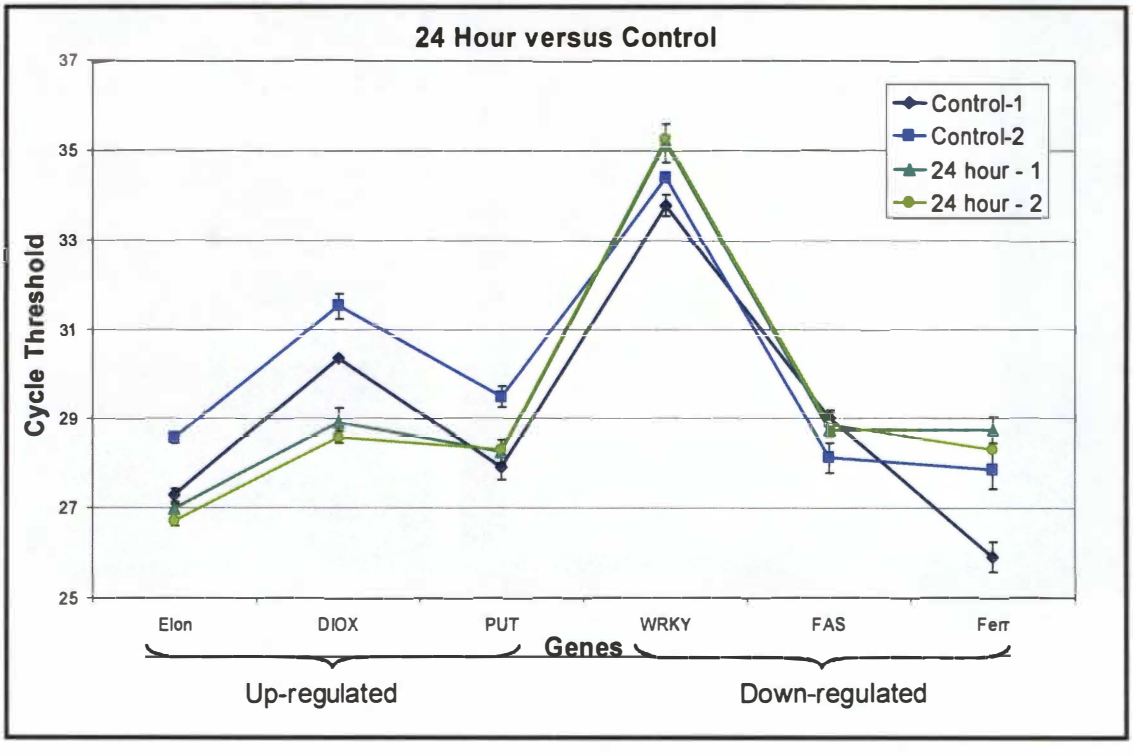

Figure 4.7: Real-time RT-PCR 24-hour versus control comparison

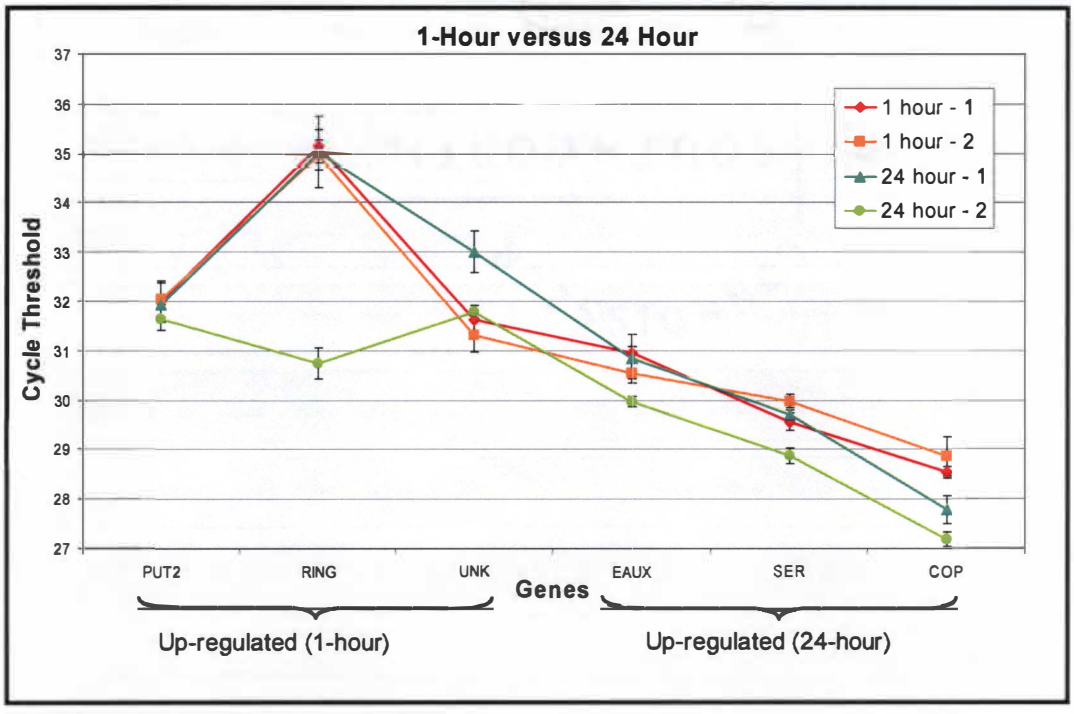

Figure 4.8: Real-time RT-PCR 1-hour versus 24-hour comparison

should also be noted that, though these 5 genes demonstrated the same *trend* as expected based on the microarray results, only one gene (DIOX) strongly validated microarray results as it was the only gene that, when cycle threshold values were averaged across reps, had a 4.88 fold higher expression level in the 24-hour treated plants compared to the control plants. The remaining four genes had a fold change of 2.8 or less when cycle thresholds were averaged across reps. Additionally, it was also noted that, in the 1-hour versus control comparison, the two control treatment reps acted in an opposite manner according to real-time RT-PCR results (FID, PROT and E2 - Fig. 4.6). Real-time RT-PCR reactions were repeated for the 1-hour versus control comparison to make sure these results were valid and the same outcome was observed (data not shown).

The resulting end products of all real-time RT-PCR reactions were run on agarose gels stained with ethidium bromide to verify that each primer set yielded single band products, as expected, for each tissue (Fig. 4.9, 4.10 and 4.11). All of the primer sets yielded a single product except FAS (Fig. 4.10). Therefore, it should not be used in the future.

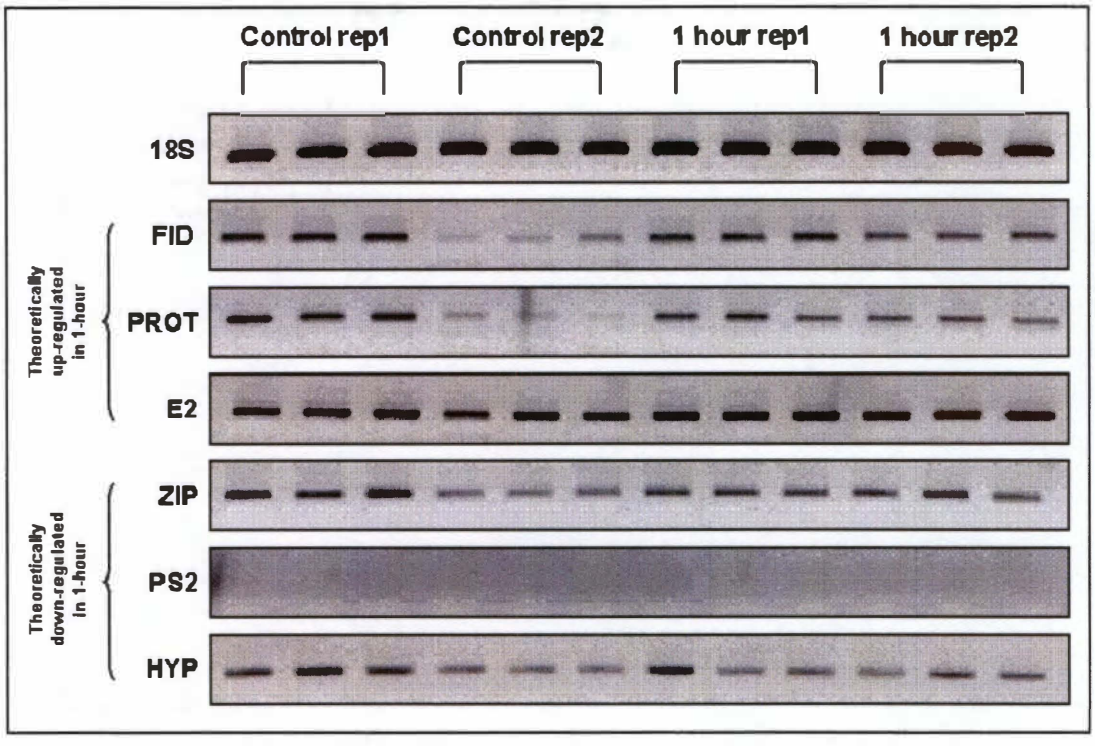

Figure 4.9: Real-time PCR end product for 1-hour versus control comparison

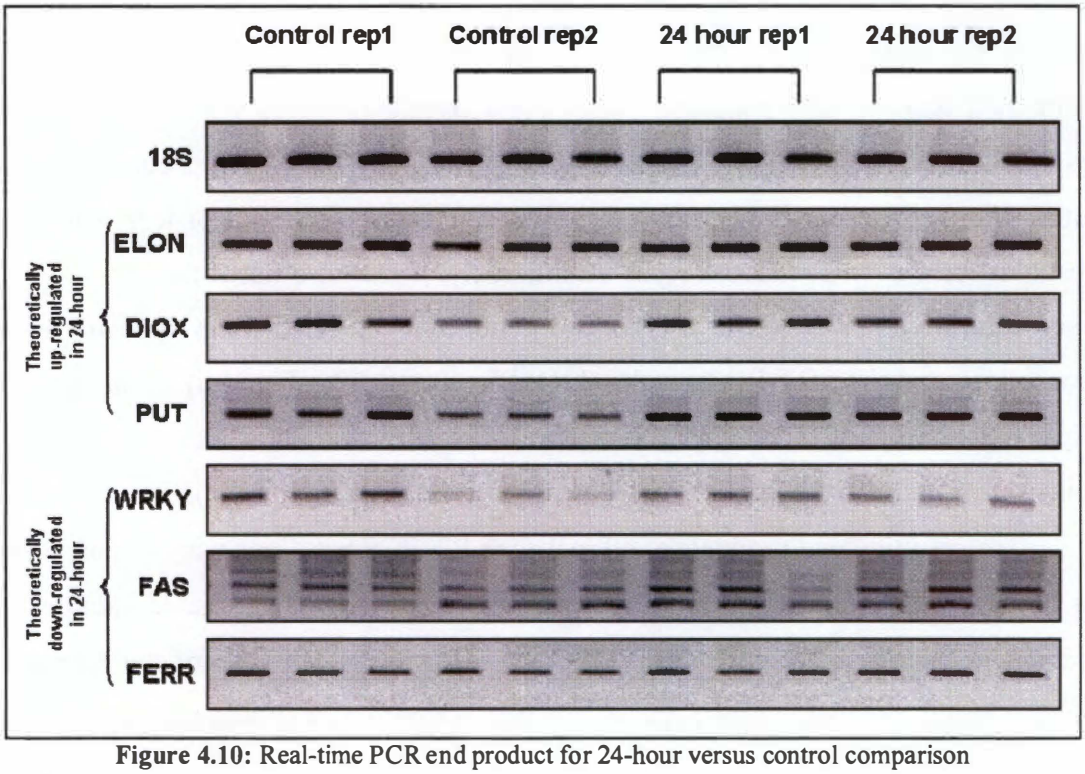

Figure 4.10: Real-time PCR end product for 24-hour versus control comparison

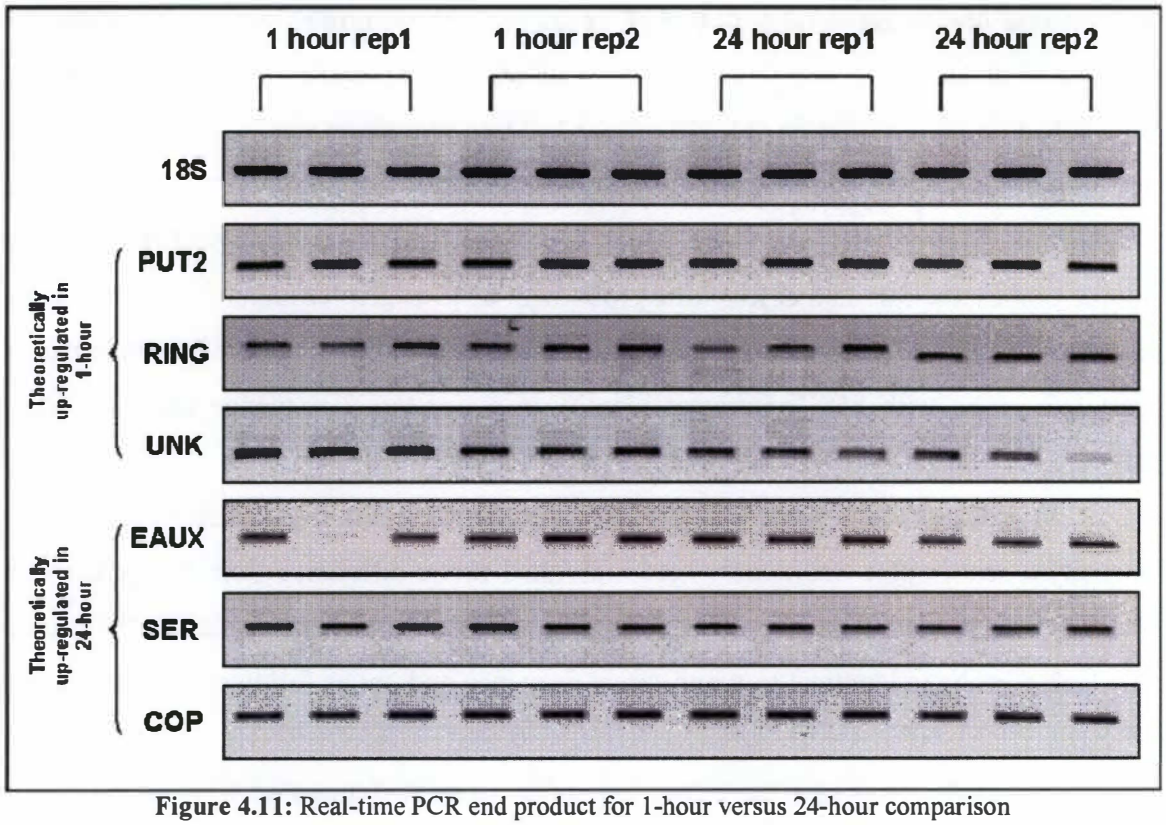

Figure 4.11: Real-time PCR end product for 1-hour versus 24-hour comparison

Auxin binding domains

Results of a goodness of fit chi-square analysis where the expected ratio of auxin transcription factor binding sites was 1:1 indicated that there does appear to be a higher number of auxin transcription factor binding sites in the significantly differentially expressed genes than would be expected by chance at a $p \geq 0.025$. Similarly, the results of a second chi-square analysis where the number of auxin transcription factor binding sites per gene was taken into account indicated that there were a higher number of auxin transcription factor binding sites in the significantly differentially expressed genes than could be expected by chance at a $p \geq 0.01$.

Probe non-target sequences

57,566 of the 164,507 probes included in this study have 85% or greater identity with one or more non-target genes. It was also found that, of the 55,971 poplar genes represented on the microarray, 10,591 genes have one probe with 85% or greater identity to least one non-target gene, 7,847 genes have two probes with 85% or greater identity to

at least one non-target gene and 10,427 genes have three probes with 85% or greater identity to at least one non-target gene. Additionally, it should be noted that 28,272 probes had one non-target hit, 11,250 probes had two non-target hits, 5,586 probes had 3 non-target hits, 23,312 had four non-target hits and 21 had five non-target hits (Fig. 4.12). Of the 568 differentially expressed genes in this particular experiment, 211 had three unique probes and 306 had at least one probe with 85% or greater identity to a non-target sequence. Of the 18 genes used to do real-time RT-PCR validation, seven of them had at least one probe with 85% or greater identity to a non-target sequence including PS2, DIOX, EAUX, ELON, PUT, Ferr and E2.

In order to see if there was a relationship between non-target identity and signal intensity a scatter plot was done where individual probe difference from the gene mean intensity was plotted against identity class (Fig. 4.13). Identity class refers to the degree to which a probe is identical to a non-target sequence and are as follows; 1 = 0-44%, 2 = 45-50%, 3 = 51-60%, 4 = 61-70%, 5 = 71-80%, 6 = 81-90%, 7 = 91-100%. According to the results of the scatter plot, there does not appear to be a relationship between the

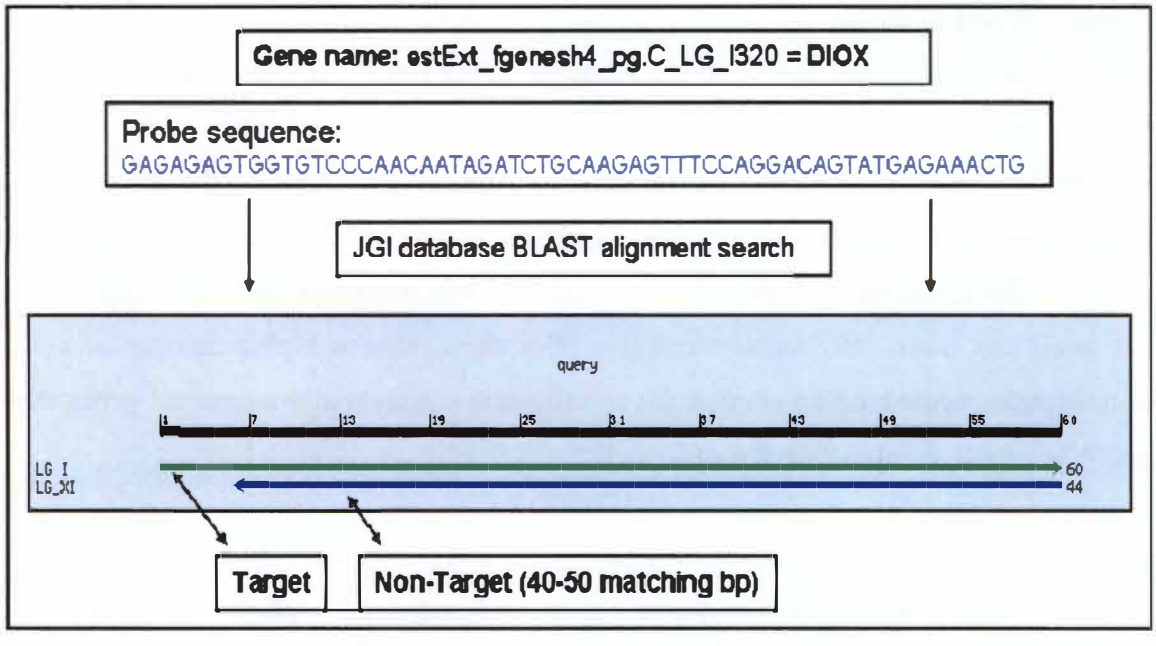

Figure 4.12: Example of probe with one non-target hit

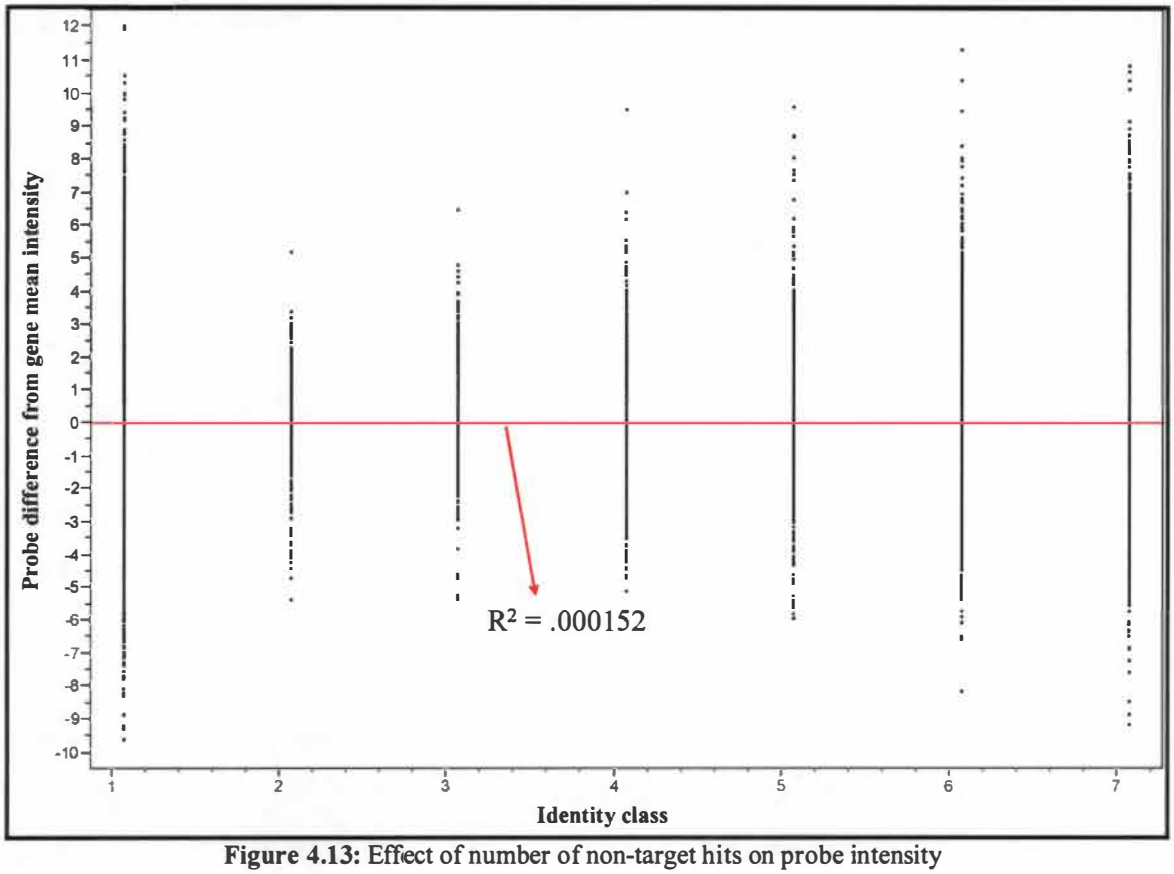

Figure 4.13: Effect of number of non-target hits on probe intensity

number of non-target hits a probe has and it's intensity on the microarray ($R^2 = 0.000152$) which is contradictory to what one might expect.

Discussion

Microarray technology has been widely used as a high throughput approach to study gene expression in many diverse organisms for a number of years. Whole genome microarray technology, in particular, is growing in popularity as whole genome sequencing of an increasing number of organisms is completed. With the whole genome sequence of poplar now available, oligonucleotide microarray technology is an ideally suited tool to elucidate information about downstream regulators of auxin response in poplar roots and, therefore, was used in this study. It was hoped that information from this study could be used to locate candidate downstream regulators of auxin response so that they could eventually be manipulated to increase carbon sequestration capacity in

poplar tree roots, thusly increasing how efficiently poplars can be used as a mitigation strategy to decrease elevated levels of atmospheric CO_2.

In this study, hybrid poplar (H11-11) root tissue was used to do a whole-genome expression analysis of genes that are up and down-regulated 1 hour and 24 hours prior to foliar exogenous auxin treatments (100μM NAA). RNA was extracted from six plants including two control biological reps, two 1 hour harvest biological reps and two 24 hour harvest biological reps and used to synthesize six target samples which were hybridized to NimbleGen™ whole genome microarray slides. The slides were scanned, raw intensity data was extracted and normalized, outliers were removed and ANOVA was applied using both SAS/STAT and GeneSpring. In order to attempt to reduce thenumber of false positives, the final list of differentially expressed genes included only those that were found to be significant ($p \leq 0.01$) using both SAS/STAT and GeneSpring.

Subsequent PCA showed that three main trends were contributing to a combined total of 100% of the variation within the dataset. This information was used to group significant genes into three clusters using a k-means clustering algorithm. Hierarchical clustering was also used to further visualize genes that appeared to have similar expression patterns. However, before an in-depth functional analysis was carried out, a sub-sampling of significantly expressed genes, 18 in total, revealed that the results of the microarray analysis did not correlate well with real-time RT-PCR analysis. Of the 18 genes on which real-time RT-PCR was performed, only 5 showed the same trend in expression levels as the microarray analysis. Additionally, of the 5, only 1 (DIOX) appeared to validate the microarray results relatively strongly.

Though it is not apparent exactly why real-time RT-PCR results did not validate results of the microarray analysis, there are two obvious factors that probably contributed to the discrepancy. These include the fact that many of the probes (57,566 out of 167,507) on the microarray are 85% or more identical to one or more non-target sequences and the fact that the experimental design did not have enough statistical power to overcome the amount of variation inherent in microarray analysis studies.

Though it is alarming that the number of probes with at least one highly identical non-target sequence is so high, it is certainly not surprising based upon what is known

about the genome of *P. trichocarpa.* Analysis of the assembled genome has revealed that a relatively recent whole-genome duplication event thought to have occurred 100 to 120 million years ago impacts around 92% of the *Populus* genome. As a result, approximately 16,000 paralogous gene pairs of the 45,555 predicted gene models have persisted in the genome (Tuskan *et al.* in review). With such a high number of closely related genes present in the genome, it is to be expected that a high number of probes would be closely similar to non-target gene sequence. This is especially true when attempting to design three 60 bp probes for each gene represented on the microarray. However, despite the probe non-target sequence problem, there was not a correlation between number of non-targets per probe and intensity on the microarray. One would expect that the intensity of a probe with one or more non-target sequences would be consistently higher than that of an apparently unique probe because of the increase in the number of potential targets able to hybridize to non-unique probe. Additionally, though 7 of the 18 differentially expressed genes had at least one or more probes with a highly identical non-target sequence, this did not appear to relate to whether or not the real-time RT-PCR results matched the microarray results. For example, the gene whose real-time RT-PCR results appeared to most strongly correlate with the microarray results, DIOX, actually has one probe with greater than 85% identity to a non-target sequence. This leads to the conclusion that another factor, low statistical power of the experimental design, also had an effect on the accuracy of the microarray results.

Low statistical power occurs when there is not enough data, due to low replicate number in this case, to accurately detect a significant effect. This is especially problematic when variation due to non-biological effects is high as is usually the case in microarray analysis. In this experiment, two biological replicates were used to try to minimize the cost of the experiment. It was hoped that applying stringent statistical criteria such as a low p-value cut-off ($p \leq 0.01$) and only retaining differentially expressed genes according to ***both*** SAS/STAT and GeneSpring ANOVA analysis would significantly reduce the number of false positives. However, as shown by results of real-time RT-PCR, this was not the case. Though it did appear as though the results of the microarray analysis were not accurate, chi-square analysis indicated that there were a

higher number of differentially expressed genes with one or more auxin transcription factor binding sites than would be expected by chance alone. This suggests that there were a number of genes that were actually differentially up or down-regulated in response to auxin in the final list of differentially expressed genes. However, the low statistical power of the experiment makes it hard to confidently determine which genes these actually are.

Conclusions

The study of downstream regulators of auxin response in poplar roots is a crucial step in understanding how *Populus* root growth can be manipulated in the future to increase carbon sequestration ability. With the genomic tools that are available, *Populus* is ideally suited for whole-genome microarray analysis studies to help answer these questions. However, because the *Populus* genome contains many highly similar paralogous genes and because of the non-biological variability generally associated with microarray technology, great care needs to be taken in the future to make sure valid and reliable results are produced.

First and foremost, a decision needs to be made about how to deal with probes that are highly identical to non-target sequences. The most logical answer initially seems to be to remove these probes from the microarray analysis. However, doing so would result in the removal of more than 10,000 genes, some of which are known to be auxin regulated, from the current microarray design. One potential option is to determine if there is a flaw in the technique used by NimbleGen™ to design the current probes. It is possible that some of the problematic probe sequences could be replaced by sequences that are, if not unique, at least less similar to non-target sequences.

The ability to identify truly differentially expressed genes can also be greatly improved if more robust statistical analysis techniques are used. For example, sufficient statistical power is of utmost importance and the extra cost associated with additional biological and technical reps is worth it if reliable results are obtained. It should be noted that, in an attempt to increase statistical power, the most current *Populus* microarray design includes a technical replicate of the array on each slide. Statistical analysis can

also be improved by including more stringent cut-off criteria to reduce the number of false positives. This can be achieved by applying a false discovery rate (FDR) cut-off (q-value) to the data in addition to a low p-value cut-off. The difference between a false positive rate (p-value) and an FDR (q-value) is that the former represents the proportion of true null hypotheses that were incorrectly rejected, while the latter represents the expected proportion of false positives among comparisons that passed the p-value cut-off (Storey and Tibshirani 2003). Lastly, it would also be advantageous to analyze individual probes separately rather than using the average intensity value of the three different probes representing the same gene. This certainly complicates the analysis process, but in the end would lead to more accurate and informative results.

In summary, though many problems were encountered during the analysis stage of this study, the technology still holds much promise as an effective tool for future *Populus* expression analysis studies. As expected when using a first generation microarray design, there will always be initial unforeseen problems that need to be worked out before the technology can be used effectively.

Bibliography

Abe, K., M. Ikeda, H. Dansako, K. Naka, K. Shimotohno, and N. Kato, 2005, cDNA microarray analysis to compare HCCV subgenomic replicon cells with their cured cells: Virus Research, v. 107, p. 73-81.

Abel, S., P. W. Oeller, and A. Theologis, 1994, Early auxin-induced genes encode short-Lived Nuclear Proteins: Proceedings of the National Academy of Sciences of the United States of America, v. 91, no. 1, p. 326-330.

Abel, S., M. D. Nguyen, and A. Theologis, 1995, The Ps-Iaa4/5-like family of early auxin-inducible messenger-RNAs in *Arabidopsis thaliana*: Journal of Molecular Biology, v. 251, no. 4, p. 533-549.

Aharoni, A., L. C. P. Keizer, H. C. Van den Broeck, R. Blanco-Portales, J. Munoz-Blanco, G. Bois, P. Smit, R. C. H. De Vos, and A. P. O'Connell, 2002, Novel insight into vascular, stress, and auxin-dependent and independent gene expression programs in strawberry, a non-climacteric fruit: Plant Physiology, v. 129, no. 3, p. 1019-1031.

Albert, T. J., J. Norton, M. Ott, T. Richmond, K. Nuwaysir, E. F. Nuwaysir, K. P. Stengele, and R. D. Green, 2003, Light-directed 5'->3' synthesis of complex oligonucleotide microarrays: Nucleic Acids Research, v. 31, no. 7.

Aloni, R., P.J. Davies ed., 1995, The induction of vascular tissue by auxin and cytokinin, in Plant Hormones and Their Role in Plant Growth Development: Dordrecht, Kluwer, p. 531-546.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990, Basic local alignment search tool: Journal of Molecular Biology, v. 215, no. 3, p. 403-410.

Anderson, J. V., and D. P. Horvath, 2001, Random sequencing of cDNAs and identification of mRNAs: Weed Science, v. 49, no. 5, p. 590-597.

Applied Biosystems. BigDye Terminator v3.1 Cycle Sequencing Kit Protocol. 2002.

Arteca, R., 1995, Plant Growth Substances: New York, Chapman and Hall.

Aspeborg, H. *et al.*, 2005, Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen: Plant Physiology, v. 137, no. 3, p. 983-997.

Audic, S., and J. M. Claverie, 1997, The significance of digital gene expression profiles: Genome Research, v. 7, no. 10, p. 986-995.

Balatinecz, J., and D. Kretschmann, D.I. Dickamnn, J.G. Isebrands, J.E. Eckenwalder, and J. Richardson eds., 2001, Properties and Utilization of Poplar Wood, in, Poplar Culture in North America: Ottawa, NRC Research Press, p. 277-291.

Ballas, N., L. M. Wong, and A. Theologis, 1993, Identification of the auxin-responsive element, AUXRE, in the primary indoleacetic acid-inducible gene, PS-IAA4/5, of pea (*Pisum sativum*): Journal of Molecular Biology, v. 233, no. 4, p. 580-596.

Barac, T., S. Taghavi, B. Borremans, A. Provoost, L. Oeyen, J. V. Colpaert, J. Vangronsveld, and D. van der Lelie, 2004, Engineered endophytic bacteria improve phytoremediation of water-soluble, volatile, organic pollutants: Nature Biotechnology, v. 22, no. 5, p. 583-588.

Barakat, A., K. Szick-Miranda, I. F. Chang, R. Guyot, G. Blanc, R. Cooke, M. Delseny, and J. Bailey-Serres, 2001, The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome: Plant Physiology, v. 127, no. 2, p. 398-415.

Barczak, A., M. W. Rodriguez, K. Hanspers, L. L. Koth, Y. C. Tai, B. M. Bolstad, T. P. Speed, and D. J. Erle, 2003, Spotted long oligonucleotide arrays for human gene expression analysis: Genome Research, v. 13, no. 7, p. 1775-1785.

Becker, J., and E. A. Craig, 1994, Heat-shock proteins as molecular chaperones: European Journal of Biochemistry, v. 219, no. 1-2, p. 11-23.

Bennett, M. J., A. Marchant, H. G. Green, S. T. May, S. P. Ward, P. A. Millner, A. R. Walker, B. Schulz, and K. A. Feldmann, 1996, Arabidopsis AUX1 gene: A permease-like regulator of root gravitropism: Science, v. 273, no. 5277, p. 948-950.

Bennett, M. J., A. Marchant, S. T. May, and R. Swarup, 1998, Going the distance with auxin: unraveling the molecular basis of auxin transport: Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, v. 353, no. 1374, p. 1511-1515.

Bent, A. F., 2000, Arabidopsis in plant transformation. Uses, mechanisms, and prospects for transformation of other species: Plant Physiology, v. 124, no. 4, p. 1540-1547.

Bhatia, S., and R. Singh, 2002, Phytohormone-mediated transformation of sugars to starch in relation to the activities of amylases, sucrose-metabolising enzymes in sorghum grain: Plant Growth Regulation, v. 36, no. 2, p. 97-104.

Bio-Rad Laboratories. iQ SYBR Green Supermix. 2005.

Blakely, L. M., M. Durham, T. A. Evans, and R. M. Blakely, 1982, Experimental studies on lateral root-formation in radish seedling roots. General-methods, developmental stages, and spontaneous formation of laterals: Botanical Gazette, v. 143, no. 3, p. 341-352.

Blanco, L., A. Bernad, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas, 1989, Highly efficient DNA-synthesis by the phage Phi29 DNA-polymerase - Symmetrical mode of DNA-replication: Journal of Biological Chemistry, v. 264, no. 15, p. 8935-8940.

Boerjan, W. et al., 1995, Superroot, A recessive mutation in Arabidopsis, confers auxin overproduction: Plant Cell, v. 7, no. 9, p. 1405-1419.

Bonaldo, M. D. F., G. Lennon, and M. B. Soares, 1996, Normalization and subtraction: Two approaches to facilitate gene discovery: Genome Research, v. 6, no. 9, p. 791-806.

Braatne, J. H., S. B. Rood, P. E. Heilman, R.F. Stettler, J. Bradshaw, and T.M. Hinckley eds., 1996, Life history, ecology and conservation of riparian cottonwoods in North America, in, Biology of *Populus* and its implications for management and conservation: Ottawa, NRC Research Press, p. 57-80.

Bradshaw, H. D., and R. F. Stettler, 1993, Molecular-genetics of growth and development in *Populus* .1. Triploidy in hybrid Poplars: Theoretical and Applied Genetics, v. 86, no. 2-3, p. 301-307.

Bradshaw, H. D., R. Ceulemans, J. Davis, and R. Stettler, 2000, Emerging model systems in plant biology: Poplar (*Populus*) as a model forest tree: Journal of Plant Growth Regulation, v. 19, no. 3, p. 306-313.

Bremer, K. *et al.*, 1998, An ordinal classification for the families of flowering plants: Annals of the Missouri Botanical Garden, v. 85, no. 4, p. 531-553.

Brinker, M., L. van Zyl, W. B. Liu, D. Craig, R. R. Sederoff, D. H. Clapham, and S. von Arnold, 2004, Microarray analyses of gene expression during adventitious root development in *Pinus contorta*: Plant Physiology, v. 135, no. 3, p. 1526-1539.

Brown, D. E., A. M. Rashotte, A. S. Murphy, J. Normanly, B. W. Tague, W. A. Peer, L. Taiz, and G. K. Muday, 2001, Flavonoids act as negative regulators of auxin transport in vivo in Arabidopsis: Plant Physiology, v. 126, no. 2, p. 524-535.

Brown, P. O., and D. Botstein, 1999, Exploring the new world of the genome with DNA microarrays: Nature Genetics, v. 21, p. 33-37.

Brundrett, M., G. Murase, and B. Kendrick, 1990, Comparative anatomy of roots and mycorrhizae of common Ontario trees: Canadian Journal of Botany-Revue, v. 68, no. 3, p. 551-578.

Brunner, A. M., V. B. Busov, and S. H. Strauss, 2004, Poplar genome sequence: functional genomics in an ecologically dominant plant species: Trends in Plant Science, v. 9, no. 1, p. 49-56.

Brunner, A. M., I. A. Yakovlev, and S. H. Strauss, 2004, Validating internal controls for quantitative plant gene expression studies: BMC Plant Biology, v. 4, p. 14-20.

Buermans, H. P. J. *et al.*, 2005, Microarray analysis reveals pivotal divergent mRNA expression profiles early in the development of either compensated ventricular hypertrophy or heart failure: Physiological Genomics, v. 21, no. 3, p. 314-323.

Burns, R. M., B. H. Honkala, 1990, Silvics of North America: Washington, D.C., U.S. Dept. of Agriculture, Forest Service.

Bustin, S. A., 2000, Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays: Journal of Molecular Endocrinology, v. 25, p. 169-193.

Bustin, S. A., 2002, Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems: Journal of Molecular Endocrinology, v. 29, no. 1.

Call, D. R., D. P. Chandler, and F. Brockman, 2001, Fabrication of DNA microarrays using unmodified oligonucleotide probes: Biotechniques, v. 30, no. 2, p. 368.

Campanoni, P., and P. Nick, 2005, Auxin-dependent cell division and cell elongation. 1-naphthaleneacetic acid and 2,4-dichlorophenoxyacetic acid activate different pathways: Plant Physiology, v. 137, no. 3, p. 939-948.

Carson, D. L., and F. C. Botha, 2000, Preliminary analysis of expressed sequence tags for sugarcane: Crop Science, v. 40, no. 6, p. 1769-1779.

Carulli, J. P. *et al.*, 1998, High throughput analysis of differential gene expression: Journal of Cellular Biochemistry, v. 286, p. 30-31.

Casimiro, I. *et al.*, 2001, Auxin transport promotes Arabidopsis lateral root initiation: Plant Cell, v. 13, no. 4, p. 843-852.

Casson, S. A., and K. Lindsey, 2003, Genes and signalling in root development: New Phytologist, v. 158, no. 1, p. 11-38.

Celenza, J. L., P. L. Grisafi, and G. R. Fink, 1995, A Pathway for lateral root-formation in *Arabidopsis thaliana*: Genes & Development, v. 9, no. 17, p. 2131-2142.

Cervera, M. T., V. Storme, B. Ivens, J. Gusmao, B. H. Liu, V. Hostyn, J. Van Slycken, M. Van Montagu, and W. Boerjan, 2001, Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers: Genetics, v. 158, no. 2, p. 787-809.

Cheadle, C., M. P. Vawter, W. J. Freed, and K. G. Becker, 2003, Analysis of microarray data using Z-score transformation: Journal of Molecular Diagnostics, v. 5, no. 2, p. 73-81.

Chen, J. G., H. Ullah, J. C. Young, M. R. Sussman, and A. M. Jones, 2001, ABP1 is required for organized cell elongation and division in Arabidopsis embryogenesis: Genes & Development, v. 15, no. 7, p. 902-911.

Chen, J. G., S. Shimomura, F. Sitbon, G. Sandberg, and A. M. Jones, 2001, The role of auxin-binding protein 1 in the expansion of tobacco leaf cells: Plant Journal, v. 28, no. 6, p. 607-617.

Chou, H. H., and M. H. Holmes, 2001, DNA sequence quality trimming and vector removal: Bioinformatics, v. 17, no. 12, p. 1093-1104.

Chou, I. T., and C. S. Gasser, 1997, Characterization of the cyclophilin gene family of *Arabidopsis thaliana* and phylogenetic analysis of known cyclophilin proteins: Plant Molecular Biology, v. 35, no. 6, p. 873-892.

Christoffels, A., A. van Gelder, G. Greyling, R. Miller, T. Hide, and W. Hide, 2001, STACK: Sequence Tag Alignment and Consensus Knowledgebase: Nucleic Acids Research, v. 29, no. 1, p. 234-238.

Chuaqui, R. F. *et al.*, 2002, Post-analysis follow-up and validation of microarray experiments: Nature Genetics, v. 32, p. 509-514.

Ciesielski, T., 1872, Untersuchungen uber die Abwartskrummung der Wurzel: Beitrage zur Biologie der Planzen, v. 1, p. 1-30.

Clonetech. Clonetech PCR-Select^{TM} cDNA Subtraction Kit User Manual. 2002.

Coba de la Peña, T., *et al.*, 1997, A Carbonic anhydrase gene is induced in the nodule primordium and its cell-specific expression is controlled by the presense of *Rhizobium* during development.: The Plant Journal, v. 11, no. 3, p. 407-420.

Cohen, J. D., and R. S. Bandurski, 1982, Chemistry and physiology of the bound auxins: Annual Review of Plant Physiology and Plant Molecular Biology, v. 33, p. 403-430.

Cram, W. H., and C. H. Lindquist, 1982, Refrigerated storage for hardwood cuttings of Willow and poplar: Tree Planters' Notes, v. 33, no. 4, p. 3-5.

Darwin, C., 1880, The Power of Movement in Plants: John Murray.

Dean, F. B., J. R. Nelson, T. L. Giesler, and R. S. Lasken, 2001, Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification: Genome Research, v. 11, no. 6, p. 1095-1099.

Deckmyn, G., B. Muys, J. G. Quijano, and R. Ceulemans, 2004, Carbon sequestration following afforestation of agricultural soils: comparing oak/beech forest to short-rotation poplar coppice combining a process and a carbon accounting model: Global Change Biology, v. 10, no. 9, p. 1482-1491.

Denef, V. J., J. Park, J. L. M. Rodrigues, T. V. Tsoi, S. A. Hashsham, and J. M. Tiedje, 2003, Validation of a more sensitive method for using spotted oligonucleotide DNA microarrays for functional genomics studies on bacterial communities: Environmental Microbiology, v. 5, no. 10, p. 933-943.

Dharmasiri, N., S. Dharmasiri, and M. Estelle, 2005, The F-box protein TIR1 is an auxin receptor: Nature, v. 435, no. 7041, p. 441-445.

Dickmann, D. I., and K. W. Stuart, 1983, The culture of poplars in eastern North America: Ann Arbor, MI, McNaughton and Gunn, Inc.

Dixon, R. K., S. Brown, R. A. Houghton, A. M. Solomon, M. C. Trexler, and J. Wisniewski, 1994, Carbon pools and flux of global forest ecosystems: Science, v. 263, no. 5144, p. 185-190.

Dong, J. X., C. H. Chen, and Z. X. Chen, 2003, Expression profiles of the *Arabidopsis* WRKY gene super-family during plant defense response: Plant Molecular Biology, v. 51, no. 1, p. 21-37.

Duggan, D. J., M. Bittner, Y. D. Chen, P. Meltzer, and J. M. Trent, 1999, Expression profiling using cDNA microarrays: Nature Genetics, v. 21, p. 10-14.

Eckenwalder, J. E., RF Stettler, J Bradshaw, PE Heilman, and TM Hinckley eds. 1996, Systematics and evolution of *Populus,* in, Biology of *Populus* and its Implications for Management and Monservation: Ottawa, NRC Research Press, p. 7-32.

Ericsson, T., L. Rytter, and E. Vapaavuori, 1996, Physiology of carbon allocation in trees: Biomass & Bioenergy, v. 11, no. 2-3, p. 115-127.

Eriksson, E. M., A. Bovy, K. Manning, L. Harrison, J. Andrews, J. De Silva, G. A. Tucker, and G. B. Seymour, 2004, Effect of the colorless non-ripening mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening: Plant Physiology, v. 136, p. 4184-4197.

Ermolaeva, O. D., and E. D. Sverdlov, 1996, Subtractive hybridization, a technique for extraction of DNA sequences distinguishing two closely related genomes: Critical analysis: Genetic Analysis-Biomolecular Engineering, v. 13, no. 2, p. 49-58.

Eulgem, T., P. J. Rushton, S. Robatzek, and I. E. Somssich, 2000, The WRKY super-family of plant transcription factors: Trends in Plant Science, v. 5, no. 5, p. 199-206.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green, 1998, Base-calling of automated sequencer traces using Phred. I. Accuracy assessment: Genome Research, v. 8, no. 3, p. 175-185.

Ewing, B., and P. Green, 1998, Base-calling of automated sequencer traces using Phred. II. Error probabilities: Genome Research, v. 8, no. 3, p. 186-194.

Farrar, J. F., and D. L. Jones, 2000, The control of carbon acquisition by roots: New Phytologist, v. 147, no. 1, p. 43-53.

Fidelity Systems. Phi29 Random Hexamer Primer. 2004.

Forster, T., D. Roy, and P. Ghazal, 2003, Experiments using microarray technology: limitations and standard operating procedures: Journal of Endocrinology, v. 178, no. 2, p. 195-204.

Freeman, W. M., S. J. Walker, and K. E. Vrana, 1999, Quantitative RT-PCR: Pitfalls and potential: Biotechniques, v. 26, no. 1, p. 112.

Friml, J., and K. Palme, 2002, Polar auxin transport - old questions and new concepts: Plant Molecular Biology, v. 49, no. 3-4, p. 273-284.

Friml, J., 2003, Auxin transport - shaping the plant: Current Opinion in Plant Biology, v. 6, no. 1, p. 7-12.

Gachon, C., A. Mingam, and B. Charrier, 2004, Real-time PCR: what relevance to plant studies?: Journal of Experimental Botany, v. 55, no. 402, p. 1445-1454.

Gao, M. J., I. A. P. Parkin, D. J. Lydiate, and A. Hannoufa, 2004, An auxin-responsive SCARECROW-like transcriptional activator interacts with histone deacetylase: Plant Molecular Biology, v. 55, no. 3, p. 417-431.

Gehrig, H., K. Winter, J. Cushman, A. Borland, and T. Taybi, 2000, An Improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides: Plant Molecular Biology Reporter, v. 18, p. 369-376.

Geldner, N., T. Hamann, and G. Jurgens, 2000, Is there a role for auxin in early embryogenesis?: Plant Growth Regulation, v. 32, no. 2-3, p. 187-191.

Gershon, D., 2004, Microarrays go mainstream: Nature Methods, v. 1, no. 3, p. 263-269.

Gibson, G., 2002, Microarrays in ecology and evolution: a preview: Molecular Ecology, v. 11, no. 1, p. 17-24.

Gil, P., Y. Liu, V. Orbovic, E. Verkamp, K. L. Poff, and P. J. Green, 1994, Characterization of the auxin-inducible SAUR-Ac1 gene for use as a molecular-genetic tool in Arabidopsis: Plant Physiology, v. 104, no. 2, p. 777-784.

Gocal, G. F. W., R. P. Pharis, E. C. Yeung, and D. Pearce, 1991, Changes after decapitation in concentrations of ndole-3-acetic-acid and abscisic-acid in the larger axillary bud of *Phaseolus vulgaris* L.cv Tender Green: Plant Physiology, v. 95, no. 2, p. 344-350.

Gray, W. M., and M. Estelle, 2000, Function of the ubiquitin-proteasome pathway in auxin response: Trends in Biochemical Sciences, v. 25, no. 3, p. 133-138.

Guilfoyle, T., G. Hagen, T. Ulmasov, and J. Murfett, 1998, How does auxin turn on genes?: Plant Physiology, v. 118, no. 2, p. 341-347.

Guo, Y., Z. Cai, and S. Gan, 2004, Transcriptome of Arabidopsis leaf senescence: Plant Cell and Environment, v. 27, no. 5, p. 521-549.

Gurley W., 2004, HSP101: A key component for the acquisition of thermotolerance in plants: The Plant Cell, v. 12, no. 457, p. 457-460.

Hagen, G., G. Martin, Y. Li, and T. J. Guilfoyle, 1991, Auxin-induced expression of the soybean GH3 promoter in transgenic tobacco plants: Plant Molecular Biology, v. 17, no. 3, p. 567-579.

Hagen, G., and T. Guilfoyle, 2002, Auxin-responsive gene expression: genes, promoters and regulatory factors: Plant Molecular Biology, v. 49, no. 3-4, p. 373-385.

Han, K. H., and *et al.*, R.F. Stettler, H.D. Bradshaw, P.E. Heilman, and T.M. Hinckley eds., 1996, Cellular and molecular biology of *Agrobacterium*-mediated transformation of plants and its application to genetic transformation of *Populus*, in, Biology of *Populus* and its Implications for Management and Conservation: Ottawa, NRC Research Press, p. 201-222.

Hansen, E. A., 1993, Soil carbon sequestration beneath hybrid poplar plantations in the north central United States: Biomass & Bioenergy, v. 5, no. 6, p. 431-436.

Heath, J., E. Ayres, M. Possell, R. D. Bardgett, H. I. J. Black, H. Grant, P. Ineson, and G. Kerstiens, 2005, Rising atmospheric CO_2 reduces sequestration of root-derived soil carbon: Science, v. 309, no. 5741, p. 1711-1713.

Heilman, P. E., and R. F. Stettler, 1985, Genetic-variation and productivity of *Populus trichocarpa* and its hybrids .2. Biomass production in a 4-year plantation: Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiere, v. 15, no. 2, p. 384-388.

Hellmann, H., and M. Estelle, 2002, Plant development: Regulation by protein degradation: Science, v. 297, no. 5582, p. 793-797.

Higo, K., Y. Ugawa, M. Iwamoto, and T. Korenaga, 1999, Plant cis-acting regulatory DNA elements (PLACE) database: 1999: Nucleic Acids Research, v. 27, no. 1, p. 297-300.

Hobbie, L., and M. Estelle, 1995, The Axr4 auxin-resistant mutants of *Arabidopsis thaliana* define a gene important for root gravitropism and lateral root initiation: Plant Journal, v. 7, no. 2, p. 211-220.

Hoek, K. *et al.*, 2004, Expression profiling reveals novel pathways in the transformation of melanocytes to melanomas.: Cancer Research, v. 64, p. 5270-5282.

Hoffman, E. P. *et al.*, 2004, Guidelines - Expression profiling - best practices for data generation and interpretation in clinical trials: Nature Reviews Genetics, v. 5, no. 3, p. 229-237.

Holzwarth, J. A., R. P. Middleton, M. Roberts, R. Mansourian, F. Raymond, and S. S. Hannah, 2005, The development of a high-density canine microarray: Journal of Heredity, v. 96, no. 7, p. 817-820.

Hooykaas, P., M. Hall, K. Libbenga, P. Hooykaas, M. Hall, and K. Libbenga, 1999, Biochemistry and Molecular Biology of Plant Hormones: The Netherlands, Elsevier Science B.V.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. Vander Linden, X. Dai, K. Maskell, and C. A. Johnson, 2001, Climate change: the Scientific Basis.: Cambridge, UK, Cambridge University Press.

Hu, P., E. L. Brodie, Y. Suzuki, H. H. McAdams, and G. L. Andersen, 2005, Whole-genome transcriptional analysis of heavy metal stresses in *Caulobacter crescentus*: Journal of Bacteriology, v. 187, no. 24, p. 8437-8449.

Hughes, T. R. et al., 2001, Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer: Nature Biotechnology, v. 19, no. 4, p. 342-347.

Imbeaud, S., and C. Auffray, 2005, 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments: Drug Discovery Today, v. 10, no. 17, p. 1175-1182.

Invitrogen Life Technologies. SuperScript III Reverase Transcriptase. 2003.

Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed, 2003, Summaries of affymetrix GeneChip probe level data: Nucleic Acids Research, v. 31, no. 4.

Isebrands, J. G., and *et al.*, 1979, Integrated Utilization of Biomass, a Case study of short-rotation intensively cultured *Populus* raw materials: TAPPI, v. 62, no. 7, p. 67-70.

Isebrands, J. G., and D. F. Karnosky, D.I. Dickamnn, J.G. Isebrands, J.E. Eckenwalder, and J. Richardson eds., 2001, Environmental benefits of poplar culture, in Poplar Culture in North America: Ottawa, NRC Research Press, p. 207-218.

Jacobs, M., and P. H. Rubery, 1988, Naturally-occurring auxin transport regulators: Science, v. 241, no. 4863, p. 346-349.

Johnson, M. A., M. A. Perez-Amador, P. Lidder, and P. J. Green, 2000, Mutants of Arabidopsis defective in a sequence-specific mRNA degradation pathway: Proceedings of the National Academy of Sciences of the United States of America, v. 97, no. 25, p. 13991-13996.

Jones Jr, S. B., and A. E. Luchsinger, M.J. Martin ed., 1986, Angiosperm families, in Plant Systematics: New York, McGraw-Hill Publishing Company, p. 293-450.

Jones, A. M., 1994, Auxin-binding proteins: Annual Review of Plant Physiology and Plant Molecular Biology, v. 45, p. 393-420.

Jones, A. M., K. H. Im, M. A. Savka, M. J. Wu, N. G. Dewitt, R. Shillito, and A. N. Binns, 1998, Auxin-dependent cell expansion mediated by over-expressed auxin-binding protein 1: Science, v. 282, no. 5391, p. 1114-1117.

Jouanin, L., A. C. M. Brasileiro, J. C. Leple, G. Pilate, and D. Cornu, 1993, Genetic-transformation - A short review of methods and their applications, results and perspectives for forest trees: Annales des Sciences Forestieres, v. 50, no. 4, p. 325-336.

Kane, M. D., T. A. Jatkoe, C. R. Stumpf, J. Lu, J. D. Thomas, and S. J. Madore, 2000, Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays: Nucleic Acids Research, v. 28, no. 22, p. 4552-4557.

Kanehisa, M., and S. Goto, 2000, KEGG: Kyoto Encyclopedia of Genes and Genomes: Nucleic Acids Research, v. 28, no. 1, p. 27-30.

Karge W.H., and *et al.*, 1998, Quantification of mRNA by polymerase chain reaction (PCR) using an internal standard and a nonradioactive detection method: Methods in Molecular Biology, v. 110, p. 43-61.

Keeling,CD, T P Whorf., 2004, Trends: A compendium of data on global change: Carbon Dioxide Information Analysis Center.

Kepinski, S., and O. Leyser, 2005, The Arabidopsis F-box protein TIR1 is an auxin receptor: Nature, v. 435, no. 7041, p. 446-451.

Kirmizis, A., S. M. Bartley, A. Kuzmichev, R. Margueron, D. Reinberg, R. Green, and P. J. Farnham, 2004, Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27: Genes & Development, v. 18, no. 13, p. 1592-1605.

Kirst, M., A. A. Myburg, J. P. G. De Leon, M. E. Kirst, J. Scott, and R. Sederoff, 2004, Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus: Plant Physiology, v. 135, no. 4, p. 2368-2378.

Klambt, D., 1990, A view about the function of auxin-binding proteins at plasma-membranes: Plant Molecular Biology, v. 14, no. 6, p. 1045-1050.

Knee, E. M., and R. P. Hangarter, 1996, Differential IAA dose response relations of the axr1 and axr2 mutants of Arabidopsis: Physiologia Plantarum, v. 98, no. 2, p. 320-324.

Kogl, F., and A. J. Haagen Smit, 1931, Uber die Chemie des Wuchsstoffs: Amsterdam Proceedings Section Science, v. 34, p. 1411-1416.

Kohler, A., C. Delaruelle, D. Martin, N. Encelot, and F. Martin, 2003, The poplar root transcriptome: analysis of 7000 expressed sequence tags: FEBS Letters, v. 542, no. 1-3, p. 37-41.

Lalonde, S., D. Wipf, and W. B. Frommer, 2004, Transport mechanisms for organic forms of carbon and nitrogen between source and sink: Annual Review of Plant Biology, v. 55, p. 341-372.

Li, J., M. Pankratz, and J. A. Johnson, 2002, Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays: Toxicological Sciences, v. 69, no. 2, p. 383-390.

Li, J., G. Brader, and E. T. Palva, 2004, The WRKY70 transcription factor: A node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense: Plant Cell, v. 16, no. 2, p. 319-331.

Liscum, E., and E. L. Stowe-Evans, 2000, Phototropism: A "simple" physiological response modulated by multiple interacting photosensory-response pathways: Photochemistry and Photobiology, v. 72, no. 3, p. 273-282.

Liscum, E., and J. W. Reed, 2002, Genetics of Aux/IAA and ARF action in plant growth and development: Plant Molecular Biology, v. 49, no. 3-4, p. 387-400.

Liu, T., J. van Staden, and W. A. Cress, 2000, Gene expression in soybean (*Glycine max* L.) roots exposed to low temperature: Isolation of cDNA clone: Plant Growth Regulation, v. 30, no. 3, p. 247-251.

Liu, Z. B., G. Hagen, and T. J. Guilfoyle, 1997, A G-box-binding protein from soybean binds to the E1 auxin-response element in the soybean GH3 promoter and contains a proline-rich repression domain: Plant Physiology, v. 115, no. 2, p. 397-407.

Lockhart, D. J. et al., 1996, Expression monitoring by hybridization to high-density oligonucleotide arrays: Nature Biotechnology, v. 14, no. 13, p. 1675-1680.

Lockhart, D. J., and E. A. Winzeler, 2000, Genomics, gene expression and DNA arrays: Nature, v. 405, no. 6788, p. 827-836.

Lu, S. E., N. Wang, J. L. Wang, Z. J. Chen, and D. C. Gross, 2005, Oligonucleotide microarray analysis of the SalA regulon controlling phytotoxin production by *Pseudomonas syringae* pv. *syringae*: Molecular Plant-Microbe Interactions, v. 18, no. 4, p. 324-333.

Magee, J. A., T. Araki, S. Patil, T. Ehrig, L. True, P. A. Humphrey, W. J. Catalona, M. A. Watson, and J. Milbrandt, 2001, Expression profiling reveals hepsin over-expression in prostate cancer: Cancer Research, v. 61, no. 15, p. 5692-5696.

Marland, G., and B. Schlamadinger, 1999, The Kyoto Protocol could make a difference for the optimal forest-based CO_2 mitigation strategy: some results from GORCAM.: Environmental Science and Policy, v. 111, p. 111-124.

Martin, F., G. A. Tuskan, S. P. DiFazio, P. Lammers, G. Newcombe, and G. K. Podila, 2004, Symbiotic sequencing for the *Populus* mesocosm: New Phytologist, v. 161, no. 2, p. 330-335.

Matsui, K., K. Hiratsu, T. Koyama, H. Tanaka, and M. Ohme-Takagi, 2005, A chimeric AtMYB23 repressor induces hairy roots, elongation of leaves and stems, and inhibition of the deposition of mucilage on seed coats in Arabidopsis: Plant and Cell Physiology, v. 46, no. 1, p. 147-155.

Matz, M. V., and S. A. Lukyanov, 1998, Different strategies of differential display: areas of application: Nucleic Acids Research, v. 26, no. 24, p. 5537-5543.

Mayer, K., and H. W. Mewes, 2001, How can we deliver the large plant genomes? Strategies and perspectives: Plant Biotechnology, v. **5**, p. 173-177.

Mcclure, B. A., G. Hagen, C. S. Brown, M. A. Gee, and T. J. Guilfoyle, 1989, Transcription, organization, and sequence of an auxin-regulated gene-cluster in soybean: Plant Cell, v. 1, no. 2, p. 229-239.

McCormick, M., 2004, High density microarrays tackle the regulome: Genetic Engineering News, v. 24, no. 16.

Mclaughlin, M. M. et al., 1992, The yeast cyclophilin multigene family - Purification, cloning and characterization of a new isoform: Gene, v. 111, no. 1, p. 85-92.

Meijer, A. H., F. J. Verbeek, E. Salas-Vidal, M. Corredor-Adamez, J. Bussman, A. M. van der Sar, G. W. Otto, R. Geisler, and H. P. Spaink, 2005, Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to *Mycobacterium marinum* infection: Molecular Immunology, v. 42, no. 10, p. 1185-1203.

Melk, A., E. S. Mansfield, S. C. Hsieh, T. Hernandez-Boussard, P. Grimm, D. C. Rayner, P. F. Halloran, and M. M. Sarwal, 2005, Transcriptional analysis of the molecular basis of human kidney aging using cDNA microarray profiling: Kidney International, v. 68, no. 6, p. 2667-2679.

Michler, C. H., and E. O. Bauer, 1991, High-frequency somatic embryogenesis from leaf tissue of *Populus* spp: Plant Science, v. 77, no. 1, p. 111-118.

Morgan, K. E., T. I. Zarembinski, A. Theologis, and S. Abel, 1999, Biochemical characterization of recombinant polypeptides corresponding to the predicted beta alpha alpha fold in Aux/IAA proteins: Febs Letters, v. 454, no. 3, p. 283-287.

Moseyko, N., T. Zhu, H. S. Chang, X. Wang, and L. J. Feldman, 2002, Transcription profiling of the early gravitropic response in Arabidopsis using high-density oligonucleotide probe microarrays: Plant Physiology, v. 130, no. 2, p. 720-728.

Muday, G. K., and P. Haworth, 1994, Tomato root-growth, gravitropism, and lateral development - Correlation with auxin transport: Plant Physiology and Biochemistry, v. 32, no. 2, p. 193-203.

Müssig, C., S. Fischer, and T. Altmann, 2002, Brassinosteroid-regulated gene expression: Plant Physiology, v. 129, no. 3, p. 1241-1251.

Nagpal, P., L. M. Walker, J. C. Young, A. Sonawala, C. Timpte, M. Estelle, and J. W. Reed, 2000, AXR2 encodes a member of the Aux/IAA protein family: Plant Physiology, v. 123, no. 2, p. 563-573.

Nalbandian, A., A. L. Y. Pang, O. M. Rennert, W. Y. Chan, N. Ravindranath, and D. Djakiew, 2005, A novel function of differentiation revealed by cDNA microarray profiling of p75(NTR)-regulated gene expression: Differentiation, v. 73, no. 8, p. 385-396.

Nelson, J. R. et al., 2002, TempliPhi, phi 29 DNA polymerase based rolling circle amplification of templates for DNA sequencing: Biotechniques, p. 44-47.

New England BioLabs. Phi29 DNA Polymerase. 2004.

Nguyen, P. V., D. I. Dickmann, K. S. Pregitzer, and R. Hendrick, 1990, Late-season changes in allocation of starch and sugar to shoots, coarse roots, and fine roots in two hybrid Poplar Clones: Tree Physiology, v. 7, no. 1-4, p. 95-105.

Normanly, J., J. D. Cohen, and G. R. Fink, 1993, *Arabidopsis thaliana* auxotrophs reveal a tryptophan-independent biosynthetic-pathway for indole-3-acetic-acid: Proceedings of the National Academy of Sciences of the United States of America, v. 90, no. 21, p. 10355-10359.

Nuwaysir, E. F. et al., 2002, Gene expression analysis using oligonucleotide arrays produced by maskless photolithography: Genome Research, v. 12, no. 11, p. 1749-1755.

Park, S., S. Oh, and K. H. Han, 2004, Large-scale computational analysis of poplar ESTs reveals the repertoire and unique features of expressed genes in the poplar genome: Molecular Breeding, v. 14, no. 4, p. 429-440.

Parker, K. E., and W. R. Briggs, 1990, Transport of indole-3-acetic-acid during gravitropism in intact maize coleoptiles: Plant Physiology, v. 94, no. 4, p. 1763-1769.

Peer, W. A., A. Bandyopadhyay, J. J. Blakeslee, S. I. Makam, R. J. Chen, P. H. Masson, and A. S. Murphy, 2004, Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with altered auxin transport in *Arabidopsis thaliana*: Plant Cell, v. 16, no. 7, p. 1898-1911.

Pina, C., F. Pinto, J. A. Feijo, and J. D. Becker, 2005, Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation: Plant Physiology, v. 138, no. 2, p. 744-756.

Pregitzer, K. S., and A. L. Friend, R.F. Stettler, H.D. Bradshaw, P.E. Heilman, and T.M. Hinckley eds., 1996, The structure and function of *Populus* root systems, in Biology of *Populus* and its Implications for Management and Conservation: Ottawa, NRC Research Press, p. 331-354.

Promega Corporation. DNase Usage Information. 2005.

Purgatto, E., F. M. Lajolo, J. R. O. do Nascimento, and B. R. Cordenunsi, 2001, Inhibition of beta-amylase activity, starch degradation and sucrose formation by indole-3-acetic acid during banana ripening: Planta, v. 212, no. 5-6, p. 823-828.

Qiagen. RNeasy Mini Handbook. 2001.

Qiagen. RNeasy MinElute Cleanup Handbook. 2003.

Quackenbush, J., 2001, Computational analysis of microarray data: Nature Reviews Genetics, v. 2, no. 6, p. 418-427.

Rabbani, M. A. *et al.*, 2003, Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analysis.: Plant Physiology, v. 133, p. 1755-1767.

Ramonell, K., M. Berrocal-Lobo, S. Koh, J. R. Wan, H. Edwards, G. Stacey, and S. Somerville, 2005, Loss-of-function mutations in chitin responsive genes show increased susceptibility to the powdery mildew pathogen *Erysiphe cichoracearum*: Plant Physiology, v. 138, no. 2, p. 1027-1036.

Reagin, M. J., T. L. Giesler, A. L. Merla, Resetar-Gerke J.M., K. M. Kapolka, and J. A. Mamone, 2003, TempliPhi: A sequencing template preparation procedure that eliminates overnight cultures and DNA purification.: Journal of Biomolecular Techniques, v. 14, no. 2, p. 143-148.

Reed, J. W., 2001, Roles and activities of Aux/IAA proteins in Arabidopsis: Trends in Plant Science, v. 6, no. 9, p. 420-425.

Reed, R. C., S. R. Brady, and G. K. Muday, 1998, Inhibition of auxin movement from the shoot into the root inhibits lateral root development in Arabidopsis: Plant Physiology, v. 118, no. 4, p. 1369-1378.

Reid, M. S., 1985, Ethylene and abscission: Hortscience, v. 20, no. 1, p. 45-50.

Rhee, S. Y. et al., 2003, The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community: Nucleic Acids Research, v. 31, no. 1, p. 224-228.

Ribnicky, D. M., N. Ilic, J. D. Cohen, and T. J. Cooke, 1996, The effects of exogenous auxins on endogenous indole-3-acetic acid metabolism - The implications for carrot somatic embryogenesis: Plant Physiology, v. 112, no. 2, p. 549-558.

Romano, P. G. N., P. Horton, and J. E. Gray, 2004, The Arabidopsis cyclophilin gene family: Plant Physiology, v. 134, no. 4, p. 1268-1282.

Rubin, E. S., R. N. Cooper, R. A. Frosch, T. H. Lee, G. Marland, A. H. Rosenfeld, and D. D. Stine, 1992, Realistic mitigation options for global warming: Science, v. 257, no. 5067, p. 148-&.

Russell, J. A., and B. H. Mccown, 1988, Recovery of plants from leaf protoplasts of hybrid-poplar and aspen clones: Plant Cell Reports, v. 7, no. 1, p. 59-62.

Sagerstrom, C. G., B. I. Sun, and H. L. Sive, 1997, Subtractive cloning: Past, present, and future: Annual Review of Biochemistry, v. 66, p. 751-783.

Salisbury,F, C Ross. Plant Physiology. Fourth edition. 1991. Belmont, Wadsworth Publishing Company.

Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989,Sambrook, J., E. F. Fritsch, and T. Maniatis Molecular Cloning: A Laboratory Manual: Cold Springs Harbor, Cold Springs Harbor Laboratory Press.

Schmittgen, T. D., B. A. Zakrajsek, A. G. Mills, V. Gorn, M. J. Singer, and M. W. Reed, 2000, Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: Comparison of endpoint and real-time methods: Analytical Biochemistry, v. 285, no. 2, p. 194-204.

Schoof, H., R. Ernst, V. Nazarov, L. Pfeifer, H. W. Mewes, and K. F. X. Mayer, 2004, MIPS Arabidopsis thaliana Database (MAtDB): an integrated biological knowledge resource for plant genomics: Nucleic Acids Research, v. 32, p. D373-D376.

Schulze, A., and J. Downward, 2001, Navigating gene expression using microarrays - a technology review: Nature Cell Biology, v. 3, no. 8, p. E190-E195.

Schulze, A., and J. Downward, 2001, Navigating gene expression using microarrays - a technology review: Nature Cell Biology, v. 3, no. 8, p. E190-E195.

Sedjo, R. A., N. R. Sampson, and J. Wisniewski, 1997,Sedjo, R. A., N. R. Sampson, and J. Wisniewski Economics of Carbon Sequestration in Forestry: Boca Raton, Florida, Lewis Publishers.

Shchepinov, M. S., S. C. CaseGreen, and E. M. Southern, 1997, Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays: Nucleic Acids Research, v. 25, no. 6, p. 1155-1161.

Shim, W. B., and L. D. Dunkle, 2002, Identification of genes expressed during cercosporin biosynthesis in *Cercospora zeae-maydis*: Physiological and Molecular Plant Pathology, v. 61, no. 4, p. 237-248.

Singh-Gasson, S., R. D. Green, Y. J. Yue, C. Nelson, F. Blattner, M. R. Sussman, and F. Cerrina, 1999, Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array: Nature Biotechnology, v. 17, no. 10, p. 974-978.

Slovin, J. P., and et al., 1999, Auxins, in PJ Hooykaas, MA Hall, and KR Libbenga eds., Biochemistry and Molecular Biology of Plant Hormones: Amsterdam, Elsevier, p. 115-135.

Sokal, R. R., and F.J. Rohlf, 2000, Miscellaneous methods, biometry: The Principles and Practice of Statistics in Biological Research: New York, W.H. Freeman and Company, p. 794-832.

Somerville, C., and S. Somerville, 1999, Plant functional genomics: Plant Biotechnology: Food and Feed, v. **285**, p. 380-383.

Srinivasan, K. *et al.*, 2005, Detection and measurement of alternative splicing using splicing-sensitive microarrays.: Methods, v. 37, p. 345-359.

Srivastava, L. M., 2002, Auxins, plant growth and development: Hormones and Environment: San Diego, Academic Press, p. 155-169.

Staswick, P. E., B. Serban, M. Rowe, I. Tiryaki, M. T. Maldonado, M. C. Maldonado, and W. Suza, 2005, Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid: Plant Cell, v. 17, no. 2, p. 616-627.

Stein, J., and P. Liang, 2002, Differential display technology: a general guide: Cellular and Molecular Life Sciences, v. 59, no. 8, p. 1235-1240.

Sterky, F. et al., 1998, Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags: Proceedings of the National Academy of Sciences of the United States of America, v. 95, no. 22, p. 13330-13335.

Stettler, R. F., 1968, Irradiated mentor pollen - Its use in remote hybridization of Black Cottonwood: Nature, v. 219, no. 5155, p. 746-&.

Storey, J. D., and R. Tibshirani, 2003, Statistical significance for genome-wide studies: Proceedings of the National Academy of Sciences of the United States of America, v. 100, no. 16, p. 9440-9445.

Strauss, S. H., 2003, Genetic technologies - Genomics, genetic engineering, and domestication of crops: Science, v. 300, no. 5616, p. 61-62.

Subramaniam, R., S. Reinold, E. K. Molitor, and C. J. Douglas, 1993, Structure, inheritance, and expression of hybrid poplar (*Populus trichocarpa* x *Populus deltoides*) phenylalanine ammonia-lyase genes: Plant Physiology, v. 102, no. 1, p. 71-83.

Sykes, K., M. J. Gething, and J. Sambrook, 1993, Proline isomerases function during heat-shock: Proceedings of the National Academy of Sciences of the United States of America, v. 90, no. 12, p. 5853-5857.

Sze, H. *et al.*, 2004, Expression patterns of a novel AtCHX gene family highlight potential roles in osmotic adjustment and K+ homeostasis in pollen development: Plant Physiology, v. 136, no. 1, p. 2532-2547.

Taiz L., and Zeiger E, 1998, Auxins, Plant Physiology: Sunderland, Sinauer Associates, Inc., p. 543-589.

Tam, Y. Y., E. Epstein, and J. Normanly, 2000, Characterization of auxin conjugates in arabidopsis. Low steady-state levels of indole-9-acetyl-aspartate indole-3-acetyl-glutamate, and indole-3-acetyl-glucose: Plant Physiology, v. 123, no. 2, p. 589-595.

Taylor, G., 2002, *Populus*: Arabidopsis for forestry. Do we need a model tree?: Annals of Botany, v. *90*, p. 681-698.

Thibaud-Nissen, F. O., R. T. Shealy, A. Khanna, and L. O. Vodkin, 2003, Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean: Plant Physiology, v. 132, no. 1, p. 118-136.

Thimann, K. V., 1977,Thimann, K. V. Hormone Action in the Whole Life of Plants: Amherst, University of Massachusetts Press.

Timpte, C., 2001, Auxin binding protein: curiouser and curiouser: Trends in Plant Science, v. 6, no. 12, p. 586-590.

Tiquia, S. M., L. Y. Wu, S. C. Chong, S. Passovets, D. Xu, Y. Xu, and J. Z. Zhou, 2004, Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples: Biotechniques, v. 36, no. 4, p. 664-+.

Turk, R., P. A. C. 'T Hoen, E. Sterrenburg, R. X. de Menezes, E. J. de Meijer, J. M. Boer, G. J. B. van Ommen, and J. T. den Dunnen, 2004, Gene expression variation between mouse inbred strains: BMC Genomics, v. 5.

Tuskan, G. A., 1998, Short-rotation woody crop supply systems in the United States: What do we know and what do we need to know?: Biomass & Bioenergy, v. 14, no. 4, p. 307-315.

Tuskan, G. A., and M. E. Walsh, 2001, Short-rotation woody crop systems, atmospheric carbon dioxide and carbon management: A US case study: Forestry Chronicle, v. 77, no. 2, p. 259-264.

Tuskan, G. A. et al., 2006, The genome of Western Black Cottonwood, *Populus trichocarpa* (Torr. & Bray ex Brayshaw): Science, v. *In review*.

Ulker, B., and I. E. Somssich, 2004, WRKY transcription factors: from DNA binding towards biological function: Current Opinion in Plant Biology, v. 7, no. 5, p. 491-498.

Ulmasov, T., G. Hagen, and T. J. Guilfoyle, 1997, AuxREs and AuxRE transcription factors: a model for auxin-responsive gene expression: Plant Physiology, v. 114, no. 3, p. 1255.

Ulmasov, T., J. Murfett, G. Hagen, and T. J. Guilfoyle, 1997, Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements: Plant Cell, v. 9, no. 11, p. 1963-1971.

Ulmasov, T., G. Hagen, and T. J. Guilfoyle, 1997, ARF1, a transcription factor that binds to auxin response elements: Science, v. 276, no. 5320, p. 1865-1868.

Ulmasov, T., G. Hagen, and T. J. Guilfoyle, 1999, Dimerization and DNA binding of auxin response factors: Plant Journal, v. 19, no. 3, p. 309-319.

van Bakel, H., and F. C. P. Holstege, 2004, In control: systematic assessment of microarray performance: EMBO Reports, v. 5, no. 10, p. 964-969.

Walia, H. et al., 2005, Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage: Plant Physiology, v. 139, no. 2, p. 822-835.

Walker, J. C., and J. L. Key, 1982, Isolation of cloned cDNAs to auxin-responsive Poly(A)+RNAs of elongating soybean hypocotyl: Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, v. 79, no. 23, p. 7185-7189.

Wang, H. Y., R. L. Malek, A. E. Kwitek, A. S. Greene, T. V. Luu, B. Behbahani, B. Frank, J. Quackenbush, and N. H. Lee, 2003, Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays: Genome Biology, v. 4, no. 1.

Watson, A., A. Mazumder, M. Stewart, and S. Balasubramanian, 1998, Technology for microarray analysis of gene expression: Current Opinion in Biotechnology, v. 9, no. 6, p. 609-614.

Weis, J. H., S. S. Tan, B. K. Martin, and C. T. Wittwer, 1992, Detection of rare Messenger-RNAs via quantitative RT-PCR: Trends in Genetics, v. 8, no. 8, p. 263-264.

White, C. A., and L. A. Salamonsen, 2005, A guide to issues in microarray analysis: application to endometrial biology: Reproduction, v. 130, no. 1, p. 1-13.

Wiese, A., N. Elzinga, and S. Smeekens, 2004, A conserved upstream open reading frame mediates sucrose-induced repression of translation: The Plant Cell, v. 16, p. 1717-1729.

Wilson, C. A., E. E. Cajulis, J. L. Green, T. M. Olsen, Y. A. Chung, M. A. Damore, J. Dering, F. J. Calzone, and D. J. Slamon, 2005, HER-2 over-expression differentially alters transforming growth factor-beta responses in luminal versus mesenchymal human breast cancer cells: Breast Cancer Research, v. 7, no. 6, p. R1058-R1079.

Winterberg, K. M., and *et al.*, 2005, Phenotypic screening of *Escherichia coli* K-12 Tn5 insertion libraries, using whole-genome oligonucleotide microarrays: Applied and Environmental Microbiology, v. 71, p. 451-459.

Winton, L. L., 1970, Shoot and tree production from Aspen tissue cultures: American Journal of Botany, v. 57, no. 8, p. 904-&.

Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules, 2001, Assessing gene significance from cDNA microarray expression data via mixed models: Journal of Computational Biology, v. 8, no. 6, p. 625-637.

Woodward, A. W., and B. Bartel, 2005, Auxin: Regulation, action, and interaction: Annals of Botany, v. 95, no. 5, p. 707-735.

Wright, A. D., M. B. Sampson, M. G. Neuffer, L. Michalczuk, J. P. Slovin, and J. D. Cohen, 1991, Indole-3-acetic-acid biosynthesis in the mutant maize orange pericarp, A tryptophan auxotroph: Science, v. 254, no. 5034, p. 998-1000.

Wu, C. M., Y. S. Lee, T. H. Wang, L. Y. Lee, W. H. Kong, E. S. Chen, M. L. Wei, Y. Liang, and T. L. Hwang, 2006, Identification of differential gene expression between intestinal and diffuse gastric cancer using cDNA microarray.: Oncology Reports, v. 15, no. 1, p. 57-64.

Wu, R. L., Y. F. Han, J. J. Hu, J. J. Fang, L. Li, M. L. Li, and Z. B. Zeng, 2000, An integrated genetic map of *Populus deltoides* based on amplified fragment length polymorphisms: Theoretical and Applied Genetics, v. 100, no. 8, p. 1249-1256.

Wu, Z. C., and J. K. Burns, 2003, Isolation and characterization of a cDNA encoding a lipid transfer protein expressed in 'Valencia' orange during abscission: Journal of Experimental Botany, v. 54, no. 385, p. 1183-1191.

Wullschleger, S. D., G. A. Tuskan, and S. P. DiFazio, 2002, Genomics and the tree physiologist: Tree Physiology, v. 22, no. 18, p. 1273-1276.

Wullschleger, S. D., S. Jansson, and G. Taylor, 2002, Genomics and forest biology: *Populus* emerges as the perennial favorite: Plant Cell, v. 14, no. 11, p. 2651-2655.

Wullschleger, S. D., and S. P. DiFazio, 2003, Emerging use of gene expression microarrays in plant physiology: Comparative and Functional Genomics, v. 4, no. 2, p. 216-224.

Yamamoto, K., and T. Sasaki, 1997, Large-scale EST sequencing in rice: Plant Molecular Biology, v. 35, no. 1-2, p. 135-144.

Yang, Y. H., and T. Speed, 2002, Design issues for cDNA microarray experiments: Nature Reviews Genetics, v. 3, no. 8, p. 579-588.

Yin, T. M., X. Y. Zhang, M. R. Huang, M. X. Wang, Q. Zhuge, S. M. Tu, L. H. Zhu, and R. L. Wu, 2002, Molecular linkage maps of the *Populus* genome: Genome, v. 45, no. 3, p. 541-555.

Yin, T. M., S. P. DiFazio, L. E. Gunter, S. S. Jawdy, W. Boerjan, and G. A. Tuskan, 2004, Genetic and physical mapping of Melampsora rust resistance genes in Populus and characterization of linkage disequilibrium and flanking genomic sequence: New Phytologist, v. 164, no. 1, p. 95-105.

Zhao, R., K. Gish, M. Murphy, Y. Yin, D. Notterman, W. Hoffman, E. Tom, D. Mack, and A. Levine, 2000, Analysis of p53-regulated gene expression patterns using oligonucleotide arrays: Genes and Development, v. 14, p. 981-993.

Zhu, T., 2003, Global analysis of gene expression using GeneChip microarrays: Current Opinion in Plant Biology, v. 6, no. 5, p. 418-425.

Zsuffa, L., E. Giordano, L. Pryor, and R. Stettler, R.F. Stettler, J. Bradshaw, P.E. Heilman, and T.M. Hinckley eds., 1996, Trends in Poplar Culture: Some global and regional perspectives, in Biology of Populus and its implications for management and conservation: Ottawa, NRC Research Press, p. 515-537.

Vita

Sara Jawdy was born to Lyle and Janice Miller of Spartansburg, Pennsylvania on October 22nd, 1977. There, she attended Corry Area High School. Upon graduating in 1996 she attended the Penn State University, studying Horticulture with an emphasis in plant breeding and genetics. After earning a B.S. from Penn State in 2000 and taking two years off from school to work as an intern at Oak Ridge National Lab in the Environmental Sciences Division, she entered the M.S. program in Plant Sciences at the University of Tennessee while continuing to work at ORNL in Dr. Gerald Tuskan's plant genomics lab.