# Architecture and Evolution of Xylem-related Gene Coexpression Networks in Poplars

by

Haktan Suren

Thesis submitted to the Faculty of

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

FORESTRY

Jason A. Holliday, Chair

Eric P. Beers

Amy M. Brunner

Roderick V. Jensen

February 14, 2013

Blacksburg, VA

**Keywords:** Populus, xylem, bioinformatics, evolution

**Architecture and Evolution of Xylem-related Gene Coexpression Networks in Poplars**

Haktan Suren

## ABSTRACT

With the advent of sequencing technologies, a variety of methods have been available day by day. Each of these methods have helped scientists to for a deeper understanding of the biological function and evolutionary constraints on the relevant genes, which can be gained through the use of modern computational approaches. Numerous approaches have being developed to advance these goals, and interaction network mapping is one of them. This method has been employed to study a variety of organisms to illustrate shared (conserved) or individual (unique) properties, and is mainly based on identifying and visualizing modules of co-expressed genes. As being a very strong candidate for such tools, co-expression gene network was used in this study to indentify the genes in wood formation of *Populus trichocarpa* with the help of the other novel bioinformatics tools such as Gene Ontology and Cytoscape.

In order to booster the accuracy of the findings, we have combined it with an evolutionary approach, synonymous and non-synonymous ratio (dN/dS) of the proteins to show the selective patterns of the genes in a comparative fashion between woody and non-woody plants.

This thesis is proposed to help plant scientists to gain insights into the genes that are involved in wood formation. By taking advantage of the computational studies have been done on this paper, one can validate the experiments along with reducing the cumbersomeness of the lab trials on the topic of wood formation in plants.

## ACKNOWLEDGMENTS

I would like to thank my chair Dr. Jason A. Holliday and my entire committee members: Dr. Eric P. Beers, Dr. Amy M. Brunner, Dr. Roderick V. Jensen for their constant help, reassurance, motivation, enthusiasm, guidance, counsel during my graduate program and for the opportunity to work on such an important project. Especially, my chair Dr. Jason A. Holliday for his insights, supports, encouragement and invaluable and detailed critiques of my drafts and for taking the time to discuss matters with me. His friendship and patience is also greatly appreciated. I feel very special and lucky to have such a great committee that works well together and is interested in their students' academic growth. I learned so much outside the classroom from them and really appreciate the time and effort each committee member put into my growth as a scientist.

I am also thankful for my fellow students, Rajesh, Regis and Mandy, our lab manager Xiaoyan and research associate Lecong and their camaraderie that I am delighted during graduate school and as a part of Forestry program.

Finally, and most importantly, I am grateful to my wife, Leman for her loving support and encouragement. I will always be thankful for her patience and understanding during this endeavor. I also wish to thank my family, my parents Recep and Betul, sister and brother, Zumrut and Emre, who were always there for me and supported me when I needed.

This research was supported by graduate research assistantship, funded by the Department of Forest Resources and Environmental Conservation and Dr. Holliday's research funds.

# TABLE OF CONTENTS

**ABSTRACT** ..... **ii**

**ACKNOWLEDGMENTS** ..... **iii**

**TABLE OF CONTENTS** ..... **iv**

**LIST OF FIGURES** ..... **v**

**LIST OF TABLES** ..... **vi**

**OBJECTIVES & JUSTIFICATION** ..... **1**

**I. LITERATURE REVIEW** ..... **3**

*I.1. Wood* ..... *3*

*I.1.1. Vascular Cambium* ..... 3

*I.1.2. Lignin* ..... 4

*I.1.3. Xylem* ..... 5

*I.1.4. Phloem* ..... 5

*I.2. Wood Biosynthesis* ..... *6*

*I.2.1. Wood cells divisions* ..... 6

*I.2.2. Radial cell expansion* ..... 7

*I.2.3. Secondary cell wall deposition* ..... 7

*I.2.4. Lignification* ..... 8

*I.3. Factors affecting wood properties and variation in wood cells* ..... *9*

*I.3.1. During The Growing Season* ..... 9

*I.3.2. During The Aging Process* ..... 10

*I.3.3. Response to Stem Displacement* ..... 10

*I.4. Molecular Mechanism & Regulation* ..... *10*

*I.5. Populus spp. As A Model Organism for Studies of Wood Formation* ..... *12*

**II. GENE INTERACTION NETWORK AND GENE PREDICTION IN XYLEM** ..... **13**

*II.1. Introduction* ..... *13*

*II.2. Materials and Methods* ..... *15*

*II.2.1. Weighted Gene Correlation Network Analysis (WGCNA)* ..... 17

*II.2.2. Gene Ontology (GO)* ..... 18

*II.3. Results and Discussion* ..... *18*

*II.4. Conclusion* ..... *33*

**III. RELATIONSHIP BETWEEN GENE CONNECTIVITY AND EVOLUTIONARY RATE IN BLACK CLUSTER** ..... **34**

*III.1. Introduction* ..... *34*

*III.2. Materials and Methods* ..... *35*

*III.3. Results and Discussion* ..... *36*

*III.4. Conclusion* ..... *45*

**LITERATURE CITED** ..... **46**

**APPENDIX A. Supplementary Tables and Figures** ..... **59**

# LIST OF FIGURES

*Figure 1: A, Drawing a transverse section of the cambial zone (CZ) showing fusiform (F), ray (R) initial cells, phloem (P) and xylem (X). B, showing the orientation of the compartments across the stem (Plomion, et al., 2001) www.plantphysiol.org Copyright American Society of Plant Biologist* ..... *7*
*Figure 2: Sample clustering to detect outliers among the samples* ..... *19*
*Figure 3: Analysis of the network topology for various soft-thresholding power β. Due to the summary mean connectivity decrease steeply along with the increasing soft thresholding power (B), it is more preferable to choose the lowest power possible that satisfies the approximate scale free topology (A).* ..... *19*
*Figure 4: Clustering dendrogram and corresponding colors of top clusters* ..... *21*
*Figure 5: Expression pattern of black, turquoise, green clusters* ..... *22*
*Figure 6: Gene co-expression network generated in Cytoscape. Colors correspond to those in Table 3 and Figure 3.* ..... *23*
*Figure 7: Average connectivity ('degree') of each cluster* ..... *24*
*Figure 8: Log-transformed p-value of the GO terms in the black cluster at cutoff >10* ..... *27*
*Figure 9: Relationship between gene connectivity and evolutionary rate in Arabidopsis thaliana* ..... *42*
*Figure_Apx 1: Expression pattern of brown, magenta and red clusters* ..... *59*
*Figure_Apx 2: Expression pattern of blue, grey, and pink cluster* ..... *60*
*Figure_Apx 3: Relationship between gene connectivity and evolutionary rate in Oryza Sativa* ..... *68*
*Figure_Apx 4: Relationship between gene connectivity and evolutionary rate in Pinus taeda* ..... *69*
*Figure_Apx 5: Relationship between gene connectivity and evolutionary rate in Populus tremula* ..... *70*

# LIST OF TABLES

*Table 1: Populus sections and species. Reprinted from (Mellerowicz, et al., 2001) with kind permission of Springer Science and Business Media. Copyright © 2001, Springer Science and Business Media.* ..... *11*
*Table 2: Gene expression data included in this study* ..... *16*
*Table 3: Number of genes that each cluster have after weight cut-off of w=0.2* ..... *20*
*Table 4: Top clusters and corresponding GO terms with p-values (top three GO terms only).* ..... *26*
*Table 5: Black cluster genes upregulated >30-fold in xylem relative to phloem tissue (A. Brunner, unpublished data).* ..... *31*
*Table 6: Gene list in black cluster annotated as "unknown" function* ..... *32*
*Table 7: Genes that have high dn/ds in comparison with non-woody organisms that do not show up as having high dn/ds when compared with woody organisms* ..... *43*
*Table_Apx 1: Genes according to their connectivity* ..... *61*
*Table_Apx 1: Genes according to their connectivity (contd.)* ..... *62*
*Table_Apx 2: GO analysis for turquoise cluster (p < E-10)* ..... *63*
*Table_Apx 3: GO analysis for green cluster (p < E-10)* ..... *63*
*Table_Apx 4: GO analysis for brown cluster (p < E-10)* ..... *63*
*Table_Apx 5: GO analysis for black cluster (p < E-10)* ..... *64*
*Table_Apx 6: Genes that are upregulated in xylem (unpublished data)* ..... *65*
*Table_Apx 6: Genes that are upregulated in xylem (unpublished data) (contd.)* ..... *66*
*Table_Apx 6: Genes that are upregulated in xylem (unpublished data) (contd.)* ..... *67*

## OBJECTIVES & JUSTIFICATION

Since the development of high-throughput gene expression technology such as microarray and high-throughput sequencing, enormous amounts of data have become publicly available. In order to make better use of these data, new analytical approaches and bioinformatics data analysis methods are constantly being developed. Constructing gene interaction networks is one of the techniques that has been widely used in recent years. Revealing the function of an unannotated protein has always been a difficult task, and gene interaction networks may provide insights into gene function by identifying co-regulated (co-expressed) transcriptional modules.

As genes on the same pathway are expected to exhibit similar expression patterns under various physiological conditions, one common way of studying microarray data is to group genes according to their expression profile similarities (Ruan, Dean, & Zhang, 2010). The goal of constructing a gene co-expression network is to identify highly correlated groups of genes across the multiple experimental samples. These groups of genes, referred to as clusters, are often overrepresented by similar functions (Xiang, Zhang, & Huang, 2012). Co-expression can be quantified by pair-wise gene expression correlation (e.g. Pearson correlation coefficient), which is based on the expression profiles of two genes. Since correlation is a symmetrical measurement, one cannot deduct the directional relationship in which one gene is downstream of another (H. Li, Sun, & Zhan, 2009). These co-expression networks are typically illustrated by nodes, representing individual genes, and edges, the weights of which reflect the correlation value of each gene pair (Mao, Van Hemert, Dash, & Dickerson, 2009). One challenge of this method is determining the threshold for correlation values that connects two genes by an edge. In order to address this issue, different network algorithms have been developed for selecting appropriate thresholds based on network theory (e.g. network topology) (Xiang, et al., 2012). Weighted gene

co-expression network analysis (WGCNA) is one of these algorithms and have been widely used to identify gene clusters that are functionally relevant to cellular processes (Langfelder & Horvath, 2008).

Gene co-expression networks have proven useful in analyzing microarray data in model organisms including yeast, mice, and humans (Mao, et al., 2009). In this study, several *Populus* spp. gene expression data sets were analyzed using co-expression network analysis, with the goal of better understanding the gene networks involved in wood formation and other processes in poplars. The putative functional relevance of the resulting clusters were investigated using the Gene Ontology (GO) classification system, and several candidate genes from these network analyses were studied using T-DNA insertion mutants in *Arabidopsis thaliana*. Finally, an evolutionary approach was taken to better understand the role that network connectivity may play in constraining positive selection, and to investigate adaptive evolution along phylogenetic lineages comprising both woody and non-woody species.

# I. LITERATURE REVIEW

## I.1. Wood

Wood is the world's most abundant natural and continuously renewable resource for the paper, timber and bioenergy industries and is expected to play a major role as a cost-effective substitute for fossil fuels. Developing wood cells not only provide fibers and lignocelluloses for bioenergy but also act as crucial sinks for excess atmospheric $CO_2$, mitigating the effects of climate change (Plomion, Leprovost, & Stokes, 2001). Wood comprises the secondary xylem of vascular plants and derives from cumulative annual activity of the cambium – a lateral meristem that continually renews the xylem and phloem over successive years of growth in woody perennials (Plomion, et al., 2001). In early spring, rapid expansion of cambial cells results in primary cell wall synthesis, and later in the season the majority of xylem cells undergo secondary wall synthesis, also called lignification. This secondary wall structure, typically comprised of 25% lignin, and 70% carbohydrates (cellulose and hemicelluloses), not only determines the physical and chemical features of wood but also its energy contents (X. Li, Wu, & Southerton, 2010; Novaes, Kirst, Chiang, Winter-Sederoff, & Sederoff, 2010).

The chemical content of the primary walled tree wood differs from the secondary walled tree. Primary walled tree wood contains 47% pectin, 23% cellulose, 18% glycans, 10% protein and 3% other material while secondary walled tree wood has 48% cellulose, 23% glycans, 19% lignin, 10% other materials (Mellerowicz, Baucher, Sundberg, & Boerjan, 2001).

### *I.1.1. Vascular Cambium*

The vascular cambium is a secondary meristem derived from the procambium and enables the radial growth of gymnosperm and angiosperm roots and shoots, and wood formation (xylogenesis). The cambium is comprised of meristematic cells organized in radial files, and

differentiates outward to produce secondary phloem and inward to produce secondary xylem. Theoretically, each file contains initial cells, called xylem mother cells, residing in the meristem that develop into xylem, and other cells, called phloem mother cells, that develop into phloem. Cambial activity ensures the perennial life of the trees through regular renewal of phloem and xylem (Lachaud, Catesson, & Bonnemain, 1999; Mellerowicz, et al., 2001; Plomion, et al., 2001). The cambium, though similar to other meristems in that its main function is cell division and differentiation, also has some unique features of its own such as having two morphologically distinct cell types: axially elongated fusiform cambial cells (FCC) and isodiametrical ray cambial cells (RCC). These cells are programmed to develop axial and horizontal cell systems in the secondary xylem and phloem (Mellerowicz, et al., 2001). Meristematic activity is usually positively correlated with growth rate. Taking into account the function of the vascular cambium, there are two important determinants for the rate of the wood production: first, the number of the xylem mother cells and, second, the duration of the cell cycle in xylem mother cells (Mellerowicz, et al., 2001).

### *I.1.2. Lignin*

Lignin is the second most abundant natural plant compound on earth after cellulose (Lacombe, Van Doorsselaere, Boerjan, Boudet, & Grima-Pettenati, 2000) and is a phenolic polymer complex most present in both the secondary cell wall and between adjacent cell walls in the middle lamellae. Lignin fills the spaces in the cell wall, thereby providing rigidity and cohesiveness of the wood between cellulose, hemicelluloses, and pectin components in xylem cells. The hydrophobic surface of lignin allows the tree to transport water to very high leaves, while also maintaining the mechanical strength of the tree. A secondary function of lignin is to

serves as a barrier against the invasion of pests and pathogens (Novaes, et al., 2010; Plomion, et al., 2001).

The biosynthesis of lignin is achieved via the phenylpropanoid pathway, which begins with phenylalanine either imported from phloem or synthesized *de novo* by the plastid-localized shikimate pathway. In addition to lignin, this pathway diverges to produce compounds involved in plant development and defense, such as flavonoids and some phenolics compounds (Lacombe, et al., 2000; Novaes, et al., 2010).

### *I.1.3. Xylem*

Xylem forms a specialized conducting channel that carries water and solutes throughout the plant and differentiated from procambium derived from apical meristem and vascular cambium. It is in charge of conducting water and nutrients obtained by roots of the plant to the shoots above ground. Besides that, it also carries the plant hormones such as abscisic acid and cytokinin. It also helps the plant maintain its rigidity and supplies structural support. Xylem develops on the adaxial or internal pole of the vascular bundles. In vascular plants, xylem is located at the center of the stem. Xylem is comprised of conducting tracheary elements called parenchyma cells, which are vessel elements and nonconducting elements called xylary fibers (Dinneny & Yanofsky, 2004; Gifford & Foster, 1989; Ye, 2002).

### *I.1.4. Phloem*

Phloem is a plant tissue that is differentiated from procambium and vascular cambium. It is the other conducting tissue that transports products of photosynthesis (sucrose) from generative tissues such as leaves to pool tissues like roots. It also provides a channel for transportation of the proteins and mRNAs, which are involved in plant growth and development. Phloem develops on the peripheral or abaxial pole of the vascular bundles. In vascular plants, phloem surrounds

the xylem tissues on the stem. Phloem is comprised of sieve elements and nonconducting elements such as parenchyma cells and fibers (Dinneny & Yanofsky, 2004; Gifford & Foster, 1989; Ye, 2002).

## I.2. Wood Biosynthesis

Wood (secondary xylem) is produced via five key steps including cell division, cell expansion, cell wall thickening, programmed cell death, and heartwood formation (Plomion, et al., 2001).

### *I.2.1. Wood cells divisions*

Xylem cells divide more than phloem cells, which explains the significance disproportion existing between phloem and xylem tissues. As it was mentioned previously, the cambial zone has two types of highly vacuolated cells give rise to fusiform and ray initials (Figure 1A). Rays are in charge of transporting the nutrients between phloem and xylem facilitated through plasmodesmata on tangential walls, while elongated fusiform initials divide longitudinally producing the secondary vascular tissues (wood elements) on the inner side in gymnosperms, and phloem cells in angiosperm dicots. Radial division of the fusiform initials also ensures an even increase of the perimeter of the cambium (Figure 1B). The ratio of fusiform and ray initials is species-specific and highly dependent on the age of the cambium (Lachaud, et al., 1999; Plomion, et al., 2001).

FCCs develop into three different cell types: axial parenchyma, fibres, and vessel elements, while RCCs give rise to two cell types: isolation and contact ray cells. Cambial cells and their immediate derivatives (e.g. xylem, phloem, vessel elements) are the main determinant of wood properties. This also suggests that the cell fate is determined at the very early stage (Mellerowicz, et al., 2001).

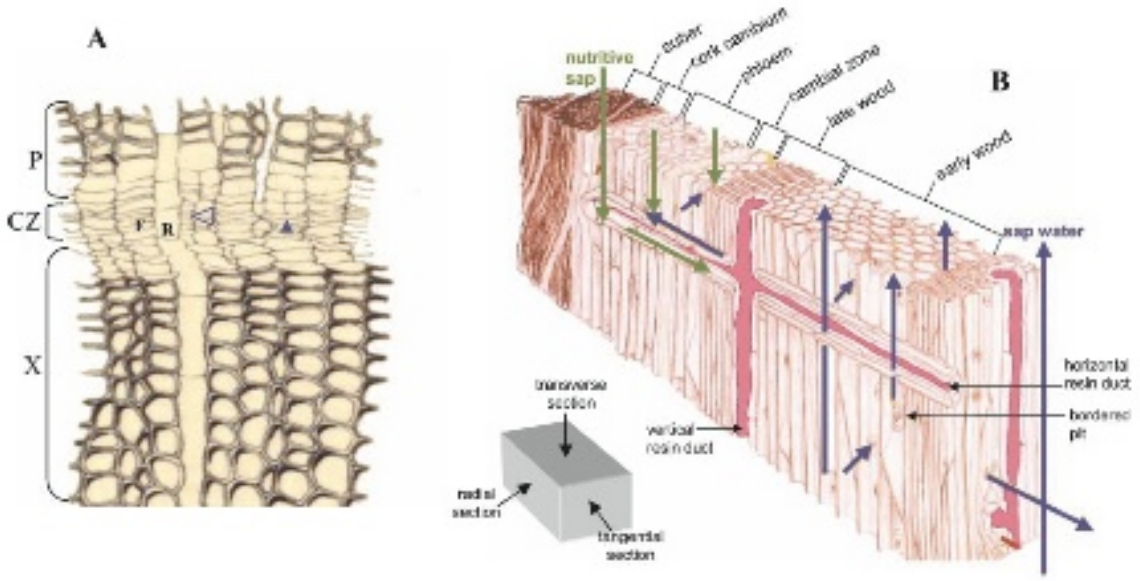



**Figure 1: A, Drawing a transverse section of the cambial zone (CZ) showing fusiform (F), ray (R) initial cells, phloem (P) and xylem (X). B, showing the orientation of the compartments across the stem (Plomion, et al., 2001) www.plantphysiol.org Copyright American Society of Plant Biologist.**

### *I.2.2. Radial cell expansion*

After xylem mother cells leave the meristem, they grow into a primary walled stage. This phase corresponds to the radial expansion (RE) zone where vessel elements and fibers display different morphological properties. While the cells remain in this zone, they enlarge to different sizes, and gain different types of polarity (Mellerowicz, et al., 2001). Due to the variations in cell enlargement among cell types, each cell type must have been programmed differently in terms of its gene regulation to maintain the cell turgor pressure and cell wall plasticity (Mellerowicz, et al., 2001).

### *I.2.3. Secondary cell wall deposition*

When xylem cells reach their final size, the formation of the secondary cell wall begins, regulated by the co-expression of numerous genes involved in biosynthesis and assembly of four major chemical compounds: polysaccharides (cellulose and hemicelluloses), lignin, cell wall

proteins and other compounds (flavonoids, terpenoids, pectins) (Plomion, et al., 2001). The orientation of the cellulose microfibrils is random, longitudinal, and deposited densely to prevent further radial expansion. The S1 (outer layer), S2 (middle layer) and S3 (inner layer) layers of the cell wall are also formed during this phase. (Gibson, 2012; Mellerowicz, et al., 2001). In populus, secondary cell wall formation starts with the vessel elements. The vessel elements forms a three-layered secondary cell wall similar to that of fibres but the S2 layer is relatively thinner. The ray cells also form three-layered secondary cell wall with three S layers (Mellerowicz, et al., 2001). After all the vessel elements have formed, they are joined together end-to-end to produce functional long vessel in wood. After that, a secondary wall is formed around the pits and perforations and eventually lignifies whereas no secondary wall and lignification in the pits and perforation occurs

### *I.2.4. Lignification*

Wood is comprised of 40%-50% cellulose, 25% hemicelluloses, and 25%-35% lignin, although these values vary among species. Lignification begins in vessel elements and detected first in middle lamella at cell corners. Lignin deposition is most intense when the S3 layer is formed and progresses towards the cell lumen until all of the wall layers have lignified (Mellerowicz, et al., 2001). When lignification is complete, xylem elements undergo apoptosis, which involves hydrolysis of the protoplast regulated by the phytohormones such as auxins and cytokinins. After this process, all of the cellular content except the secondary cell wall is hydrolyzed (Mellerowicz, et al., 2001; Plomion, et al., 2001).

Because of its importance to wood quality and, more recently, bioenergy, lignin biosynthesis is probably the most studied pathway unique to woody perennials. The interest in lignin biosynthesis has led to the identification and cloning of several structural and regulatory genes.

Nevertheless, due to the complexity of the process, our ability to engineer lignin biosynthesis is incomplete suggesting that there are still unknown players in this pathway (Plomion, et al., 2001). Recent studies showed that a large number of genes involved in the lignin biosynthesis pathway significantly and differentially expressed in the developing xylem such as phenylalanine ammonia-lyase (*PAL*), cinnamate 4-hydroxylase (*C4H*) promoter, 4-coumarate:CoA ligase (4CL) and cinnamoyl-coenzyme A reductase (COMT) (Baucher, et al., 1996; Baucher, et al., 1995; Franke, et al., 2000; Hu, et al., 1999; Jouanin, et al., 2000; Lapierre, et al., 1999; Mellerowicz, et al., 2001; Meyermans, et al., 2000; Ranocha, et al., 2000; Tsai, et al., 1998; Zhong, Ripperger, & Ye, 2000).

### I.3. Factors affecting wood properties and variation in wood cells

A distinction is commonly made between two highly divergent taxa with regard to their wood properties – the softwood gymnosperms and the hardwood angiosperms – although there are some exceptions to this demarcation. Wood variability is due to the heterogeneity of the cell types and their diverse structures and may also be variable within the same species due to anatomical, physical and chemical differences that have both genetic and environmental underpinnings (Plomion, et al., 2001).

#### *I.3.1. During The Growing Season*

Annual cambium activity is affected by temperature and photoperiod in temperate region. Early wood is formed early in the growing season whereas late wood is formed in late summer or autumn when the temperature and daylight are favorable for the tree. The transition between one type wood cells to another is rapid and therefore visible between two season but not within a single growing season (Plomion, et al., 2001). Xylem cell length increases from minimum to

maximum during the transition from earlywood to latewood, and cell wall crytallinity also increases from earlywood to latewood (Mellerowicz, et al., 2001).

*I.3.2. During The Aging Process*

The passage from juvenile wood to mature wood is one of the factors that affect wood formation. The main difference results from the density which is the number of the cells per unit volume and microfibrils angle which is the angle between the long axis of the fibres and tracheids and cellulose microfibrils winding around the cell (helically). Longitudinal variation exists in vessel length and fiber along the tree trunk and the radial diameter of vessel elements increases from pith to bark. Because of that, the xylem cell volume increases along with age of the cambium (Barnett & Bonham, 2004; Mellerowicz, et al., 2001; Plomion, et al., 2001).

*I.3.3. Response to Stem Displacement*

When the environment is unfavorable (e.g. prevailing winds, snow, slop), an abnormal type of wood forms as a part of the developmental process to enable the tree to find a more favorable position or orientation. This type of wood formation is associated with eccentric radial growth of the displaced stem (Plomion, et al., 2001). During the bending (leaning) process, a high rate of cell production is observed in the tension area, whereas wood formation on the opposite side is inhibited. Other anatomical changes include the reduction in vessel size and density and formation of gelatinous fibres (Mellerowicz, et al., 2001).

**I.4. Molecular Mechanism & Regulation**

Plant hormones, when applied to plant exogenously, affect cambial growth in many aspects such as cell division, cell expansion, cell differentiation and even cell chemistry (Little & Savidge, 1987; Mellerowicz, et al., 2001). Auxin is one of the key regulators of xylogenesis. It is the only hormone that is adequate to stimulate cell differentiation of vascular elements on its

own. Recent studies have shown that there is positive correlation between auxin concentration and mersitemetic activity, meristem size, and xylem cell production. It has also been shown that auxin enhances lignification process (Fukuda & Komamine, 1980; Mellerowicz, et al., 2001; Phillips & Arnott, 1983; Savidge, 1983; Sundberg & Little, 1990; Uggla, Moritz, Sandberg, & Sundberg, 1996; Q. Wang, Little, & Oden, 1997). Auxin also works with other plant hormones such as gibberellins. When applied with auxin, gibberellins stimulates meristematic activity along with xylem fibre elongation (Aloni, 1979). Although it is well known that cytokinins promote cell division in plant, their role has not been very clear in cambial growth (Little & Savidge, 1987). Ethylene also stimulates cambial cell division via auxin support (Abel, Nguyen, Chow, & Theologis, 1995; Eklund & Little, 1996; Mellerowicz, et al., 2001). Genes involved in phythormonal biosynthesis and regulations are controlled by a strict transcriptional regulation during every step of differentiation. In addition, continuously changing environment also contributes to this process. Due to this interaction, variations can be detected not only among different species but also within same species (Villalobos, et al., 2012).

**Table 1: *Populus* sections and species. Reprinted from (Mellerowicz, et al., 2001) with kind permission of Springer Science and Business Media. Copyright © 2001, Springer Science and Business Media.**

| Section | Species |
|---|---|
| Leuce (aspen type) | P. grandidentata (Michx.) |
| | P. alba (L.) |
| | P. tremula (L.) |
| | P. tremuloides (Michx.) |
| Aigeiros (cottonwood or poplar type) | P. deltoides (Bartr. ex. Marsh.) |
| | P. sargentii (Dode) |
| | P. fremontii (Wats.) |
| | P. nigra (L.) |
| Tacamahaca (balsam poplar type) | P. balsamifera (L.) |
| | P. maximowiczii (Henry) |
| | P. trichocarpa (Torr. & Gray) |
| | P. angustifolia (James) |
| Leucoides (swamp poplar type) | P. heterophylla (L.) |
| Turanga | P. euphratica (Olivier) |

### I.5. *Populus* spp. As A Model Organism for Studies of Wood Formation

*Populus* species have emerged as the premier model of woody perennials for studies of tree biology in general and wood formation in particular. The *Populus* genus is very diverse, comprising thirty species native to the Northern hemisphere (Table 1). This diversity facilitates genetic improvement for economically important traits related to growth rate, adaptability, paper quality, wood properties, etc. (Mellerowicz, et al., 2001; Tuskan, et al., 2006). Because *Populus* species occupy diverse environments with respect to temperature, longitude, daylength, and exhibit differences in developmental features such as presence of juvenile and mature phases, and secondary extensive growth, the genus is an amenable model system for many questions related to tree biology, in addition to the economic and ecological importance of the genus.

*Populus* has a number of features that make it an amenable model system for genetics: a modest genome size of only 550 Mb that is fully sequenced, ease of propagation from rooted cuttings, relative ease of gene transformation via *Agrobacterium*-mediated technique, and availability of the genetic tools (sequence, microarray data) (Mellerowicz, et al., 2001; Tuskan, et al., 2006). These advantages of *Populus* spp. facilitate research in diverse areas such as association mapping, quantitative trait locus (QTL) mapping, molecular biology and map based cloning, and bioinformatics (Breen, Glenn, Yeager, & Olson, 2009; Frewen, et al., 2000; Induri, et al., 2012; Morreel, et al., 2006; Rae, Tricker, Bunn, & Taylor, 2007; Sjodin, Street, Sandberg, Gustafsson, & Jansson, 2009; Street, et al., 2006; Woolbright, et al., 2008; R. Wu, Bradshaw, & Stettler, 1997; R. Wu, et al., 2003; Yin, et al., 2002).

## II. GENE INTERACTION NETWORK AND GENE PREDICTION IN XYLEM

### II.1. Introduction

Gene co-expression networks may be determined from microarray expression experiments where a number of experimental conditions are profiled. These networks rely on a "guilt by association" heuristic and many related studies have been published based on this method with a variety of organisms (Childs, Davidson, & Buell, 2011; Dewey, et al., 2011; Fuller, et al., 2007; Horvath & Dong, 2008; Lin, et al., 2011; Mochida, Uehara-Yamaguchi, Yoshida, Sakurai, & Shinozaki, 2011; Movahedi, Van de Peer, & Vandepoele, 2011; Prieto, Risueno, Fontanillo, & De las Rivas, 2008; Reverter, et al., 2006; Torkamani, Dean, Schork, & Thomas, 2010).

In recent years, advances in high-throughput technologies (e.g. gene expression microarrays) have resulted in an enormous increase in genomics data, and the challenge becomes one of interpreting these data in a biological context (Maere, Heymans, & Kuiper, 2005). The accumulation of data from these high-throughput experiments allows for a deeper understanding of the biological function and evolutionary constraints on the relevant genes, which can be gained through the use of modern computational approaches (Jordan, Marino-Ramirez, Wolf, & Koonin, 2004). Numerous approaches have been developed to advance these goals, and interaction network mapping is one of them. This method has been employed to study a variety of organisms to illustrate shared (conserved) or individual (unique) properties, and is mainly based on identifying and visualizing modules of co-expressed genes (Oldham, Horvath, & Geschwind, 2006).

Weighted gene co-expression network analysis (WGCNA) is an open source R-package to analyze high-dimensional data. This software is a systems biology tool to identify correlated expression patterns among genes across multiple microarray samples, which facilitates

functional annotation. WGCNA focuses on modules of the genes rather than expression of individual genes, and as such may reveal relationships between a module and a specific biological function in the cell, e.g., the xylem development network. This feature makes WGCNA a very attractive software to identify candidate genes relevant to a particular process of interest, allowing for analysis of network architecture, module detection, calculation of topological properties, and visualization (Langfelder & Horvath, 2008). Annotation of clusters revealed by WGCNA may be achieved by subsequently identifying statistically overrepresented Gene Ontology (GO) terms (Maere, et al., 2005).

Understanding the genes, and their interactions, that underlie wood formation provides the foundation for genetic manipulation to modify its structure for a variety of end uses, including conversion for bioenergy, pulp and paper, and saw timber. Several studies have employed gene expression microarrays to better understand which genes are involved in wood formation. Schrader et al. (2004) generated a high-resolution transcript profile across the cambial zone of aspen (*Populus tremula*) for more than 13,000 genes across six anatomically homogenous cell layers in the meristem zone and proposed the position of stem and proliferating cells in the cambial zone (Jarmo Schrader, et al., 2004). Goue et al. (2008) separated two neighboring initials that play a major role in wood production: fusiform cambial cells (FCCs), which produce the axial cell system, and ray cambial cells (RCCs), which produce the rays, from the cambial meristem of poplar (*Populus trichocarpa x Populus deltoides)* and found that photosynthesis genes were enriched in RCCs. They also denoted other cell type-specific expression patterns such as xyloglucan and pectin metabolism in RCCs and FCCs (Goué, et al., 2008). Dharmawardhana et al. (2010) used the most comprehensive poplar microarray to date to study stem segments that spanned primary to secondary growth, and identified differentially expressed

genes during stem development (Dharmawardhana, Brunner, & Strauss, 2010). We employed these data for co-expression network analysis to better understand the expression networks underlying wood formation in *Populus*.

### II.2. Materials and Methods

Co-expression of poplar genes was studied based on microarray data of three different studies comprising 28 experimental conditions, which were obtained from Schrader et al. (2004), Goué et al. (2008) and Dharmawardhana et al. (2010). The first study included a high-resolution transcript profile across the cambial meristem of *Populus tremula*. In this study, the cambial region of the aspen dissected out eleven different sections and the expression profile of the 13,000 genes were measured across each section (prefixed as A from end of the phloem to end of xylem, prefixed as B from start of phloem to start of xylem). The positions annotated by B3, B4, B6, A3 are located in phloem, whereas B8, B9, A4, A5, A6, A7 are in cambial zone, and B10, A9, A10 in the xylem. In addition, expression data was generated from the cambial zone, expansion zone, zone of secondary cell wall formation and late maturation (shown from A to E), as well as meristem tissue from the apex, cambium, root tip, and mature leaf (Jarmo Schrader, et al., 2004). In another study, Schrader et al, studied transcriptome remodeling in cambial meristem during dormancy in *Populus tremula*. Gene expressions in active growth were compared with those in dormancy (Ratio A.D) (J. Schrader, et al., 2004). Goue et al. analyzed the two neighboring initials of the meristem of poplar (*Populus trichocarpa x Populus deltoides)* – fusiform cambial cells (FCCs) and ray cambial cells (RCCs) – and produced a transcriptional database where differential expressions of the two cell types were shown (Goué, et al., 2008). Finally, genome-wide transcriptome analysis of *Populus trichocarpa* of the progression from primary to secondary stem development was used. In this study, stem segments from successive

internodes below the apex (plastochron indices 2,3,4, and 5) and further down the stem (internode 9) were analyzed (Dharmawardhana, et al., 2010).

The R (Team, 2010) statistical package Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder & Horvath, 2008) was used to analyze these data, and the results were visualized in Cytoscape (Shannon, et al., 2003). The expression profiles of a merged dataset including a total of 10,164 genes from the four different Poplar studies were used. Each study and the corresponding data are summarized in Table 2. Annotations were gathered from the Joint Genome Institute (JGI) *P. trichocarpa* genome build version 1.0 (Grigoriev, et al., 2012; Tuskan, et al., 2006) database using a custom script to conduct automated reciprocal BLASTX.

In order to verify the computational study, knockout experiments (reverse genetics) were conducted to identify the genes that are involved in xylem biosynthesis. As a first step, knockout seed stocks that were ordered from TAIR, (Lamesch, et al., 2012) were planted and kept in the growth chamber between 3-4 weeks. Once they were grown enough to develop siliques, stem samples were carefully taken from each gene knockout plant. Stem samples were stabilized in agar gel horizontally and cross-sectional samples (100μ thick) were obtained by using microtomes. Then, each sample was visualized and captured under a light microscope under 50X and 100X magnification.

**Table 2: Gene expression data included in this study**

| Description | Conditions | Reference |
|---|---|---|
| High-resolution transcript profile across the cambial zone of aspen. | 22 | (Jarmo Schrader, et al., 2004) |
| Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome | 1 | (J. Schrader, et al., 2004) |
| Cell type specific gene expression patterns between ray and fusiform initials within the cambial meristem | 1 | (Goué, et al., 2008) |
| Transcriptome analysis of the transition from primary to secondary stem development in Populus trichocarpa. | 5 | (Dharmawardhana, et al., 2010) |
| TOTAL | 28 | |

### *II.2.1. Weighted Gene Correlation Network Analysis (WGCNA)*

The WGCNA package first takes the gene sets and evaluates them according to the number of missing fields and those with zero variance, and eliminates those genes that have significant numbers of missing values and zero variance. The minimum fraction of non-missing samples for a gene is set to 0.5 in this case. The software then checks if there are any outliers among the samples by a hierarchical clustering method using constant-height tree cut. In our case, samples were clustered by averaging the expression values, and the corresponding dendrograms were then used to detect outliers. Next, a parameter called soft-thresholding power (β) was calculated, which is a cutoff value to calculate adjacency of the genes/clusters under a scale-free network (Langfelder & Horvath, 2008). Like many other network types, the degree distribution of a biological network follows a power law: the fraction of nodes in the network increases as the number of connections gets larger. A tradeoff is involved in choosing the proper β. Larger β values can provide a better network, but networks with large β values take more time to calculate. Sometimes it is impossible to calculate the network on a standard computer since the program relies on high processor speed. Hence, the challenge is choosing the best β with an acceptable scale-free model fit. With the selected β value, the genes were clustered according to their adjacency and co-expression value. WGCNA also has the capability to produce cluster dendrograms and a feature to visualize gene network using a heatmap plot. Finally, WGCNA creates output files that allow viewing in Cytoscape, which is a powerful pipeline software for visualizing a variety of types of network data and integrating with the other information, such as annotations, expression level, etc.

*II.2.2. Gene Ontology (GO)*

In order to functionally characterize the clusters, we annotated them using Gene Ontology (GO). For this purpose, we used BINGO to identify GO terms that were significantly overrepresented in a cluster. We were thus able to annotate the co-expression network interactively within Cytoscape. To detect overrepresented GO terms, the hypergeometric test was used, and in order to control type 1 error (false positive) rate, Benjamini and Hochberg correction is applied with 0.05 significance level on biological process category only (Maere, et al., 2005).

**II.3. Results and Discussion**

During data cleaning, 536 genes were excluded from the analysis due to missing values and zero variance across the samples. No outlier samples were identified (Figure 2). The soft-thresholding power parameter, β, was set at 12 a compromise between a scale-free network fit and the highest power possible (roughly r=0.9). β values larger than 12 gave a slightly better fit for the network, but not substantially more so, and values larger than 14 resulted in a poorer fit (Figure 3A). The mean connectivity declined steeply with increasing soft-thresholding values (Figure 3B). Hence, it is advantageous to choose the lowest β that meets the scale-free topology criterion.

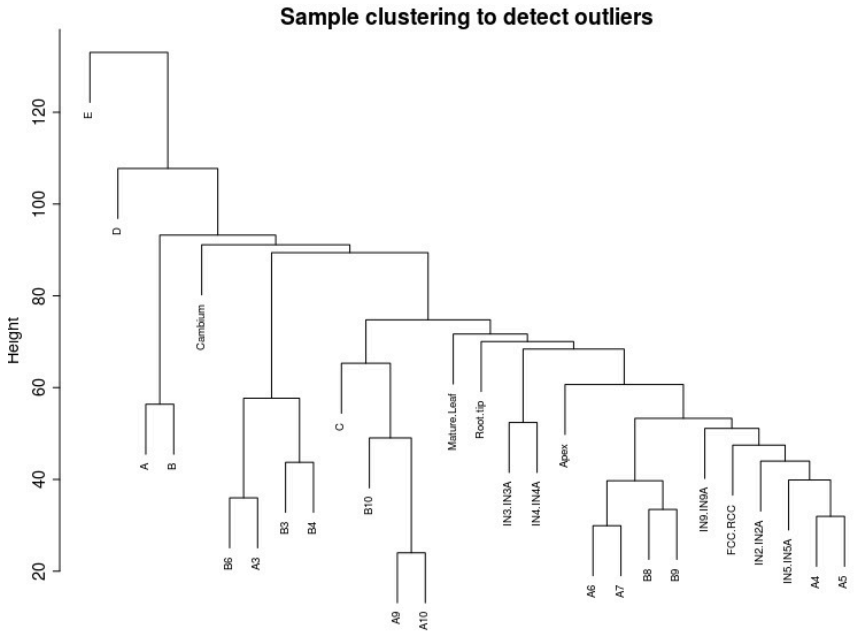



**Figure 2: Sample clustering to detect outliers among the samples.**

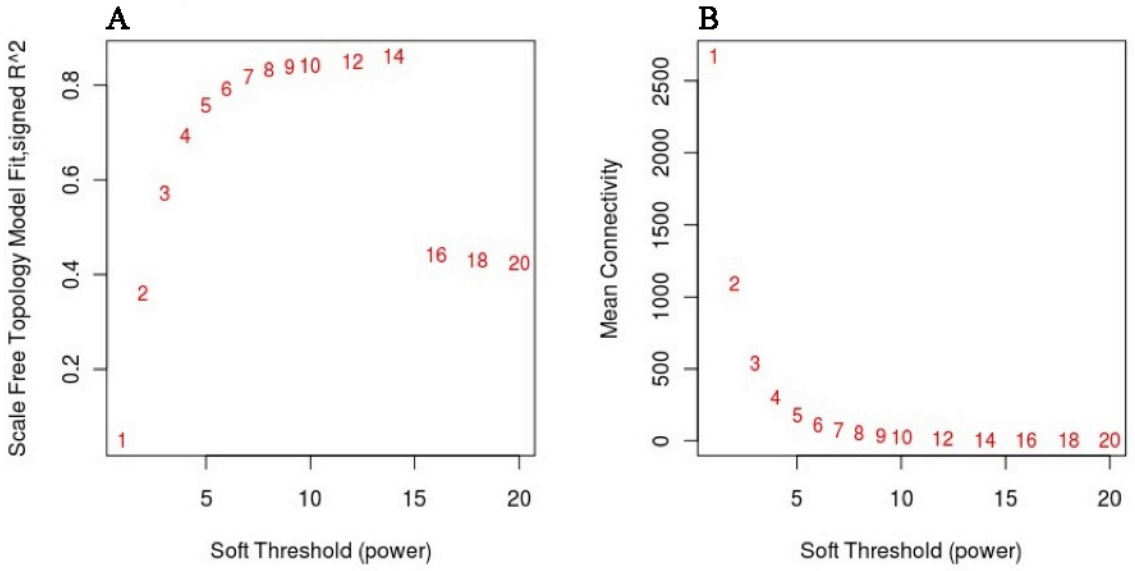



**Figure 3: Analysis of the network topology for various soft-thresholding power β. Due to the summary mean connectivity decrease steeply along with the increasing soft thresholding power (B), it is more preferable to choose the lowest power possible that satisfies the approximate scale free topology (A).**

The resulting network comprised 17 clusters with between 45 and 1466 genes. Modules one and two consist of almost 20% of all of the genes (module 0 is reserved for genes outside of

other modules), 1364 in this case. After a weight cutoff (w=0.2), which is a threshold of adjacency of included edges, was applied each group was minimized and assigned to a particular color code (Table 3). Higher weights resulted in a more complex network in terms of the number of genes included. The turquoise, green, brown, black and magenta clusters had the largest number of genes (Table 3 and Figure 4).

The robustness of the clustering algorithm has been shown as depicting the expression pattern of each individual cluster in Figure 5. Higher degrees of similarity in gene expression patterns increase the likelihood that the genes will be grouped into the same cluster. Expression pattern for the other major clusters can be found in APPENDIX A.

From the dendrogram, it is apparent that the turquoise cluster had the greatest connectivity given its minimal branching structure. Most genes in this cluster fell into a single clade, although a second small clade was evident (Figure 4). By contrast, the black cluster had two distinct clades with a relatively high degree of within-clade connectivity, and the green cluster was highly branched, indicating less connectivity. Interestingly, the magenta and brown clusters had a high degree of inter-cluster connectivity relative to the other clusters (Figure 4 and Figure 6).

**Table 3: Number of genes that each cluster have after weight cut-off of w=0.2**

| **Cluster** | **#of genes** | **Cluster** | **# of genes** | **Cluster** | **# of genes** |
|---|---|---|---|---|---|
| turquoise | 410 | blue | 31 | lightcyan | 8 |
| green | 275 | grey | 44 | tan | 7 |
| brown | 212 | pink | 23 | cyan | 6 |
| black | 158 | purple | 19 | greenyellow | 3 |
| magenta | 116 | yellow | 15 | Total | 1399 |
| red | 63 | midnightblue | 9 | | |

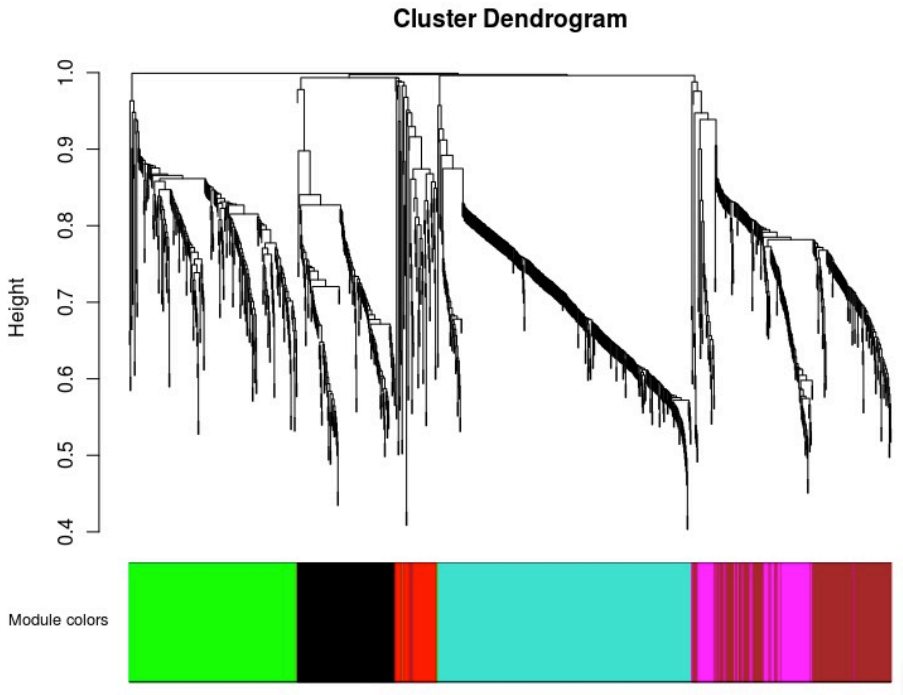



**Figure 4: Clustering dendrogram and corresponding colors of top clusters.**

Following analysis with WGCNA, we used Cytoscape to visualize the results (Figure 6), and the software BINGO was used to identify GO terms overrepresented within each cluster. The turquoise cluster was the largest, and was significantly enriched with genes involved in translation and gene expression. This result was expected because in most of the cases, the largest cluster is comprised of housekeeping genes, which are involved various functions in cell metabolism (Lercher, Urrutia, & Hurst, 2002). Genes in this cluster also have higher connectivity since they tend to be constitutively expressed (Figure 7). Gene list according to their connectivity can be found in APPENDIX A.

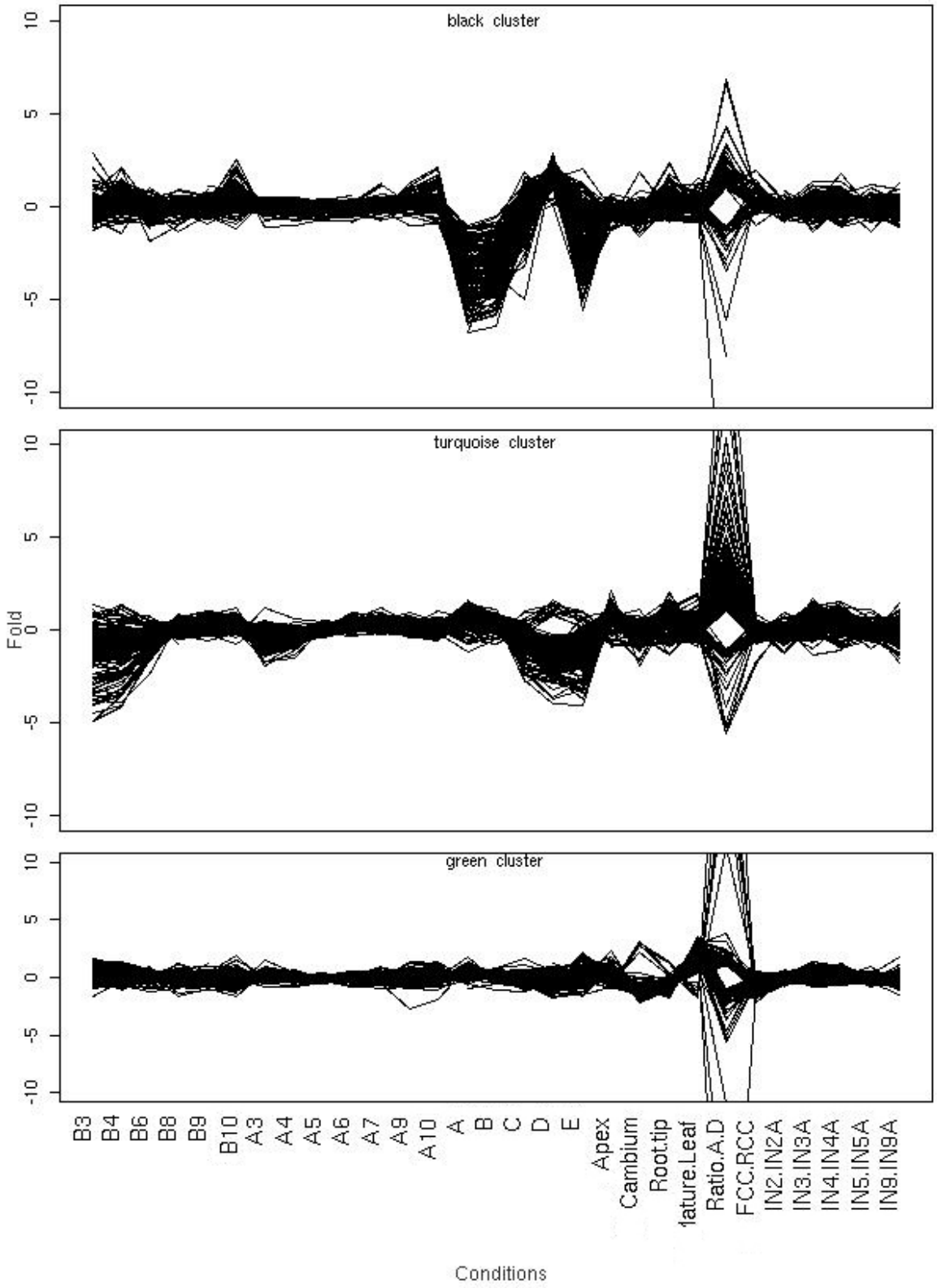



**Figure 5: Expression pattern of black, turquoise, green clusters**

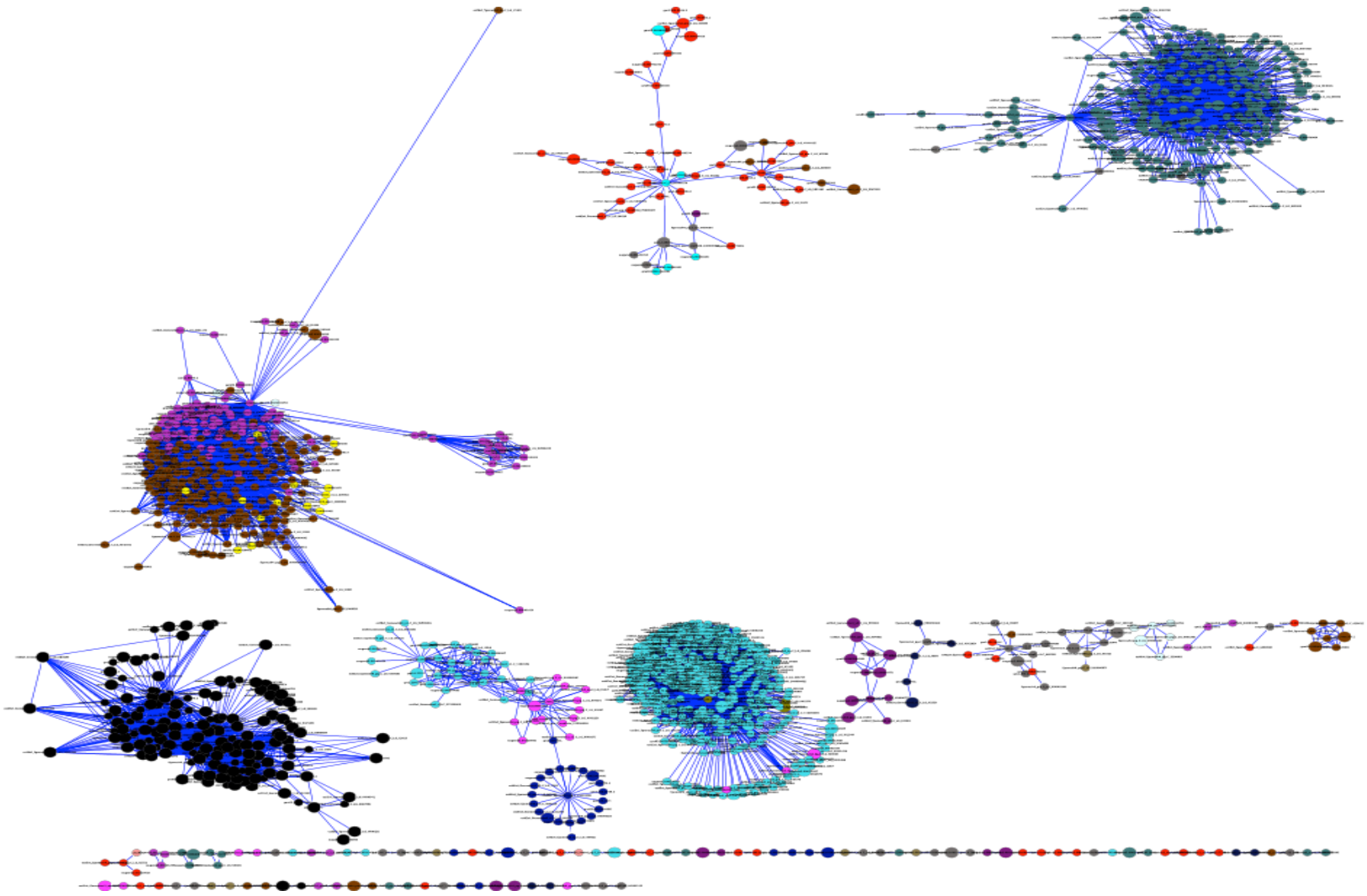


**Figure 6: Gene co-expression network generated in Cytoscape. Colors correspond to those in Table 3 and Figure 3.**

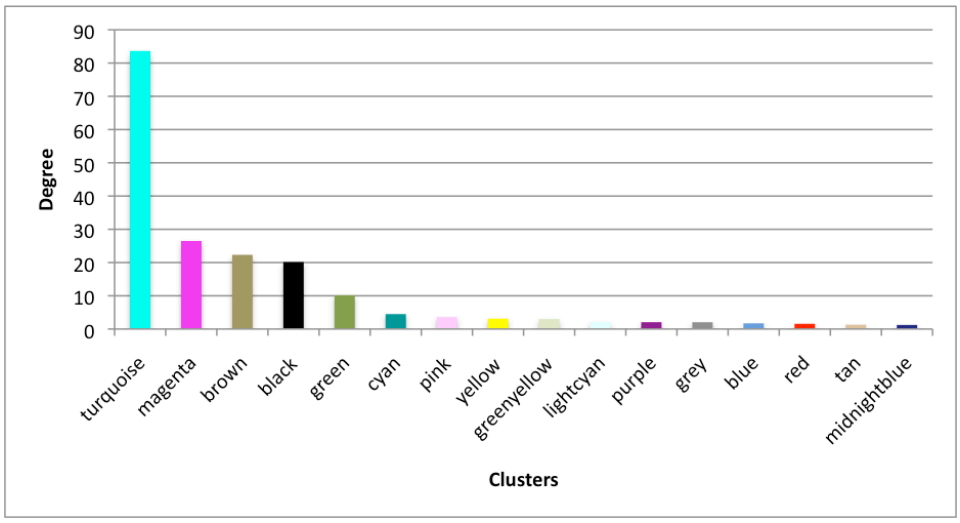



**Figure 7: Average connectivity ('degree') of each cluster**

The turquoise cluster is enriched for the structural and cellular activities and most of the genes are located in cytosol and ribosome according to the cellular component analysis. The green cluster was significantly enriched for GO terms that include photosynthesis, light reaction and precursor metabolites and energy, the brown cluster for response to water deprivation, response to water, and response to abiotic stimulus, and the black cluster for secondary cell wall biogenesis, cell wall biogenesis, and cellular cell wall organization. The remaining clusters were not enriched for any GO terms (Table 4). More detailed results can be found in APPENDIX A

Expression patterns of the individual clusters provide some clues about their function. Genes in the turquoise cluster are downregulated in the phloem and xylem (B3, B4, B6, B10, A3, A9, A10, A) whereas they are upregulated in the cambial zone, suggesting greater biosynthesis activity in the cambial zone that in the neighboring tissues (B8, B9, A4, A5, A6, A7, B, C, D, E). The genes in the green cluster were upregulated in the phloem (B3, B4, B6, A3) and mature leaves, but flat in the cambial zone (B8, B9, A4, A5, A6, A7) and xylem (B10, A9, A10), and slightly downregulated in the root tip. These results align well with the result from gene ontology

that the genes in this cluster are involved in the photosynthetic process and located in chloroplasts according to cellular component analysis. For the brown cluster, tissue-specific expression was observed. These genes were highly upregulated in the phloem (B3, B4, B6, A3), and highly downregulated in the xylem (B10, A9, A10). Significant expression pattern change was observed during the life cycle of the plant, which is the transition from active phase to dormancy (see the sample name: Ratio.A.D). Most of the genes in this cluster were differentially expressed in the dormancy phase of the plant. Interestingly, GO analysis showed that the majority of the genes in this cluster are highly overrepresented by response to stress conditions. The black cluster revealed a pattern of gene expression suggestive of involvement in wood formation, as expression levels increased through the tissue transition from phloem (A3, A4, A5, A6, A7) to xylem (A9, A10), as well as from the cambial zone (A) to the late maturation zone of the cell wall (B, C, D) (downregulated in the cambial zone but increased in the zone of secondary cell wall formation and decreased in the maturation zone (E)). This variation and differentiation between non-woody to woody tissue is explained well by the gene ontology annotation as well. According to GO analysis, secondary cell wall biogenesis and cell wall organization and biogenesis have been overrepresented significantly (e = e-17). Furthermore, cellular component analysis showed that the majority of the genes are located in the cell periphery, plasma membrane, golgi, apoplast, and cytoskleton. The magenta cluster shows very steep changes in the transition from the phloem to the cambial zone and again from the cambial zone to the xylem. Genes are downregulated in phloem (B3), highly downregulated in xylem (B10, A9, A10) and upregulated in the cambial zone (B8, B9, A4, A5, A6, A7, A). Unfortunately, GO analysis has not shed significant light on this relationship.

**Table 4: Top clusters and corresponding GO terms with p-values (top three GO terms only).**

| Cluster | GO ID | p-value | Description |
|---|---|---|---|
| turqoise | 6412 | 3.11E-49 | translation |
| turqoise | 9059 | 2.05E-46 | macromolecule biosynthetic process |
| turqoise | 34645 | 6.47E-46 | cellular macromolecule biosynthetic process |
| green | 15979 | 4.11E-35 | photosynthesis |
| green | 19684 | 6.28E-17 | photosynthesis, light reaction |
| green | 6091 | 3.57E-12 | generation of precursor metabolites and energy |
| brown | 9414 | 7.22E-04 | response to water deprivation |
| brown | 9415 | 7.22E-04 | response to water |
| brown | 9628 | 3.79E-02 | response to abiotic stimulus |
| black | 9834 | 2.02E-17 | secondary cell wall biogenesis |
| black | 42546 | 8.05E-17 | cell wall biogenesis |
| black | 70882 | 2.52E-16 | cellular cell wall organization or biogenesis |
| pink | 90357 | 2.11E-02 | regulation of tryptophan metabolic process |
| pink | 90358 | 2.11E-02 | positive regulation of tryptophan metabolic process |
| pink | 48016 | 2.11E-02 | inositol phosphate-mediated signaling |

As the black cluster was significantly enriched with genes involved in cell wall biogenesis and metabolism, we chose to focus on this cluster. GO terms to which a large number of black cluster genes were annotated include cell wall macromolecule metabolism, as well as primary and secondary cell wall organization and biogenesis (Figure 8).

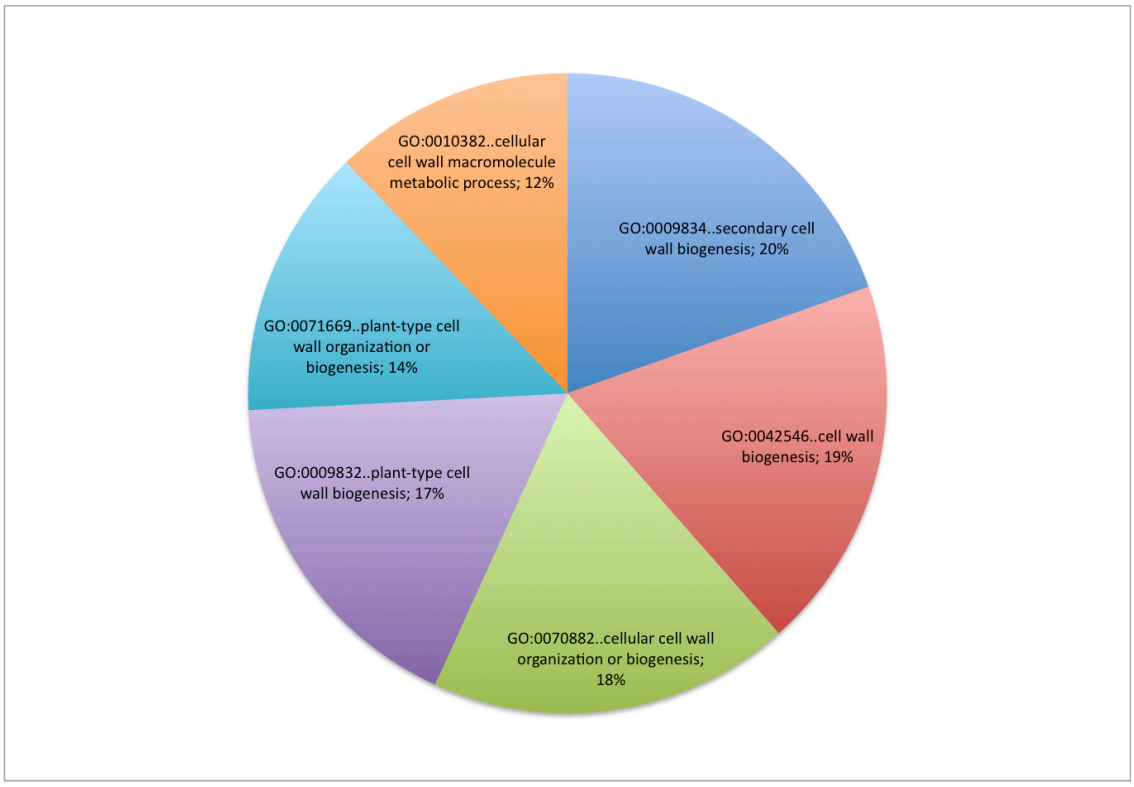



**Figure 8: Log-transformed p-value of the GO terms in the black cluster at cutoff >10**

A large number of genes in this cluster have been previously characterized as being involved in cell wall formation, and an unpublished microarray study showed that many were upregulated in xylem relative to phloem tissue (Table 5 and APPENDIX A). In total, 72% of the genes in the black cluster were upregulated in the xylem.

AT1G20850, when it was first studied, was considered to be one of the apoptotic (cell death) genes (Oh, Park, & Han, 2003). Later, it was identified in apoptoplastic fluids of *Arabidopsis thaliana* rosettes as a cell wall protein by mass spectrometry (Boudart, et al., 2005). The comparative studies have shown that this gene is differentially expressed in tension wood, possibly having a role in the formation of fibres with thick inner cell wall layers (Andersson-Gunneras, et al., 2006). Recent studies argued that it regulates cysteine endopeptidase in xylem

by being highly expressed during xylem differentiation (Ko, Yang, Park, Lerouxel, & Han, 2007). The high level expression of this gene in xylem tissues aligned very well with those studies.

AT3G49260 was first mentioned as one of the members of IQD (calmodulin-binding nuclear protein) gene families that promotes glucosinilate accumulation and plant defense in Arabidopsis thaliana (Abel, Savchenko, & Levy, 2005). Later, it was also shown that it regulates pathogen response and cell cycling during a virus infection (Ascencio-Ibanez, et al., 2008). Recently, it has also been shown that it was duplicated in the genome so the knockout mutants of this gene will not show any phenotype due to either the compensation role of the duplicated copy or triggering the alternative pathway. As it is well known, wood healing is a major mechanism for plant defense; thus, one can expect that these genes which are involved in secondary cell wall biogenesis will have high expression during the infection and healing processes.

AT1G20850 was first revealed as a cysteine peptidase in xylem in 2000 (Zhao, Johnson, Kositsup, & Beers, 2000). Later, reverse genetic studies showed that it is one of the genes responsible in the programmed cell death stage of the sequential events of secondary growth (Oh, et al., 2003). With the help of genome-wide expression profiling, it has been shown that it is differentially expressed in xylem six-fold more than phloem and non-vascular tissues (Zhao, Craig, Petzold, Dickerman, & Beers, 2005). Recent studies on cell wall thickening and secondary cell wall growth have shown the similar results (Andersson-Gunneras, et al., 2006; Boudart, et al., 2005; Hanada, et al., 2011; Mitsuda, Seki, Shinozaki, & Ohme-Takagi, 2005; Oh, et al., 2003; Ohashi-Ito, Kubo, Demura, & Fukuda, 2005; Suh, et al., 2005).

AT5G23750 was shown to be differentially expressed in tension wood and wood comparative studies as possible inducer of fibers with cell wall layer (Andersson-Gunneras, et al., 2006). This study also suggest that the rigidity and strength of the wood during mechanical induction or

stress can be maintained and sustained by excessive biosynthesis of the secondary cells. It should also be noted that in conifers tension/compression wood is distinguished from other wood by higher levels of lignin. In short, this particular gene might be involved in lignin biosynthesis. Recently, the same gene in another study was also identified as one of the drought induced transcripts (C. E. C. E. Wong, et al., 2006).

AT5G67210 was first identified as irregular xylem (IRX) likely involved in xylan biosynthesis (Jensen, et al., 2011). A similar study was performed where the (protein?/gene?) was named IRX15L due to its similarity to AT3G50220 (IRX15) (D. Brown, et al., 2011).

AT5G12250 was first identified as beta-6 tubulin involved in cell wall biosynthesis differentially expressed in xylem versus bark (Oh, et al., 2003). It was shown in pollen transcriptome study as it is constitutively expressed but not expressed in pollen (Becker, Boavida, Carneiro, Haury, & Feijo, 2003). The same result was obtained by the RT-PCR study in Arabidopsis (Usadel, et al., 2004). A more recent study has shown that it is also categorized as proteins responsive to saline stress (Jiang, Yang, Harris, & Deyholos, 2007).

AT5G03760 was first shown to be involved in cell wall biosynthesis by encoding the catalytic subunits of cellulose synthase (Roberts & Bushoven, 2007). And then, it was described as cellulose sythase-like protein and named CSLA9 (Goubet, et al., 2009).

AT5G54690 was first shown to function in secondary cell wall formation identified by a knockout experiment and named irx8 mutant (Bischoff, Cookson, Wu, & Scheible, 2009; D. M. Brown, et al., 2007; D. M. Brown, Zeef, Ellis, Goodacre, & Turner, 2005; Q. Li, et al., 2011; Persson, et al., 2007). Later, it was discovered that it is specifically expressed in fibres and vessels and exhibits similar expression pattern to the organ development (Pena, et al., 2007).

Borner et al. showed that AT3G27200 is one of the glycosylphosphatidylinositol-ancgored proteins involved in root development, cell wall integrity and adhesion (Borner, Lilley, Stevens, & Dupree, 2003; Borner, Sherrier, Stevens, Arkin, & Dupree, 2002).

AT1G30900 was described as a member of a novel plant specific (ZIM) GATA factor gene family, which ultimately up-regulates cell wall modification genes in elongated plants (Ascencio-Ibanez, et al., 2008; Shikata, et al., 2004).

AT2G28110 was identified by generating a knockout mutant which decreases secondary cell wall formation and also the expression profile of this gene is correlated with frigida genes (irx7, fra8) (D. M. Brown, et al., 2007; D. M. Brown, et al., 2005; Lee, O'Neill, Tsumuraya, Darvill, & Ye, 2007; Lee, Teng, Huang, Zhong, & Ye, 2009; C. Lee, R. Zhong, et al., 2007; Persson, et al., 2007; A. M. Wu, et al., 2010; Zhong, et al., 2005; G. K. Zhou, et al., 2006).

AT1G03080 was exhibited in tension wood versus normal wood comparative genome studies as a differentially expressed myosin class 11 heavy chain, which is involved in secondary cell wall biosynthesis (Andersson-Gunneras, et al., 2006).

It was first shown that the knockout mutant of AT2G38080 exhibits weak irx phenotype (irx12), which is involved in xylem vessel formation and secondary wall formation of treachery elements containing interfascicular fibers and xylem cells but not in parenchymatous pith cells (Andersson-Gunneras, et al., 2006; D. M. Brown, et al., 2005; Ko, et al., 2007; Koizumi, Yokoyama, & Nishitani, 2009; Kubo, et al., 2005; Mitsuda, et al., 2005; Sawa, Demura, Horiguchi, Kubo, & Fukuda, 2005; J. Zhou, Lee, Zhong, & Ye, 2009). In some other studies, it was also mentioned that it is involved in lignin biosynthesis (Berthet, et al., 2011; Ranocha, et al., 1999; Sibout, et al., 2005; J. Zhou, et al., 2009).

AT2G40270 is identified as putative kinase family protein in several studies but none of them have mentioned about its role in secondary wall function or xylem (Ascencio-Ibanez, et al., 2008; Dievart & Clark, 2003; Goda, et al., 2004; Hoth, et al., 2003; Navarro, et al., 2004; Shiu & Bleecker, 2003).

**Table 5: Black cluster genes upregulated >30-fold in xylem relative to phloem tissue (A. Brunner, unpublished data).**

| Poplar Gene Model | AT Homologue | Fold Change | Description |
|---|---|---|---|
| fgenesh4_pg.C_LG_II000045 | AT1G20850.1 | 85.63 | tracheary element vacuolar protein |
| eugene3.00150136 | AT3G49260.2 | 62.68 | alpha/beta-Hydrolases superfamily protein |
| estExt_Genewise1_v1.C_LG_V2210 | AT1G20850.1 | 59.3 | xylem cysteine peptidase 2 (XCP2) |
| estExt_fgenesh4_pm.C_LG_XV0452 | AT5G23750.2 | 51.98 | remorin family protein |
| eugene3.00050506 | AT5G67210.1 | 49.18 | Encode a DUF579 |
| grail3.0018029802 | AT5G12250.1 | 44.32 | beta-tubulin |
| estExt_fgenesh4_pm.C_LG_VIII0087 | AT5G03760.1 | 40.22 | encodes a beta-mannan synthase |
| estExt_fgenesh4_pm.C_LG_XIII0357 | AT5G54690.1 | 36.76 | putative galacturonosyltransferase activity |
| gw1.I.3002.1 | AT3G27200.1 | 36.25 | Cupredoxin superfamily protein |
| estExt_fgenesh4_pm.C_LG_III0520 | AT1G30900.1 | 34.78 | vacuolar sorting receptor |
| grail3.0001137701 | AT2G28110.1 | 34.78 | a member of glycosyltransferase family |
| eugene3.00031337 | AT1G03080.1 | 33.82 | Kinase interacting |
| estExt_fgenesh4_pg.C_LG_VIII0541 | AT2G38080.1 | 32.67 | laccase activity |
| estExt_fgenesh4_pg.C_LG_I0941 | No Hits | 31.12 | Remorin family protein |
| gw1.I.2205.1 | AT2G40270.1 | 30.91 | Protein kinase family protein |
| gw1.VIII.2926.1 | AT1G13635.1 | 30.7 | DNA glycosylase superfamily protein |
| estExt_Genewise1_v1.C_LG_VI2188 | AT5G17420.1 | 30.48 | Encodes a xylem-specific cellulose synthase |

AT5G17420 has been extensively studied for years and it is very well concluded that AT5G17420 encodes cellulose synthase subunit AtCesA7 that is also essential for cellulose synthesis during secondary cell wall formation. This locus has been screened several times for

cell wall properties. It is also known as irregular xylem (irx3), fragile fiber (fra5) and murus 10 (mur10) (Atanassov, Pittman, & Turner, 2009; Betancur, et al., 2010; Bischoff, et al., 2009; Bosca, et al., 2006; D. M. Brown, et al., 2007; D. M. Brown, et al., 2005; Carroll, et al., 2012; Ha, et al., 2002; Hernandez-Blanco, et al., 2007; Jones, Ennos, & Turner, 2001; Kim, et al., 2012; Liang, et al., 2010; Samuga & Joshi, 2002; Szyjanowicz, et al., 2004; Taylor, 2007; Taylor, Howells, Huttly, Vickers, & Turner, 2003; Zhong, Morrison, Freshour, Hahn, & Ye, 2003).

In contrast, many members of the black cluster had no annotation either in Arabidopsis or poplar, the Arabidopsis orthologs of which are annotated either as "unknown protein" or "protein of unknown function" (Table 6). We were not able to discover further information about those genes by using GeneMANIA (Mostafavi, Ray, Warde-Farley, Grouios, & Morris, 2008).

**Table 6: Gene list in black cluster annotated as "unknown" function**

| Poplar Gene Model | AT Homologue | Fold Change | Annotation |
|---|---|---|---|
| estExt_Genewise1_v1.C_LG_XIV1980 | AT2G41610 | 8.06 | unknown protein |
| grail3.0140003901 | AT3G13275 | 11.00 | unknown protein; |
| gw1.X.5904.1 | AT5G43150 | 16.68 | unknown protein; |
| grail3.0004021203 | AT1G27690 | 3.73 | Protein of unknown function (DUF620) |
| estExt_fgenesh4_pm.C_LG_XI0311 | AT4G27435 | 6.92 | Protein of unknown function (DUF1218) |
| fgenesh4_pm.C_LG_XV000296 | AT4G24910 | 10.34 | Protein of unknown function (DUF579) |
| eugene3.00110990 | AT5G54240 | 18.38 | Protein of unknown function (DUF1223) |
| gw1.IX.4902.1 | AT5G60720 | 20.97 | Protein of unknown function (DUF547) |
| gw1.86.114.1 | AT1G09610 | 28.44 | Protein of unknown function (DUF579) |
| grail3.0096001401 | AT3G14170 | 10.06 | Plant protein of unknown function (DUF936) |
| eugene3.00050506 | AT5G67210 | 49.18 | Encode a DUF579 (domain of unknown function 579) |
| estExt_Genewise1_v1.C_LG_I2969 | AT3G01810 | 5.86 | Molecular function unknown |
| eugene3.00061618 | AT5G11890 | 21.26 | Molecular function unknown |
| estExt_Genewise1_v1.C_LG_XIV2984 | AT2G42700 | NA | Molecular function unknown |

In an attempt to better understand the possible functions of these genes in xylem formation, we identified Arabidopsis T-DNA insertion lines for the relevant orthologs, and phenotyped them. However there were no obviously abonormal xylem phenotypes in these plants. This could be due to more subtle effects of these genes.

### II.4. Conclusion

This study showed the underlying molecular mechanism of wood formation and the complexity of the process. Even a single process/function in a plant is maintained by the up- and down-regulation of the hundreds and thousands of genes together. As we illustrated in this study, it is possible to show all the players in any biological pathway at statistically significant levels by simply generating a gene network. However, this does not necessarily help to identify individual gene functions in the pathway accurately.

For future study, new bioinformatics tools need to be introduced in science to ease the study of the additive effects of the genes. Furthermore, new simulation techniques are needed to reduce the cumbersome work of dealing with many genes for lab trials.

# III. RELATIONSHIP BETWEEN GENE CONNECTIVITY AND EVOLUTIONARY RATE IN BLACK CLUSTER

## III.1. Introduction

Protein-protein interaction degree (PPID) has been shown to be one of the most important predictors determining evolutionary rate (Pang, Cheng, Xuan, Sheng, & Ma, 2010). It has been claimed that the proteins with high gene connectivity (hubs) are less dispensable physiologically, in other words, more likely to be essential (e.g. knockout-inviable), and having multiple binding sites makes them slow evolving (Batada, Hurst, & Tyers, 2006).

A negative correlation between evolutionary rate and gene connectivity was first introduced by Fraser at al (H. B. Fraser, Hirsh, Steinmetz, Scharfe, & Feldman, 2002). Soon after, there have been many debates on whether this relationship is merely an artifact or confounded by other genomic variables (Pang, et al., 2010). This controversy has been demonstrated by some studies are biased towards finding more interaction for abundant proteins and abundant proteins evolve more slowly (Bloom & Adami, 2004).

Scientists have focused more on hub genes, which are defined as genes that have many interaction partners, in different species and criticized the robustness of the method by inventing different techniques (Aragues, Sali, Bonet, Marti-Renom, & Oliva, 2007; Batada, Reguly, et al., 2006; Batada, et al., 2007; Bertin, et al., 2007; H. Fraser, 2005).

Similar findings were observed in studies of human and mouse gene expression where it was suggested that the hubs of the co-expression network evolve more slowly on average than those with fewer co-expressed partners (the spokes), and genes that are co-expressed have similar evolutionary rates (Jordan, et al., 2004)

Similar to other studies shown above, in this study, co-expression network analysis has been used to better understand the interaction between natural selection and network connectivity.

In this study, the dN/dS ratio has been used to define the selection pressure acting on protein-coding regions. This technique is commonly used due to its wide-range generic application to different organisms, the robustness of the statistics embedded in its algorithm, and the simplicity of its use. This measure is simply identified by taking the ratio of the rate of substitutions occurred at non-silent sites (dN), which possibly experience selection, to the ratio of substitutions at silent sites (dS), which are recognized neutral. Having a ratio larger than one can be interpreted as neutral selection and promotes the protein sequence changes where as having less than one silences the protein changes. Although this method was meant to be used for analyzing protein sequences of divergent species, it may be applied even within a single population. In that regard, we aimed to identify the genes that are involved in wood formation by comparing their evolutionary rates in woody and non-woody plants (Kryazhimskiy & Plotkin, 2008).

### III.2. Materials and Methods

In order to study the relationship between gene connectivity and evolutionary rate, we focused on all the genes that are possibly related in xylem formation, which is the black cluster from previous section.

To evaluate the hypothesis that network connectivity should be correlated with evolutionary rate, reciprocal tblastx (Altschul, et al., 1997) was used to identify homologous genes in *Arabidopsis thaliana*, *Oryza Sativa, Populus tremula* and *Pinus taeda*. These species were chosen both for their available genomic resources and to represent the diverse life histories of plants (i.e., monocots and dicots, and angiosperm and gymnosperm trees). After ORFs of each

genes were found, pairwise dN/dS ratios were calculated using the CODEML function in PAML (Z. Yang, 2007). As a codeml parameter we used pairwise maximum likelihood method (ML), calculated single dN/dS ratio for all lineages (branches) and assumed equal codon frequencies across the sequences (1/61).

To do this, we employed a custom script that aligns orthologous ORFs sequences and calculates the dN/dS ratio (Stajich, 2005). Higher dN/dS ratios indicate a relative excess of non-synonymous mutations and a higher evolutionary rate (positive selection), and small values for dN/dS indicate purifying selection or a slower evolutionary rate.

### III.3. Results and Discussion

Based on our data, we failed to show negative correlation between evolutionary rate and gene connectivity directly. This might be due to having only small set of samples. We studied around 150 genes but it is obviously not enough for representing all genes in wood formation process. There is a significant chance that we failed to show the negative correlation due to very high variation among the subset we analyzed (black cluster). This statement also explains having trend-wise or local correlations in the data. Therefore, we binned the data by various degree intervals and, in every case, we were able to show negative correlation as expected. Surprisingly, the strength of the negative correlation increased as the degree interval is enlarged. Figure 9 shows the existence of negative correlation between gene connectivity and evolutionary rate in *Arabidopsis thaliana*. Similar analysis for the other organisms can be found in APPENDIX A. The relationship in *Oryza Sativa, Populus tremula* and *Pinus taeda* is not as strong as that of *Arabidopsis thaliana* due to having less sequence information.

There is no evidence for positive selection found between *Arabidopsis thaliana* and *Populus trichocarpa*. It was thought that in order to carry out the non-woody to woody transition, non-

synonymous changes across the gene should precede. In comparisons with *Oryza*, three genes showed positive selection – that is, a dN/dS value higher than one (gw1.I.5485.1, grail3.0039027801, and estExt_fgenesh4_pg.C_LG_VIII0293 or their Arabidopsis homologue, AT1G32770.1, AT3G62020.1, and AT2G37040.1, respectively).

As expected, comparison with the woody-plants yielded slow evolution rate or negative selection since the majority of the amino acid changes resulted in synonymous changes rather than non-synonymous.

There are some genes that have high dN/dS ratios in comparison with *Arabidopsis thaliana* and *Oryza Sativa* that do not show up as having high dN/dS ratios when compared with *Pinus taeda*. This result very well aligned with the conclusion that non-woody to woody transition requires positive selection on genes related to wood formation (Table 7). AT1G06650 has been shown to be one of the smallest gene clusters (also known as regulon) that regulate a specific biological process including other genes such as AT1G76020 and AT5G13720. Unfortunately the cluster function has not yet been identified (Mentzen & Wurtele, 2008). The same gene had been shown to be one of the re-watering induced transcripts in another study before (C. E. Wong, et al., 2006). This might explain the relative difference in re-watering or drought tolerance between woody and non-woody plants. AT1G32100 was identified as one of the genes involved in lignan biosynthesis. As it is very well known, it is biosynthesized and deposited in significant amounts in heartwood, which is the wood close to the pith (Nakatsubo, Mizutani, Suzuki, Hattori, & Umezawa, 2008). AT1G75280 has been shown to be related to some important biological responses such as detoxification, stress, and secondary metabolism (Mueller, et al., 2008). AT1G77280 was identified first as one of the members of the protein kinase family along with At1g21590 and At5g63940 with a very strong bootstrap support (Kerk, Bulgrien, Smith, &

Gribskov, 2003). In another study, it was listed as a differentially expressed gene by constitutive expression of a plant transcription factor involved in jasmonate-mediated and salicylate-mediated plant defense (J. Li, Brader, & Palva, 2004). It was also noted that it is regulated by auxins, a family of plant hormones (Goda, et al., 2004). AT2G29130 is commonly known as Lac2, which maintains laccase activity in cell and located in the apoplast and the endomembrane system. It has been well studied that this gene responds to water deprivation. It was also noted that it was down-regulated when the plants were placed horizontally due to mechanical response (Andersson-Gunneras, et al., 2006; Koizumi, et al., 2009). In different study, it was demonstrated that it is co-regulated with the At2g35700 (AtERF38) which is the gene more intensively expressed in siliques and floral stems within tissues that undergo cell wall modifications and lignin biosynthesis. In the same study, it was annotated as a core xylem-specific gene (Lasserre, Jobet, Llauro, & Delseny, 2008; Sibout, et al., 2005). Brown et al. listed this gene with genes that exhibit similar expression patterns to secondary cell wall-specific cellulose synthase genes (IRREGULAR XYLEM) IRX1 and IRX3 (D. M. Brown, et al., 2005). AT2G38080 is commonly known as IRX12 or LAC4 (J. Zhou, et al., 2009), which is shown in several studies to regulate secondary cell wall thickening by affecting the morphology of tracheary elements that are located in specifically immature xylem vessel (D. M. Brown, et al., 2005; Kubo, et al., 2005; Mitsuda, et al., 2005; Sibout, et al., 2005; C. Yang, et al., 2007). Another comparative study illustrated that it is differentially expressed in tension wood compared with normal wood (Andersson-Gunneras, et al., 2006; Koizumi, et al., 2009). AT2G40270 was reported as one of the leucine-rich repeat receptor-like kinases of the protein kinase family (Dievart & Clark, 2003). Later on it was noted that it was one of the genes that are up-regulated and then showed transient response to endogenous cytokinin (Hoth, et al., 2003). In another study, it was presented that it is

specifically regulated by brassinosteroid, which is a plant hormone promoted in stem elongation and cell division (Goda, et al., 2004). AT2G40370, commonly known as LAC5, was first reported to have a role in polymerization of lignin during secondary xylem formation due to being highly expressed in xylem but not in bark (Berthet, et al., 2011; Oh, et al., 2003). Later, it was suggested that it is one of the laccase-like multicopper oxidase (LMCO) genes that is expressed in consistently high levels in all tissues (McCaig, Meagher, & Dean, 2005). It was also noted that LAC5 is ubiquitously expressed and its regulation is affected by copper supply (Abdel-Ghany & Pilon, 2008). AT2G42880 was listed one of the MAPK genes, which is highly conserved and plays important roles in regulating a variety of plant biological process (Hamel, et al., 2006). Unfortunately, their gene function in woody plants has not been studied well. AT3G13310 is a Dnaj protein that is mainly responsible for protein folding. It has been shown that its regulation is dependent on the presence of light and oxygen (Branco-Price, Kawaguchi, Ferreira, & Bailey-Serres, 2005; Ma, et al., 2002). AT5G03760 is commonly known as cellulose synthase-like (CSLA9), which was reported to be mainly expressed in stem, vascular leaf, hypocotyl and root etc (Zhu, Nam, Carpita, Matthysse, & Gelvin, 2003). Although, there is no study focused on AT5G06740, GO analysis suggests that is involved in protein phosphorylation and transition metal ion transport and located in plasma membrane that facilitates carbohydrate binding. According to GO analysis, AT5G38940 is the gene that responds to salt stress located in the cell wall apoplast (Boudart, et al., 2005). It is suggested that it is suppressed by plant hormone abscisic acid (ABA) (Xin, Zhao, & Zheng, 2005). In another study, it was also reported that it is down-regulated after cytokinin benzyladenine treatment (D. J. Lee, et al., 2007). AT5G40020 was reported as one of the genes that responses to mechanical stimulation (Kimbrough, Salinas-Mondragon, Boss, Brown, & Sederoff, 2004). Later, it was listed as a

xylem-specific gene that has a function in plant defense (Ko, Beers, & Han, 2006). It was also noted that it is co-regulated with cellulose synthase genes (CESA4, 7 and 8) (Srinivasasainagendra, Page, Mehta, Coulibaly, & Loraine, 2008). AT5G44030 has been intensively studied and reported to be involved in secondary cell wall biogenesis and located in the plasma membrane of the cell by several researchers (Bosca, et al., 2006; D. M. Brown, et al., 2005; Carroll, et al., 2012; Heyndrickx & Vandepoele, 2012; Kim, et al., 2012; Ko, et al., 2006; Ohashi-Ito, et al., 2005; Roberts & Bushoven, 2007; Sibout, et al., 2005; Somerville, 2006). AT5G44790 has been reported in several studies to have a vital role in the ethylene response pathway via the ATP biosynthetic process (Alonso, Hirayama, Roman, Nourizadeh, & Ecker, 1999; Binder, Rodriguez, & Bleecker, 2010; del Pozo, Cambiazo, & Gonzalez, 2010; Fujimoto, Ohta, Usui, Shinshi, & Ohme-Takagi, 2000; Himelblau & Amasino, 2000; Hirayama, et al., 1999; Y. Wang, et al., 2008; Woeste & Kieber, 2000). AT5G54160 is commonly known as caffeate o-methyl transferase 1 (COMT1) and known to be involved in lignification (Bischoff, et al., 2009; Do, et al., 2007; Quentin, et al., 2009; Rogers, et al., 2005; Sibout, et al., 2005; Vanholme, et al., 2010). It was also reported that it regulates shoot apical meristem and vascular cambium (Groover, et al., 2006). It is also noted that it is regulated by plant pathogen defense responses oligogalacturonides (Ferrari, et al., 2007). This gene is known to be down-regulated when the plant is exposed to ultraviolet B radiation stress (Hectors, Prinsen, De Coen, Jansen, & Guisez, 2007). AT5G60020, commonly known as LAC17, is another gene involved in the lignin biosynthetic process. It has been intensively studied and reported to have laccase activity based on enzyme assay and possibly to contribute to cell wall biosynthesis (Berthet, et al., 2011; Che, Lall, & Howell, 2007; Groover, et al., 2006; Krishnaswamy, Verma, Rahman, & Kav, 2011; McCaig, et al., 2005; Minic, et al., 2009; Mutwil, Obro, Willats, & Persson, 2008; Pourcel, et al.,

2005; Srinivasasainagendra, et al., 2008). Lastly, there is not very much known about AT5G60570. It is reported that it interacts with Apoptosis signal-regulating kinase protein and shows changes in gene expression during pollen germination and tube growth in Arabidopsis (Schumann, Navarro-Quezada, Ullrich, Kuhl, & Quint, 2011; Y. Wang, et al., 2008).

As has been shown above from the standpoint of biological functions, recent studies about those genes confirm that there is a very clear distinction between woody and non-woody plants that characterize the features of either of the groups. For instance, AT1G06650, AT1G32100, AT2G29130, AT2G38080, AT2G40370, AT5G03760, AT5G44030, AT5G54160, and AT5G60020 are possibly the altered genes that most likely evolved later in woody plants not only because they possess woody plant characteristics such as resistance to water deprivation, cell wall thickening and modifications, lignin biosynthesis and so on, but also because they are highly expressed and present in the cell wall of xylem cells. Finally, AT1G75280, AT1G77280, AT2G40270, AT2G42880, AT3G13310, AT5G06740, AT5G44790, and AT5G60570 are some of the genes that are hard to distinguish solely by looking at their annotation. There is nothing known about them except they are protein kinases and some are involved in plant stress responses.

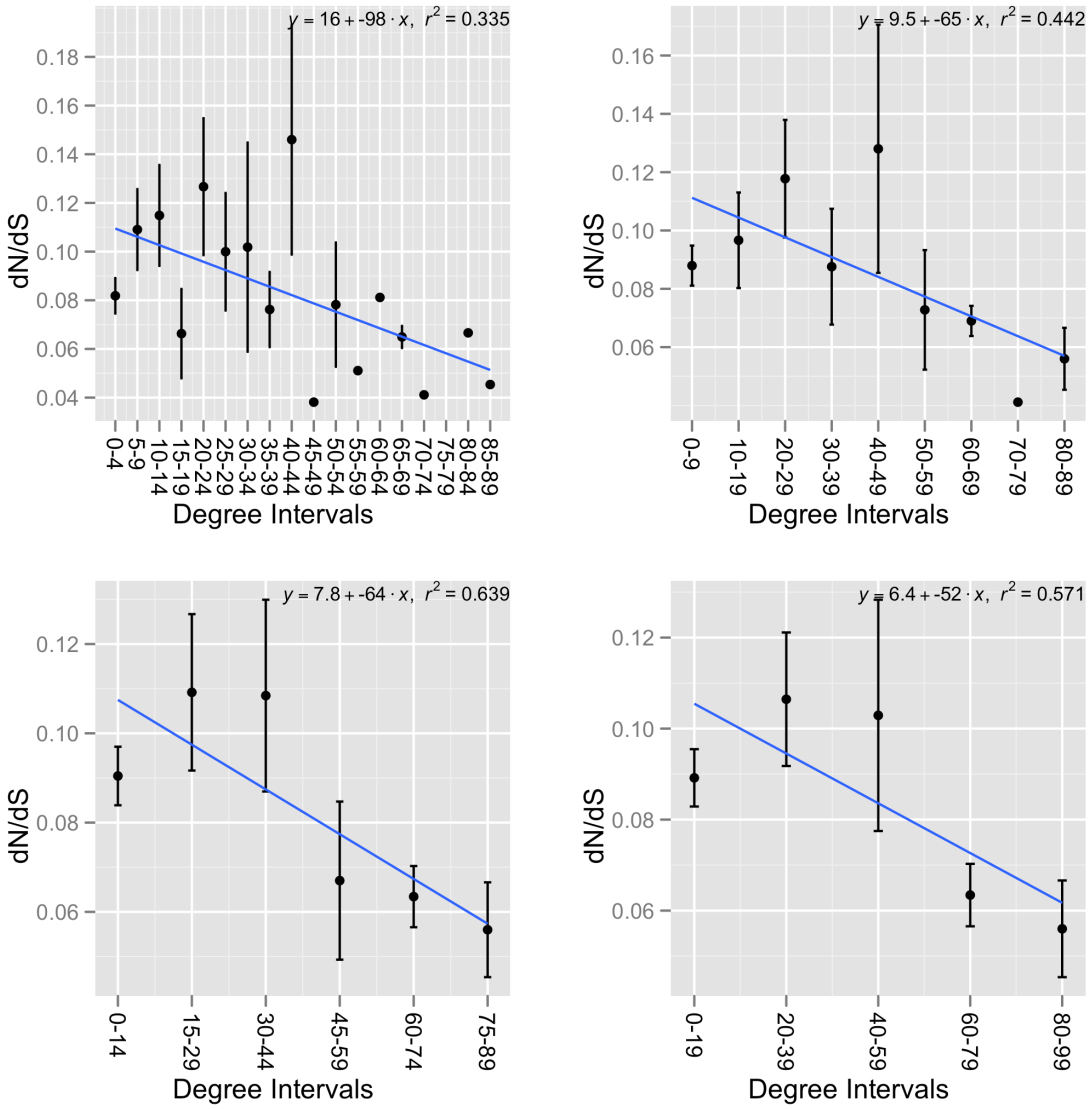



**Figure 9: Relationship between gene connectivity and evolutionary rate in *Arabidopsis thaliana***

**Table 7: Genes that have high dn/ds in comparison with non-woody organisms that do not show up as having high dn/ds when compared with woody organisms**

| Poplar Gene ID | Arabidopsis Thaliana Gene ID | Arabidopsis Thaliana dN/dS | Oryza Sativa dN/dS | Pinus taeda dN/dS | Gene Description |
|---|---|---|---|---|---|
| gw1.X.5490.1 | AT1G06650.2 | 0.10872 | 0.08861 | 0.06362 | encodes a protein whose sequence is similar to 2-oxoglutarate-dependent dioxygenase |
| grail3.0018006801 | AT1G32100.1 | 0.05984 | 0.14672 | 0.039 | Encodes a pinoresinol reductase involved in lignan biosynthesis. Expressed strongly in roots and less strongly in stems. Shows specificity for pinoresinol and not lariciresinol. |
| estExt_fgenesh4_pm.C_LG_II0164 | AT1G75280.1 | 0.11177 | 0.28786 | 0.08049 | isoflavone reductase, putative, identical to SP:P52577 Isoflavone reductase homolog P3 (EC 1.3.1.-) Arabidopsis thaliana; contains Pfam profile PF02716: isoflavone reductase. Involved in response to oxidative stress. |
| estExt_Genewise1_v1.C_LG_V0045 | AT1G77280.1 | 0.18091 | 0.11245 | 0.03591 | Protein kinase protein with adenine nucleotide alpha hydrolases-like domain; |
| fgenesh4_pg.C_LG_IX001228 | AT2G29130.1 | 0.11143 | 0.23534 | 0.01391 | putative laccase, knockout mutant had reduced root elongation under PEG-induced dehydration |
| gw1.I.247.1 | AT2G29130.1 | 0.11652 | 0.28072 | 0.0131 | putative laccase, knockout mutant had reduced root elongation under PEG-induced dehydration |
| estExt_fgenesh4_pg.C_LG_VIII0541 | AT2G38080.1 | 0.06755 | 0.20658 | 0.01588 | LAC4 appears to have laccase activity based on enzyme assays performed using lac4 mutants. |
| estExt_Genewise1_v1.C_LG_XVI3501 | AT2G38080.1 | 0.04538 | 0.30294 | 0.01475 | LAC4 appears to have laccase activity based on enzyme assays performed using lac4 mutants. |
| eugene3.00161066 | AT2G38080.1 | 0.03814 | 0.25122 | 0.02142 | LAC4 appears to have laccase activity based on enzyme assays performed using lac4 mutants. |
| fgenesh4_pg.C_LG_VI000783 | AT2G38080.1 | 0.04135 | 0.1567 | 0.02363 | LAC4 appears to have laccase activity based on enzyme assays performed using lac4 mutants. |
| gw1.I.2205.1 | AT2G40270.1 | 0.18091 | 0.13038 | 0.03929 | Protein kinase family protein; |
| estExt_fgenesh4_pm.C_LG_VIII0291 | AT2G40370.1 | 0.05034 | 0.24218 | 0.01505 | putative laccase, a member of laccase family of genes (17 members in Arabidopsis). |
| estExt_fgenesh4_pm.C_LG_II0282 | AT2G42880.1 | 0.10759 | 0.06738 | 0.03566 | member of MAP Kinase |
| eugene3.00060017 | AT3G13310.1 | 0.08657 | 0.10449 | 0.05887 | Chaperone DnaJ-domain superfamily protein; |
| estExt_fgenesh4_pm.C_LG_VIII0087 | AT5G03760.1 | 0.06165 | 0.16344 | 0.02288 | encodes a beta-mannan synthase that is required for agrobacterium-mediated plant genetic transformation involves a complex interaction between the bacterium and the host plant. |

| | | | | | |
|---|---|---|---|---|---|
| gw1.I.5741.1 | AT5G06740.1 | 0.04072 | 0.07597 | 0.03387 | Concanavalin A-like lectin protein kinase family protein; |
| gw1.6246.4.1 | AT5G38940.1 | 0.06091 | 0.12837 | 0.01297 | RmlC-like cupins superfamily protein; |
| gw1.XVII.817.1 | AT5G40020.1 | 0.07174 | 0.47858 | 0.04202 | Pathogenesis-related thaumatin superfamily protein; |
| eugene3.00002636 | AT5G44030.1 | 0.05595 | 0.04695 | 0.00874 | Encodes a cellulose synthase involved in secondary cell wall biosynthesis. Confers resistance towards bacterial and fungal pathogens, independent of salicylic acid, ethylene and jasmonate signaling. |
| estExt_Genewise1_v1.C_290004 | AT5G44790.1 | 0.10937 | 0.19577 | 0.00991 | ATP dependent copper transporter vital for ethylene response pathway |
| estExt_fgenesh4_pm.C_LG_XII0129 | AT5G54160.1 | 0.07378 | 0.21743 | 0.0516 | A caffeic acid/5-hydroxyferulic acid O-methyltransferase. Interacts with 14-4-3 proteins in yeast 2 hybrid assay. |
| estExt_fgenesh4_pg.C_LG_VI0684 | AT5G60020.1 | 0.09755 | 0.27306 | 0.04262 | LAC17 appears to have laccase activity based on enzyme assays performed using lac17 mutants. Notably, these mutants appear to have a reduced deposition of G lignin units. LAC17 is expressed in interfascicular fibers and likely contributes to lignin biosynthesis, and hence, cell wall biosynthesis, there. |
| fgenesh4_pg.C_LG_XIX000486 | AT5G60570.1 | 0.14694 | 0.1337 | 0.07494 | Galactose oxidase/kelch repeat superfamily protein;. |

### III.4. Conclusion

As mentioned in the introduction, we were able to show in this study that the hub genes evolve more slowly on average. In order to do better analysis, more sequence information is essential. Unfortunately, current sequence information in plants, especially gymnosperms, is not sufficient to provide useful information in bioinformatics. Plant scientist should consider sequencing different tree species and contribute to public databases more in the future. Lab trials towards identifying protein functions should be another essential future study in order to help bioinformaticians to verify their findings and make more generalized and robust conclusions.

**LITERATURE CITED**


Abdel-Ghany, S. E., & Pilon, M. (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. *J Biol Chem, 283*(23), 15932-15945.

Abel, S., Nguyen, M. D., Chow, W., & Theologis, A. (1995). ASC4, A PRIMARY INDOLEACETIC ACID-RESPONSIVE GENE ENCODING 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE IN ARABIDOPSIS-THALIANA - STRUCTURAL CHARACTERIZATION, EXPRESSION IN ESCHERICHIA-COLI, AND EXPRESSION CHARACTERISTICS IN RESPONSE TO AUXIN. *Journal of Biological Chemistry, 270*(32), 19093-19099.

Abel, S., Savchenko, T., & Levy, M. (2005). Genome-wide comparative analysis of the IQD gene families in Arabidopsis thaliana and Oryza sativa. *BMC Evol Biol, 5*, 72.

Aloni, R. (1979). ROLE OF AUXIN AND GIBBERELLIN IN DIFFERENTIATION OF PRIMARY PHLOEM FIBERS. *Plant Physiology, 63*(4), 609-614.

Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., & Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science, 284*(5423), 2148-2152.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.

Andersson-Gunneras, S., Mellerowicz, E. J., Love, J., Segerman, B., Ohmiya, Y., Coutinho, P. M., et al. (2006). Biosynthesis of cellulose-enriched tension wood in Populus: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J, 45*(2), 144-165.

Aragues, R., Sali, A., Bonet, J., Marti-Renom, M., & Oliva, B. (2007). Characterization of protein hubs by inferring interacting motifs from protein interactions. *PLoS Comput Biol, 3*, 1761 - 1771.

Ascencio-Ibanez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R., et al. (2008). Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol, 148*(1), 436-454.

Atanassov, II, Pittman, J. K., & Turner, S. R. (2009). Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of Arabidopsis secondary cell walls. *J Biol Chem, 284*(6), 3833-3841.

Barnett, J. R., & Bonham, V. A. (2004). Cellulose microfibril angle in the cell wall of wood fibres. *Biological Reviews, 79*(2), 461-472.

Batada, N., Hurst, L., & Tyers, M. (2006). Evolutionary and physiological importance of hub proteins. *PLoS Comput Biol, 2*, e88.

Batada, N., Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B., Hurst, L., et al. (2006). Stratus not altocumulus: a new view of the yeast protein interaction network. *PLoS Biol, 4*, e317.

Batada, N., Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B., Hurst, L., et al. (2007). Still stratus not altocumulus: further evidence against the date/party hub distinction. *PLoS Biol, 5*, e154.

Baucher, M., Chabbert, B., Pilate, G., Van Doorsselaere, J., Tollier, M. T., Petit-Conil, M., et al. (1996). Red Xylem and Higher Lignin Extractability by Down-Regulating a Cinnamyl Alcohol Dehydrogenase in Poplar. *Plant Physiol, 112*(4), 1479-1490.

Baucher, M., Van Doorsselaere, J., Gielen, J., Van Montagu, M., Inze, D., & Boerjan, W. (1995). Genomic nucleotide sequence of an Arabidopsis thaliana gene encoding a cinnamyl alcohol dehydrogenase. *Plant Physiol, 107*(1), 285-286.

Becker, J. D., Boavida, L. C., Carneiro, J., Haury, M., & Feijo, J. A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiol, 133*(2), 713-725.

Berthet, S., Demont-Caulet, N., Pollet, B., Bidzinski, P., Cezard, L., Le Bris, P., et al. (2011). Disruption of LACCASE4 and 17 results in tissue-specific alterations to lignification of Arabidopsis thaliana stems. *Plant Cell, 23*(3), 1124-1137.

Bertin, N., Simonis, N., Dupuy, D., Cusick, M., Han, J., Fraser, H., et al. (2007). Confirmation of organized modularity in the yeast interactome. *PLoS Biol, 5*, e153.

Betancur, L., Singh, B., Rapp, R. A., Wendel, J. F., Marks, M. D., Roberts, A. W., et al. (2010). Phylogenetically distinct cellulose synthase genes support secondary wall thickening in arabidopsis shoot trichomes and cotton fiber. *J Integr Plant Biol, 52*(2), 205-220.

Binder, B. M., Rodriguez, F. I., & Bleecker, A. B. (2010). The copper transporter RAN1 is essential for biogenesis of ethylene receptors in Arabidopsis. *J Biol Chem, 285*(48), 37263-37270.

Bischoff, V., Cookson, S. J., Wu, S., & Scheible, W. R. (2009). Thaxtomin A affects CESA-complex density, expression of cell wall genes, cell wall composition, and causes ectopic lignification in Arabidopsis thaliana seedlings. *J Exp Bot, 60*(3), 955-965.

Bloom, J., & Adami, C. (2004). Evolutionary rate depends on number of protein-protein interactions independently of gene expression level: response. *BMC Evol Biol, 4*, 14.

Borner, G. H., Lilley, K. S., Stevens, T. J., & Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. *Plant Physiol, 132*(2), 568-577.

Borner, G. H., Sherrier, D. J., Stevens, T. J., Arkin, I. T., & Dupree, P. (2002). Prediction of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A genomic analysis. *Plant Physiol, 129*(2), 486-499.

Bosca, S., Barton, C. J., Taylor, N. G., Ryden, P., Neumetzler, L., Pauly, M., et al. (2006). Interactions between MUR10/CesA7-dependent secondary cellulose biosynthesis and primary cell wall structure. *Plant Physiol, 142*(4), 1353-1363.

Boudart, G., Jamet, E., Rossignol, M., Lafitte, C., Borderies, G., Jauneau, A., et al. (2005). Cell wall proteins in apoplastic fluids of Arabidopsis thaliana rosettes: identification by mass spectrometry and bioinformatics. *Proteomics, 5*(1), 212-221.

Branco-Price, C., Kawaguchi, R., Ferreira, R. B., & Bailey-Serres, J. (2005). Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Ann Bot, 96*(4), 647-660.

Breen, A. L., Glenn, E., Yeager, A., & Olson, M. S. (2009). Nucleotide diversity among natural populations of a North American poplar (Populus balsamifera, Salicaceae). *New Phytol, 182*(3), 763-773.

Brown, D., Wightman, R., Zhang, Z., Gomez, L. D., Atanassov, I., Bukowski, J.-P., et al. (2011). Arabidopsis genes IRREGULAR XYLEM (IRX15) and IRX15L encode DUF579-

containing proteins that are essential for normal xylan deposition in the secondary cell wall. *The Plant Journal, 66*(3), 401-413.
Brown, D. M., Goubet, F., Wong, V. W., Goodacre, R., Stephens, E., Dupree, P., et al. (2007). Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *Plant J, 52*(6), 1154-1168.
Brown, D. M., Zeef, L. A., Ellis, J., Goodacre, R., & Turner, S. R. (2005). Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell, 17*(8), 2281-2295.
Carroll, A., Mansoori, N., Li, S., Lei, L., Vernhettes, S., Visser, R. G., et al. (2012). Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants. *Plant Physiol, 160*(2), 726-737.
Che, P., Lall, S., & Howell, S. H. (2007). Developmental steps in acquiring competence for shoot development in Arabidopsis tissue culture. *Planta, 226*(5), 1183-1194.
Childs, K. L., Davidson, R. M., & Buell, C. R. (2011). Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS One, 6*(7), e22196.
del Pozo, T., Cambiazo, V., & Gonzalez, M. (2010). Gene expression profiling analysis of copper homeostasis in Arabidopsis thaliana. *Biochem Biophys Res Commun, 393*(2), 248-252.
Dewey, F. E., Perez, M. V., Wheeler, M. T., Watt, C., Spin, J., Langfelder, P., et al. (2011). Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circ Cardiovasc Genet, 4*(1), 26-35.
Dharmawardhana, P., Brunner, A., & Strauss, S. (2010). Genome-wide transcriptome analysis of the transition from primary to secondary stem development in Populus trichocarpa. *BMC Genomics, 11*(1), 150.
Dievart, A., & Clark, S. E. (2003). Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. *Curr Opin Plant Biol, 6*(5), 507-516.
Dinneny, J. R., & Yanofsky, M. F. (2004). Vascular patterning: xylem or phloem? *Curr Biol, 14*(3), R112-114.
Do, C. T., Pollet, B., Thevenin, J., Sibout, R., Denoue, D., Barriere, Y., et al. (2007). Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta, 226*(5), 1117-1129.
Eklund, L., & Little, C. H. A. (1996). Laterally applied Ethrel causes local increases in radial growth and indole-3-acetic acid concentration in Abies balsamea shoots. *Tree Physiology, 16*(5), 509-513.
Ferrari, S., Galletti, R., Denoux, C., De Lorenzo, G., Ausubel, F. M., & Dewdney, J. (2007). Resistance to Botrytis cinerea induced in Arabidopsis by elicitors is independent of salicylic acid, ethylene, or jasmonate signaling but requires PHYTOALEXIN DEFICIENT3. *Plant Physiol, 144*(1), 367-379.
Franke, R., McMichael, C. M., Meyer, K., Shirley, A. M., Cusumano, J. C., & Chapple, C. (2000). Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase. *Plant J, 22*(3), 223-234.
Fraser, H. (2005). Modularity and evolutionary constraint on proteins. *Nat Genet, 37*, 351 - 352.
Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., & Feldman, M. W. (2002). Evolutionary rate in the protein interaction network. *Science, 296*(5568), 750-752.

Frewen, B. E., Chen, T. H., Howe, G. T., Davis, J., Rohde, A., Boerjan, W., et al. (2000). Quantitative trait loci and candidate gene mapping of bud set and bud flush in populus. *Genetics, 154*(2), 837-845.

Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H., & Ohme-Takagi, M. (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell, 12*(3), 393-404.

Fukuda, H., & Komamine, A. (1980). DIRECT EVIDENCE FOR CYTODIFFERENTIATION TO TRACHEARY ELEMENTS WITHOUT INTERVENING MITOSIS IN A CULTURE OF SINGLE CELLS ISOLATED FROM THE MESOPHYLL OF ZINNIA-ELEGANS. *Plant Physiology, 65*(1), 61-64.

Fuller, T. F., Ghazalpour, A., Aten, J. E., Drake, T. A., Lusis, A. J., & Horvath, S. (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome, 18*(6-7), 463-472.

Gibson, L. J. (2012). The hierarchical structure and mechanics of plant materials. *J R Soc Interface, 9*(76), 2749-2766.

Gifford, E. M., & Foster, A. S. (1989). *Morphology and Evolution of Vascular Plants*: W. H. Freeman.

Goda, H., Sawa, S., Asami, T., Fujioka, S., Shimada, Y., & Yoshida, S. (2004). Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in Arabidopsis. *Plant Physiol, 134*(4), 1555-1573.

Goubet, F., Barton, C. J., Mortimer, J. C., Yu, X., Zhang, Z., Miles, G. P., et al. (2009). Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant J, 60*(3), 527-538.

Goué, N., Lesage-Descauses, M.-C., Mellerowicz, E. J., Magel, E., Label, P., & Sundberg, B. (2008). Microgenomic analysis reveals cell type-specific gene expression patterns between ray and fusiform initials within the cambial meristem of Populus. *New Phytologist, 180*(1), 45-56.

Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., et al. (2012). The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res, 40*(Database issue), D26-32.

Groover, A. T., Mansfield, S. D., DiFazio, S. P., Dupper, G., Fontana, J. R., Millar, R., et al. (2006). The Populus homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Mol Biol, 61*(6), 917-932.

Ha, M. A., MacKinnon, I. M., Sturcova, A., Apperley, D. C., McCann, M. C., Turner, S. R., et al. (2002). Structure of cellulose-deficient secondary cell walls from the irx3 mutant of Arabidopsis thaliana. *Phytochemistry, 61*(1), 7-14.

Hamel, L. P., Nicole, M. C., Sritubtim, S., Morency, M. J., Ellis, M., Ehlting, J., et al. (2006). Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci, 11*(4), 192-198.

Hanada, K., Sawada, Y., Kuromori, T., Klausnitzer, R., Saito, K., Toyoda, T., et al. (2011). Functional compensation of primary and secondary metabolites by duplicate genes in Arabidopsis thaliana. *Mol Biol Evol, 28*(1), 377-382.

Hectors, K., Prinsen, E., De Coen, W., Jansen, M. A., & Guisez, Y. (2007). Arabidopsis thaliana plants acclimated to low dose rates of ultraviolet B radiation show specific changes in

morphology and gene expression in the absence of stress symptoms. *New Phytol, 175*(2), 255-270.
Hernandez-Blanco, C., Feng, D. X., Hu, J., Sanchez-Vallet, A., Deslandes, L., Llorente, F., et al. (2007). Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *Plant Cell, 19*(3), 890-903.
Heyndrickx, K. S., & Vandepoele, K. (2012). Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiol, 159*(3), 884-901.
Himelblau, E., & Amasino, R. M. (2000). Delivering copper within plant cells. *Curr Opin Plant Biol, 3*(3), 205-210.
Hirayama, T., Kieber, J. J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., et al. (1999). RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. *Cell, 97*(3), 383-393.
Horvath, S., & Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol, 4*(8), e1000117.
Hoth, S., Ikeda, Y., Morgante, M., Wang, X., Zuo, J., Hanafey, M. K., et al. (2003). Monitoring genome-wide changes in gene expression in response to endogenous cytokinin reveals targets in Arabidopsis thaliana. *FEBS Lett, 554*(3), 373-380.
Hu, W. J., Harding, S. A., Lung, J., Popko, J. L., Ralph, J., Stokke, D. D., et al. (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol, 17*(8), 808-812.
Induri, B. R., Ellis, D. R., Slavov, G. T., Yin, T., Zhang, X., Muchero, W., et al. (2012). Identification of quantitative trait loci and candidate genes for cadmium tolerance in Populus. *Tree Physiol, 32*(5), 626-638.
Jensen, J. K., Kim, H., Cocuron, J. C., Orler, R., Ralph, J., & Wilkerson, C. G. (2011). The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in Arabidopsis. *Plant J, 66*(3), 387-400.
Jiang, Y., Yang, B., Harris, N. S., & Deyholos, M. K. (2007). Comparative proteomic analysis of NaCl stress-responsive proteins in Arabidopsis roots. *J Exp Bot, 58*(13), 3591-3607.
Jones, L., Ennos, A. R., & Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. *Plant J, 26*(2), 205-216.
Jordan, I., Marino-Ramirez, L., Wolf, Y., & Koonin, E. (2004). Conservation and coevolution in the scale-free human gene coexpression network. *Molecular Biology and Evolution, 21*(11), 2058.
Jouanin, L., Goujon, T., de Nadai, V., Martin, M. T., Mila, I., Vallet, C., et al. (2000). Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity. *Plant Physiol, 123*(4), 1363-1374.
Kerk, D., Bulgrien, J., Smith, D. W., & Gribskov, M. (2003). Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiol, 131*(3), 1209-1219.
Kim, W. C., Ko, J. H., Kim, J. Y., Kim, J. M., Bae, H. J., & Han, K. H. (2012). MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in Arabidopsis. *Plant J*.
Kimbrough, J. M., Salinas-Mondragon, R., Boss, W. F., Brown, C. S., & Sederoff, H. W. (2004). The fast and transient transcriptional network of gravity and mechanical stimulation in the Arabidopsis root apex. *Plant Physiol, 136*(1), 2790-2805.

Ko, J. H., Beers, E. P., & Han, K. H. (2006). Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in Arabidopsis thaliana. *Mol Genet Genomics, 276*(6), 517-531.

Ko, J. H., Yang, S. H., Park, A. H., Lerouxel, O., & Han, K. H. (2007). ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana. *Plant J, 50*(6), 1035-1048.

Koizumi, K., Yokoyama, R., & Nishitani, K. (2009). Mechanical load induces upregulation of transcripts for a set of genes implicated in secondary wall formation in the supporting tissue of Arabidopsis thaliana. *J Plant Res, 122*(6), 651-659.

Krishnaswamy, S., Verma, S., Rahman, M. H., & Kav, N. N. (2011). Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in Arabidopsis. *Plant Mol Biol, 75*(1-2), 107-127.

Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genet, 4*(12), e1000304.

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., et al. (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev, 19*(16), 1855-1860.

Lachaud, S., Catesson, A. M., & Bonnemain, J. L. (1999). Structure and functions of the vascular cambium. *C R Acad Sci III, 322*(8), 633-650.

Lacombe, E., Van Doorsselaere, J., Boerjan, W., Boudet, A. M., & Grima-Pettenati, J. (2000). Characterization of cis-elements required for vascular expression of the cinnamoyl CoA reductase gene and for protein-DNA complex formation. *Plant J, 23*(5), 663-676.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res, 40*(Database issue), D1202-1210.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics, 9*(1), 559.

Lapierre, C., Pollet, B., Petit-Conil, M., Toval, G., Romero, J., Pilate, G., et al. (1999). Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiol, 119*(1), 153-164.

Lasserre, E., Jobet, E., Llauro, C., & Delseny, M. (2008). AtERF38 (At2g35700), an AP2/ERF family transcription factor gene from Arabidopsis thaliana, is expressed in specific cell types of roots, stems and seeds that undergo suberization. *Plant Physiol Biochem, 46*(12), 1051-1061.

Lee, C., O'Neill, M. A., Tsumuraya, Y., Darvill, A. G., & Ye, Z. H. (2007). The irregular xylem9 mutant is deficient in xylan xylosyltransferase activity. *Plant Cell Physiol, 48*(11), 1624-1634.

Lee, C., Teng, Q., Huang, W., Zhong, R., & Ye, Z. H. (2009). The F8H glycosyltransferase is a functional paralog of FRA8 involved in glucuronoxylan biosynthesis in Arabidopsis. *Plant Cell Physiol, 50*(4), 812-827.

Lee, C., Zhong, R., Richardson, E. A., Himmelsbach, D. S., McPhail, B. T., & Ye, Z. H. (2007). The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. *Plant Cell Physiol, 48*(12), 1659-1672.

Lee, D. J., Park, J. Y., Ku, S. J., Ha, Y. M., Kim, S., Kim, M. D., et al. (2007). Genome-wide expression profiling of ARABIDOPSIS RESPONSE REGULATOR 7(ARR7) overexpression in cytokinin response. *Mol Genet Genomics, 277*(2), 115-137.

Lercher, M. J., Urrutia, A. O., & Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet, 31*(2), 180-183.

Li, H., Sun, Y., & Zhan, M. (2009). Exploring pathways from gene co-expression to network dynamics. *Methods Mol Biol, 541*, 249-267.

Li, J., Brader, G., & Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell, 16*(2), 319-331.

Li, Q., Min, D., Wang, J. P., Peszlen, I., Horvath, L., Horvath, B., et al. (2011). Down-regulation of glycosyltransferase 8D genes in Populus trichocarpa caused reduced mechanical strength and xylan content in wood. *Tree Physiol, 31*(2), 226-236.

Li, X., Wu, H. X., & Southerton, S. G. (2010). Seasonal reorganization of the xylem transcriptome at different tree ages reveals novel insights into wood formation in Pinus radiata. *New Phytol, 187*(3), 764-776.

Liang, Y. K., Xie, X., Lindsay, S. E., Wang, Y. B., Masle, J., Williamson, L., et al. (2010). Cell wall composition contributes to the control of transpiration efficiency in Arabidopsis thaliana. *Plant J, 64*(4), 679-686.

Lin, W. D., Liao, Y. Y., Yang, T. J., Pan, C. Y., Buckhout, T. J., & Schmidt, W. (2011). Coexpression-based clustering of Arabidopsis root genes predicts functional modules in early phosphate deficiency signaling. *Plant Physiol, 155*(3), 1383-1402.

Little, C. H. A., & Savidge, R. A. (1987). THE ROLE OF PLANT-GROWTH REGULATORS IN FOREST TREE CAMBIAL GROWTH. *Plant Growth Regulation, 6*(1-2), 137-169.

Ma, L., Gao, Y., Qu, L., Chen, Z., Li, J., Zhao, H., et al. (2002). Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in Arabidopsis. *Plant Cell, 14*(10), 2383-2398.

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics, 21*(16), 3448.

Mao, L., Van Hemert, J. L., Dash, S., & Dickerson, J. A. (2009). Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics, 10*, 346.

McCaig, B. C., Meagher, R. B., & Dean, J. F. (2005). Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in Arabidopsis thaliana. *Planta, 221*(5), 619-636.

Mellerowicz, E. J., Baucher, M., Sundberg, B., & Boerjan, W. (2001). Unravelling cell wall formation in the woody dicot stem. *Plant Mol Biol, 47*(1-2), 239-274.

Mentzen, W. I., & Wurtele, E. S. (2008). Regulon organization of Arabidopsis. *BMC Plant Biol, 8*, 99.

Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., Busson, R., et al. (2000). Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. *J Biol Chem, 275*(47), 36899-36909.

Minic, Z., Jamet, E., San-Clemente, H., Pelletier, S., Renou, J. P., Rihouey, C., et al. (2009). Transcriptomic analysis of Arabidopsis developing stems: a close-up on cell wall genes. *BMC Plant Biol, 9*, 6.

Mitsuda, N., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2005). The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell, 17*(11), 2993-3006.

Mochida, K., Uehara-Yamaguchi, Y., Yoshida, T., Sakurai, T., & Shinozaki, K. (2011). Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. *Plant Cell Physiol, 52*(5), 785-803.

Morreel, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., et al. (2006). Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. *Plant J, 47*(2), 224-237.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., & Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol, 9 Suppl 1*, S4.

Movahedi, S., Van de Peer, Y., & Vandepoele, K. (2011). Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in Arabidopsis and rice. *Plant Physiol, 156*(3), 1316-1330.

Mueller, S., Hilbert, B., Dueckershoff, K., Roitsch, T., Krischke, M., Mueller, M. J., et al. (2008). General detoxification and stress responses are mediated by oxidized lipids through TGA transcription factors in Arabidopsis. *Plant Cell, 20*(3), 768-785.

Mutwil, M., Obro, J., Willats, W. G., & Persson, S. (2008). GeneCAT--novel webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res, 36*(Web Server issue), W320-326.

Nakatsubo, T., Mizutani, M., Suzuki, S., Hattori, T., & Umezawa, T. (2008). Characterization of Arabidopsis thaliana pinoresinol reductase, a new type of enzyme involved in lignan biosynthesis. *J Biol Chem, 283*(23), 15550-15557.

Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T., et al. (2004). The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. *Plant Physiol, 135*(2), 1113-1128.

Novaes, E., Kirst, M., Chiang, V., Winter-Sederoff, H., & Sederoff, R. (2010). Lignin and biomass: a negative correlation for wood formation and lignin content in trees. *Plant Physiol, 154*(2), 555-561.

Oh, S., Park, S., & Han, K. H. (2003). Transcriptional regulation of secondary growth in Arabidopsis thaliana. *J Exp Bot, 54*(393), 2709-2722.

Ohashi-Ito, K., Kubo, M., Demura, T., & Fukuda, H. (2005). Class III homeodomain leucine-zipper proteins regulate xylem cell differentiation. *Plant Cell Physiol, 46*(10), 1646-1656.

Oldham, M., Horvath, S., & Geschwind, D. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proceedings of the National Academy of Sciences, 103*(47), 17973.

Pang, K., Cheng, C., Xuan, Z., Sheng, H., & Ma, X. (2010). Understanding protein evolutionary rate by integrating gene co-expression with protein interactions. *BMC Syst Biol, 4*, 179.

Pena, M. J., Zhong, R., Zhou, G. K., Richardson, E. A., O'Neill, M. A., Darvill, A. G., et al. (2007). Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell, 19*(2), 549-563.

Persson, S., Caffall, K. H., Freshour, G., Hilley, M. T., Bauer, S., Poindexter, P., et al. (2007). The Arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and

homogalacturonan, which are essential for secondary cell wall integrity. *Plant Cell, 19*(1), 237-255.
Phillips, R., & Arnott, S. M. (1983). STUDIES ON INDUCED TRACHEARY ELEMENT DIFFERENTIATION IN CULTURED-TISSUES OF TUBERS OF THE JERUSALEM ARTICHOKE, HELIANTHUS-TUBEROSUS. *Histochemical Journal, 15*(5), 427-436.
Plomion, C., Leprovost, G., & Stokes, A. (2001). Wood formation in trees. *Plant Physiol, 127*(4), 1513-1523.
Pourcel, L., Routaboul, J. M., Kerhoas, L., Caboche, M., Lepiniec, L., & Debeaujon, I. (2005). TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. *Plant Cell, 17*(11), 2966-2980.
Prieto, C., Risueno, A., Fontanillo, C., & De las Rivas, J. (2008). Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PLoS One, 3*(12), e3911.
Quentin, M., Allasia, V., Pegard, A., Allais, F., Ducrot, P. H., Favery, B., et al. (2009). Imbalanced lignin biosynthesis promotes the sexual reproduction of homothallic oomycete pathogens. *PLoS Pathog, 5*(1), e1000264.
Rae, A. M., Tricker, P. J., Bunn, S. M., & Taylor, G. (2007). Adaptation of tree growth to elevated CO2: quantitative trait loci for biomass in Populus. *New Phytol, 175*(1), 59-69.
Ranocha, P., Bourgis, F., Ziemak, M. J., Rhodes, D., Gage, D. A., & Hanson, A. D. (2000). Characterization and functional expression of cDNAs encoding methionine-sensitive and -insensitive homocysteine S-methyltransferases from Arabidopsis. *J Biol Chem, 275*(21), 15962-15968.
Ranocha, P., McDougall, G., Hawkins, S., Sterjiades, R., Borderies, G., Stewart, D., et al. (1999). Biochemical characterization, molecular cloning and expression of laccases - a divergent gene family - in poplar. *Eur J Biochem, 259*(1-2), 485-495.
Reverter, A., Hudson, N. J., Wang, Y., Tan, S. H., Barris, W., Byrne, K. A., et al. (2006). A gene coexpression network for bovine skeletal muscle inferred from microarray data. *Physiol Genomics, 28*(1), 76-83.
Roberts, A. W., & Bushoven, J. T. (2007). The cellulose synthase (CESA) gene superfamily of the moss Physcomitrella patens. *Plant Mol Biol, 63*(2), 207-219.
Rogers, L. A., Dubos, C., Surman, C., Willment, J., Cullis, I. F., Mansfield, S. D., et al. (2005). Comparison of lignin deposition in three ectopic lignification mutants. *New Phytol, 168*(1), 123-140.
Ruan, J., Dean, A. K., & Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Syst Biol, 4*, 8.
Samuga, A., & Joshi, C. P. (2002). A new cellulose synthase gene (PtrCesA2) from aspen xylem is orthologous to Arabidopsis AtCesA7 (irx3) gene associated with secondary cell wall synthesis. *Gene, 296*(1-2), 37-44.
Savidge, R. A. (1983). THE ROLE OF PLANT HORMONES IN HIGHER-PLANT CELLULAR-DIFFERENTIATION .2. EXPERIMENTS WITH THE VASCULAR CAMBIUM, AND SCLEREID AND TRACHEID DIFFERENTIATION IN THE PINE, PINUS-CONTORTA. *Histochemical Journal, 15*(5), 447-466.
Sawa, S., Demura, T., Horiguchi, G., Kubo, M., & Fukuda, H. (2005). The ATE genes are responsible for repression of transdifferentiation into xylem cells in Arabidopsis. *Plant Physiol, 137*(1), 141-148.

Schrader, J., Moyle, R., Bhalerao, R., Hertzberg, M., Lundeberg, J., Nilsson, P., et al. (2004). Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome. *Plant J, 40*(2), 173-187.

Schrader, J., Nilsson, J., Mellerowicz, E., Berglund, A., Nilsson, P., Hertzberg, M., et al. (2004). A High-Resolution Transcript Profile across the Wood-Forming Meristem of Poplar Identifies Potential Regulators of Cambial Stem Cell Identity. *Plant Cell, 16*(9), 2278-2292.

Schumann, N., Navarro-Quezada, A., Ullrich, K., Kuhl, C., & Quint, M. (2011). Molecular evolution and selection patterns of plant F-box proteins with C-terminal kelch repeats. *Plant Physiol, 155*(2), 835-850.

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research, 13*(11), 2498.

Shikata, M., Matsuda, Y., Ando, K., Nishii, A., Takemura, M., Yokota, A., et al. (2004). Characterization of Arabidopsis ZIM, a member of a novel plant-specific GATA factor gene family. *J Exp Bot, 55*(397), 631-639.

Shiu, S. H., & Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol, 132*(2), 530-543.

Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L., et al. (2005). CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. *Plant Cell, 17*(7), 2059-2076.

Sjodin, A., Street, N. R., Sandberg, G., Gustafsson, P., & Jansson, S. (2009). The Populus Genome Integrative Explorer (PopGenIE): a new resource for exploring the Populus genome. *New Phytol, 182*(4), 1013-1025.

Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol, 22*, 53-78.

Srinivasasainagendra, V., Page, G. P., Mehta, T., Coulibaly, I., & Loraine, A. E. (2008). CressExpress: a tool for large-scale mining of expression data from Arabidopsis. *Plant Physiol, 147*(3), 1004-1016.

Stajich, J. (2005). pairwise_kaks - script to calculate pairwise Ka,Ks for a set of sequences. from http://www.bioperl.org/wiki/Main_Page

Street, N. R., Skogstrom, O., Sjodin, A., Tucker, J., Rodriguez-Acosta, M., Nilsson, P., et al. (2006). The genetics and genomics of the drought response in Populus. *Plant J, 48*(3), 321-341.

Suh, M. C., Samuels, A. L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., et al. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. *Plant Physiol, 139*(4), 1649-1665.

Sundberg, B., & Little, C. H. A. (1990). TRACHEID PRODUCTION IN RESPONSE TO CHANGES IN THE INTERNAL LEVEL OF INDOLE-3-ACETIC-ACID IN 1-YEAR-OLD SHOOTS OF SCOTS PINE. *Plant Physiology, 94*(4), 1721-1727.

Szyjanowicz, P. M., McKinnon, I., Taylor, N. G., Gardiner, J., Jarvis, M. C., & Turner, S. R. (2004). The irregular xylem 2 mutant is an allele of korrigan that affects the secondary cell wall of Arabidopsis thaliana. *Plant J, 37*(5), 730-740.

Taylor, N. G. (2007). Identification of cellulose synthase AtCesA7 (IRX3) in vivo phosphorylation sites--a potential role in regulating protein degradation. *Plant Mol Biol, 64*(1-2), 161-171.

Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., & Turner, S. R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proc Natl Acad Sci U S A, 100*(3), 1450-1455.

Team, R. D. C. (2010). R: A Language and Environment for Statistical Computing. from http://www.R-project.org

Torkamani, A., Dean, B., Schork, N. J., & Thomas, E. A. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res, 20*(4), 403-412.

Tsai, C. J., Popko, J. L., Mielke, M. R., Hu, W. J., Podila, G. K., & Chiang, V. L. (1998). Suppression of O-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown wood phenotypes. *Plant Physiol, 117*(1), 101-112.

Tuskan, G., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science, 313*(5793), 1596.

Uggla, C., Moritz, T., Sandberg, G., & Sundberg, B. (1996). Auxin as a positional signal in pattern formation in plants. *Proceedings of the National Academy of Sciences of the United States of America, 93*(17), 9282-9286.

Usadel, B., Schluter, U., Molhoj, M., Gipmans, M., Verma, R., Kossmann, J., et al. (2004). Identification and characterization of a UDP-D-glucuronate 4-epimerase in Arabidopsis. *FEBS Lett, 569*(1-3), 327-331.

Vanholme, R., Ralph, J., Akiyama, T., Lu, F., Pazo, J. R., Kim, H., et al. (2010). Engineering traditional monolignols out of lignin by concomitant up-regulation of F5H1 and down-regulation of COMT in Arabidopsis. *Plant J, 64*(6), 885-897.

Villalobos, D. P., Diaz-Moreno, S. M., Said el, S. S., Canas, R. A., Osuna, D., Van Kerckhoven, S. H., et al. (2012). Reprogramming of gene expression during compression wood formation in pine: Coordinated modulation of S-adenosylmethionine, lignin and lignan related genes. *BMC Plant Biol, 12*, 100.

Wang, Q., Little, C. H. A., & Oden, P. C. (1997). Control of longitudinal and cambial growth by gibberellins and indole-3-acetic acid in current-year shoots of Pinus sylvestris. *Tree Physiology, 17*(11), 715-721.

Wang, Y., Zhang, W. Z., Song, L. F., Zou, J. J., Su, Z., & Wu, W. H. (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. *Plant Physiol, 148*(3), 1201-1211.

Woeste, K. E., & Kieber, J. J. (2000). A strong loss-of-function mutation in RAN1 results in constitutive activation of the ethylene response pathway as well as a rosette-lethal phenotype. *Plant Cell, 12*(3), 443-455.

Wong, C. E., Li, Y., Labbe, A., Guevara, D., Nuin, P., Whitty, B., et al. (2006). Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in Thellungiella, a close relative of Arabidopsis. *Plant Physiol, 140*(4), 1437-1450.

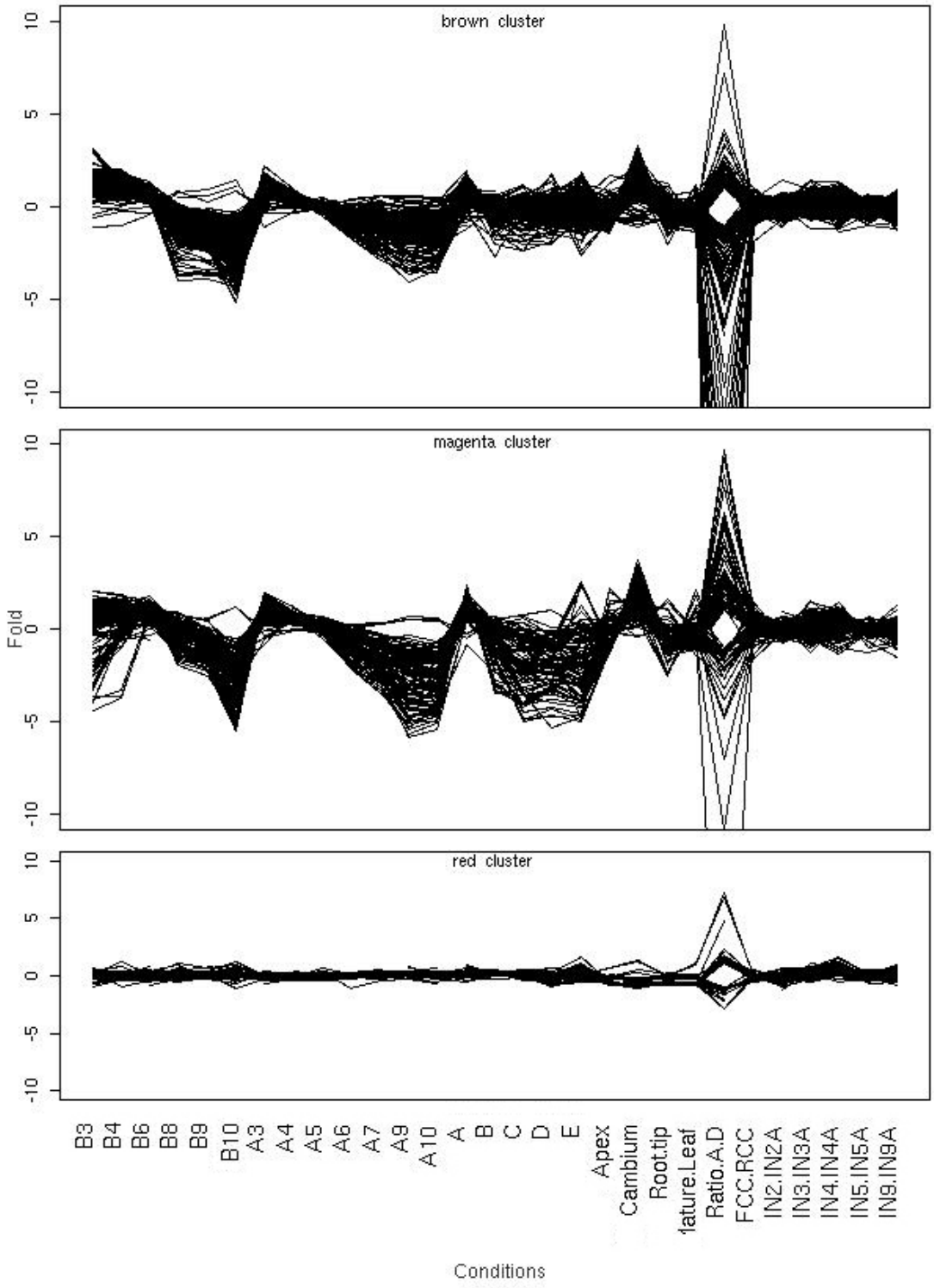
Wong, C. E. C. E., Li, Y. Y., Labbe, A. A., Guevara, D. D., Nuin, P. P., Whitty, B. B., et al. (2006). Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in Thellungiella, a close relative of Arabidopsis. *Plant physiology (Bethesda), 140*(4), 1437.

Woolbright, S. A., Difazio, S. P., Yin, T., Martinsen, G. D., Zhang, X., Allan, G. J., et al. (2008). A dense linkage map of hybrid cottonwood (Populus fremontii x P. angustifolia)

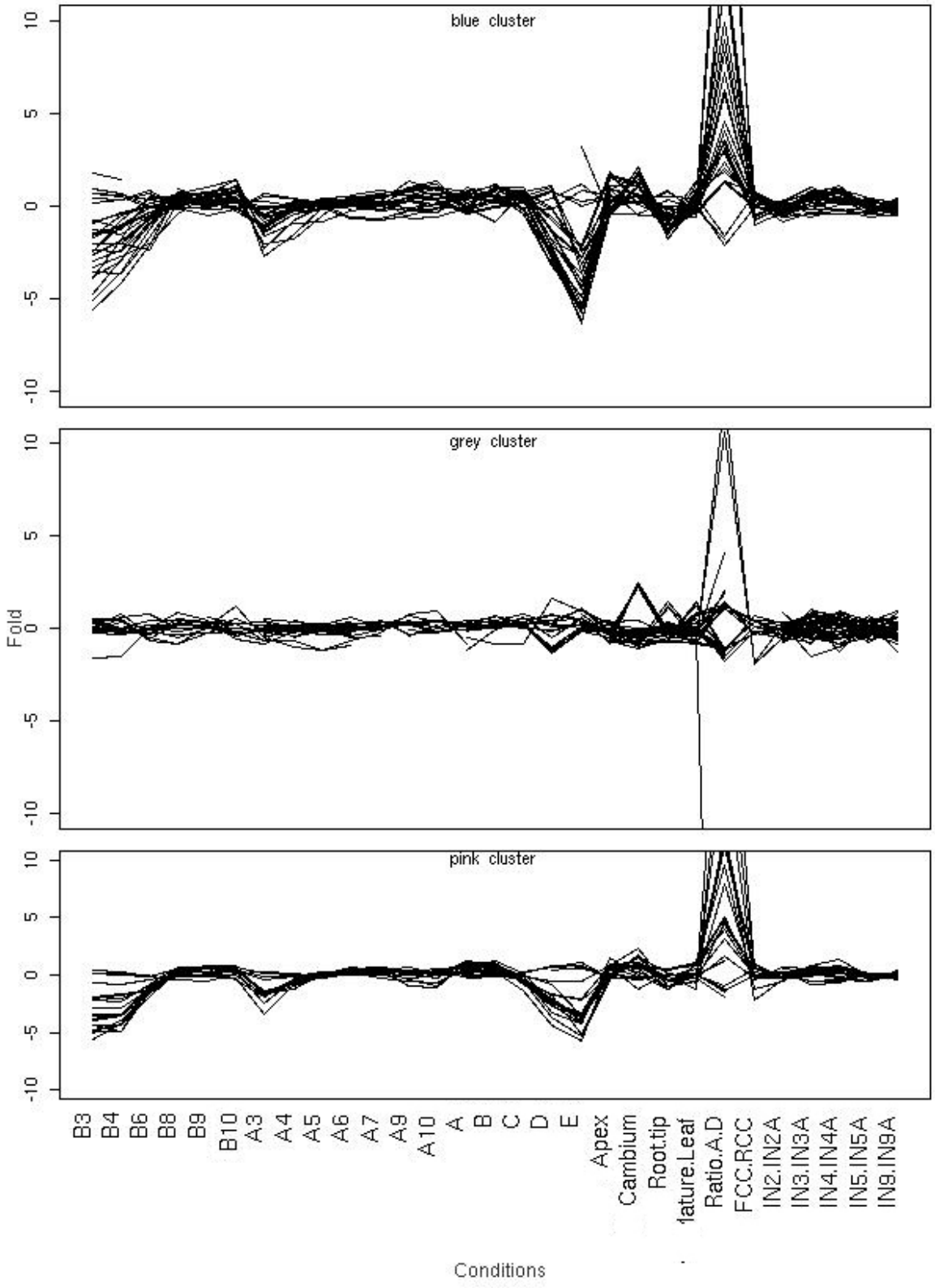
contributes to long-term ecological research and comparison mapping in a model forest tree. *Heredity (Edinb), 100*(1), 59-70.
Wu, A. M., Hornblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., et al. (2010). Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant Physiol, 153*(2), 542-554.
Wu, R., Bradshaw, H., & Stettler, R. (1997). Molecular genetics of growth and development in Populus (Salicaceae). v. mapping quantitative trait loci affecting leaf variation. *Am J Bot, 84*(2), 143.
Wu, R., Ma, C. X., Yang, M. C., Chang, M., Littell, R. C., Santra, U., et al. (2003). Quantitative trait loci for growth trajectories in Populus. *Genet Res, 81*(1), 51-64.
Xiang, Y., Zhang, C. Q., & Huang, K. (2012). Predicting glioblastoma prognosis networks using weighted gene co-expression network analysis on TCGA data. *BMC Bioinformatics, 13 Suppl 2*, S12.
Xin, Z., Zhao, Y., & Zheng, Z. L. (2005). Transcriptome analysis reveals specific modulation of abscisic acid signaling by ROP10 small GTPase in Arabidopsis. *Plant Physiol, 139*(3), 1350-1365.
Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G., & Wilson, Z. A. (2007). Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *Plant Cell, 19*(2), 534-548.
Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*(8), 1586.
Ye, Z. H. (2002). Vascular tissue differentiation and pattern formation in plants. *Annu Rev Plant Biol, 53*, 183-202.
Yin, T., Zhang, X., Huang, M., Wang, M., Zhuge, Q., Tu, S., et al. (2002). Molecular linkage maps of the Populus genome. *Genome, 45*(3), 541-555.
Zhao, C., Craig, J. C., Petzold, H. E., Dickerman, A. W., & Beers, E. P. (2005). The xylem and phloem transcriptomes from secondary tissues of the Arabidopsis root-hypocotyl. *Plant Physiol, 138*(2), 803-818.
Zhao, C., Johnson, B. J., Kositsup, B., & Beers, E. P. (2000). Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. *Plant Physiol, 123*(3), 1185-1196.
Zhong, R., Morrison, W. H., 3rd, Freshour, G. D., Hahn, M. G., & Ye, Z. H. (2003). Expression of a mutant form of cellulose synthase AtCesA7 causes dominant negative effect on cellulose biosynthesis. *Plant Physiol, 132*(2), 786-795.
Zhong, R., Pena, M. J., Zhou, G. K., Nairn, C. J., Wood-Jones, A., Richardson, E. A., et al. (2005). Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. *Plant Cell, 17*(12), 3390-3408.
Zhong, R., Ripperger, A., & Ye, Z. H. (2000). Ectopic deposition of lignin in the pith of stems of two Arabidopsis mutants. *Plant Physiol, 123*(1), 59-70.
Zhou, G. K., Zhong, R., Richardson, E. A., Morrison, W. H., 3rd, Nairn, C. J., Wood-Jones, A., et al. (2006). The poplar glycosyltransferase GT47C is functionally conserved with Arabidopsis Fragile fiber8. *Plant Cell Physiol, 47*(9), 1229-1240.
Zhou, J., Lee, C., Zhong, R., & Ye, Z. H. (2009). MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *Plant Cell, 21*(1), 248-266.

Zhu, Y., Nam, J., Carpita, N. C., Matthysse, A. G., & Gelvin, S. B. (2003). Agrobacterium-mediated root transformation is inhibited by mutation of an Arabidopsis cellulose synthase-like gene. *Plant Physiol, 133*(3), 1000-1010.

## APPENDIX A. Supplementary Tables and Figures




**Figure_Apx 1: Expression pattern of brown, magenta and red clusters**




**Figure_Apx 2: Expression pattern of blue, grey, and pink cluster**

**Table_Apx 1: Genes according to their connectivity**

| Gene Name | Homologue | Cluster | Degree | Annotation |
|---|---|---|---|---|
| estExt_fgenesh4_pg.C_LG_VI1104 | AT1G09690.1 | turquoise | 246 | Translation protein SH3-like family protein; |
| estExt_Genewise1_v1.C_LG_V0746 | AT1G15930.1 | turquoise | 238 | Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein; |
| eugene3.00011799 | AT1G18540.1 | turquoise | 224 | Ribosomal protein L6 family protein; |
| estExt_fgenesh4_pg.C_1220024 | AT1G24070.1 | brown | 237 | encodes a gene similar to cellulose synthase |
| eugene3.01500073 | AT1G43170.8 | turquoise | 273 | Encodes a cytoplasmic ribosomal protein |
| eugene3.00150364 | AT1G48630.1 | turquoise | 235 | Encodes a protein with similarity to mammalian RACKs |
| grail3.0020027101 | AT1G55510.1 | brown | 245 | branched-chain alpha-keto acid decarboxylase E1 beta |
| estExt_Genewise1_v1.C_LG_XV2068 | AT1G67430.1 | turquoise | 270 | Ribosomal protein L22p/L17e family protein; |
| estExt_Genewise1_v1.C_LG_XII1497 | AT1G67430.1 | turquoise | 246 | Ribosomal protein L22p/L17e family protein; |
| estExt_fgenesh4_pg.C_LG_IX0938 | AT1G74050.1 | turquoise | 226 | Ribosomal protein L6 family protein; |
| estExt_fgenesh4_pg.C_LG_VIII0532 | AT1G74270.1 | turquoise | 284 | Ribosomal protein L35Ae family protein; |
| grail3.0019011601 | AT2G27530.2 | turquoise | 231 | Encodes ribosomal protein L10aP |
| estExt_fgenesh4_pm.C_LG_IX0111 | AT2G27710.4 | turquoise | 262 | 60S acidic ribosomal protein family; |
| gw1.I.7425.1 | AT2G32060.2 | turquoise | 273 | Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein; |
| estExt_fgenesh4_kg.C_LG_II0025 | AT2G34480.1 | turquoise | 230 | Ribosomal protein L18ae/LX family protein; |
| eugene3.00100938 | AT2G37870.1 | turquoise | 220 | Bifunctional inhibitor/lipid-transfer protein/seed storage |
| eugene3.00111242 | AT2G41480.1 | turquoise | 238 | Peroxidase superfamily protein; |
| estExt_fgenesh4_pg.C_1650014 | AT2G47610.1 | turquoise | 321 | Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein; |
| gw1.VIII.659.1 | AT3G04920.1 | turquoise | 263 | Ribosomal protein S24e family protein; |
| gw1.XIV.2529.1 | AT3G05560.3 | turquoise | 247 | Ribosomal L22e protein family; |
| estExt_fgenesh4_pg.C_LG_VIII0948 | AT3G06700.3 | turquoise | 290 | Ribosomal L29e protein family; |
| gw1.X.3379.1 | AT3G10610.1 | turquoise | 268 | Ribosomal S17 family protein; |
| grail3.0101010901 | AT3G12490.2 | turquoise | 228 | Encodes a protein with cysteine proteinase inhibitor activity |
| estExt_fgenesh4_pm.C_LG_III0216 | AT3G49910.1 | turquoise | 255 | Translation protein SH3-like family protein; |
| eugene3.00091453 | AT3G52580.1 | turquoise | 262 | Ribosomal protein S11 family protein; |
| estExt_fgenesh4_kg.C_LG_IV0022 | AT3G53020.1 | turquoise | 242 | RPL24B encodes ribosomal protein L24 |
| estExt_fgenesh4_pg.C_LG_III1004 | AT3G53020.1 | turquoise | 229 | RPL24B encodes ribosomal protein L24 |
| grail3.0035003102 | AT3G60770.1 | turquoise | 245 | Ribosomal protein S13/S15; |
| eugene3.00051184 | AT4G09800.1 | turquoise | 230 | encodes a ribosomal protein S18C, |
| eugene3.00020474 | AT4G09800.1 | turquoise | 227 | encodes a ribosomal protein S18C, |
| estExt_Genewise1_v1.C_LG_XI0817 | AT4G10450.1 | turquoise | 262 | Ribosomal protein L6 family; |
| gw1.I.3503.1 | AT4G13170.1 | turquoise | 309 | Ribosomal protein L13 family protein; |
| estExt_fgenesh4_pg.C_LG_XVI0119 | AT4G15000.1 | turquoise | 243 | Ribosomal L27e protein family; |

**Table_Apx 1: Genes according to their connectivity (contd.)**

| Gene Name | Homologue | Cluster | Degree | Annotation |
|---|---|---|---|---|
| eugene3.00060200 | AT4G15000.2 | turquoise | 283 | Ribosomal L27e protein family; |
| grail3.2024000101 | AT4G16720.1 | turquoise | 376 | Ribosomal protein L23/L15e family protein; |
| grail3.0029006102 | AT4G16720.1 | turquoise | 242 | Ribosomal protein L23/L15e family protein; |
| estExt_fgenesh4_pm.C_400015 | AT4G16720.1 | turquoise | 228 | Ribosomal protein L23/L15e family protein; |
| estExt_Genewise1_v1.C_2040056 | AT4G18100.1 | turquoise | 228 | Ribosomal protein L32e; |
| gw1.XVI.3868.1 | AT5G07090.1 | turquoise | 262 | Ribosomal protein S4 (RPS4A) family protein; |
| estExt_fgenesh4_pm.C_LG_V0048 | AT5G10360.2 | turquoise | 270 | RPS6A and RPS6B are fully redundant and essential during gametogenesis |
| estExt_Genewise1_v1.C_660486 | AT5G22440.2 | turquoise | 271 | Ribosomal protein L1p/L10e family; |
| eugene3.00012975 | AT5G23160.1 | turquoise | 281 | unknown protein; |
| estExt_fgenesh4_kg.C_LG_II0054 | AT5G23740.1 | turquoise | 248 | Encodes a putative ribosomal protein S11 (RPS11-beta) |
| estExt_fgenesh4_pm.C_LG_X0057 | AT5G27770.1 | turquoise | 245 | Ribosomal L22e protein family; |
| estExt_fgenesh4_pg.C_1200093 | AT5G39740.2 | turquoise | 249 | Encodes a ribosomal protein RPL5B. |
| estExt_fgenesh4_pg.C_LG_VI1248 | AT5G45775.2 | turquoise | 233 | Ribosomal L5P family protein; |
| estExt_Genewise1_v1.C_LG_IX3421 | AT5G59240.1 | turquoise | 246 | Ribosomal protein S8e family protein; |
| estExt_fgenesh4_kg.C_LG_VI0025 | AT5G60670.1 | turquoise | 286 | Ribosomal protein L11 family protein; |

**Table_Apx 2: GO analysis for turquoise cluster (p < E-10)**

| GO ID | p-value | Description |
|---|---|---|
| 6412 | 3.11E-49 | translation |
| 9059 | 2.05E-46 | macromolecule biosynthetic process |
| 34645 | 6.47E-46 | cellular macromolecule biosynthetic process |
| 10467 | 4.20E-40 | gene expression |
| 44249 | 8.22E-35 | cellular biosynthetic process |
| 9058 | 6.60E-33 | biosynthetic process |
| 44267 | 2.79E-31 | cellular protein metabolic process |
| 19538 | 4.29E-28 | protein metabolic process |
| 44260 | 1.85E-27 | cellular macromolecule metabolic process |
| 43170 | 2.89E-25 | macromolecule metabolic process |
| 44237 | 1.00E-20 | cellular metabolic process |
| 9987 | 2.42E-19 | cellular process |
| 44238 | 7.51E-18 | primary metabolic process |
| 8152 | 1.49E-13 | metabolic process |

**Table_Apx 3: GO analysis for green cluster (p < E-10)**

| GO ID | p-value | Description |
|---|---|---|
| 15979 | 4.11E-35 | photosynthesis |
| 19684 | 6.28E-17 | photosynthesis, light reaction |
| 6091 | 3.57E-12 | generation of precursor metabolites and energy |
| 6778 | 7.35E-10 | porphyrin metabolic process |
| 33013 | 8.58E-10 | tetrapyrrole metabolic process |

**Table_Apx 4: GO analysis for brown cluster (p < E-10)**

| GO ID | p-value | Description |
|---|---|---|
| 9414 | 7.22E-04 | response to water deprivation |
| 9415 | 7.22E-04 | response to water |
| 9628 | 3.79E-02 | response to abiotic stimulus |
| 5985 | 4.07E-02 | sucrose metabolic process |
| 55075 | 4.07E-02 | potassium ion homeostasis |
| 10118 | 4.07E-02 | stomatal movement |
| 30007 | 4.07E-02 | cellular potassium ion homeostasis |
| 50896 | 4.07E-02 | response to stimulus |

**Table_Apx 5: GO analysis for black cluster (p < E-10)**

| GO ID | p-value | Description |
|---|---|---|
| 9834 | 2.02E-17 | secondary cell wall biogenesis |
| 42546 | 8.05E-17 | cell wall biogenesis |
| 70882 | 2.52E-16 | cellular cell wall organization or biogenesis |
| 9832 | 1.64E-15 | plant-type cell wall biogenesis |
| 71669 | 2.26E-12 | plant-type cell wall organization or biogenesis |
| 10382 | 4.30E-11 | cellular cell wall macromolecule metabolic process |
| 71554 | 1.19E-10 | cell wall organization or biogenesis |
| 33692 | 2.71E-10 | cellular polysaccharide biosynthetic process |
| 10417 | 2.71E-10 | glucuronoxylan biosynthetic process |
| 10413 | 2.71E-10 | glucuronoxylan metabolic process |
| 45492 | 2.71E-10 | xylan biosynthetic process |
| 10383 | 3.09E-10 | cell wall polysaccharide metabolic process |
| 271 | 3.42E-10 | polysaccharide biosynthetic process |
| 9698 | 4.78E-10 | phenylpropanoid metabolic process |
| 44264 | 6.91E-10 | cellular polysaccharide metabolic process |

**Table_Apx 6: Genes that are upregulated in xylem (unpublished data)**

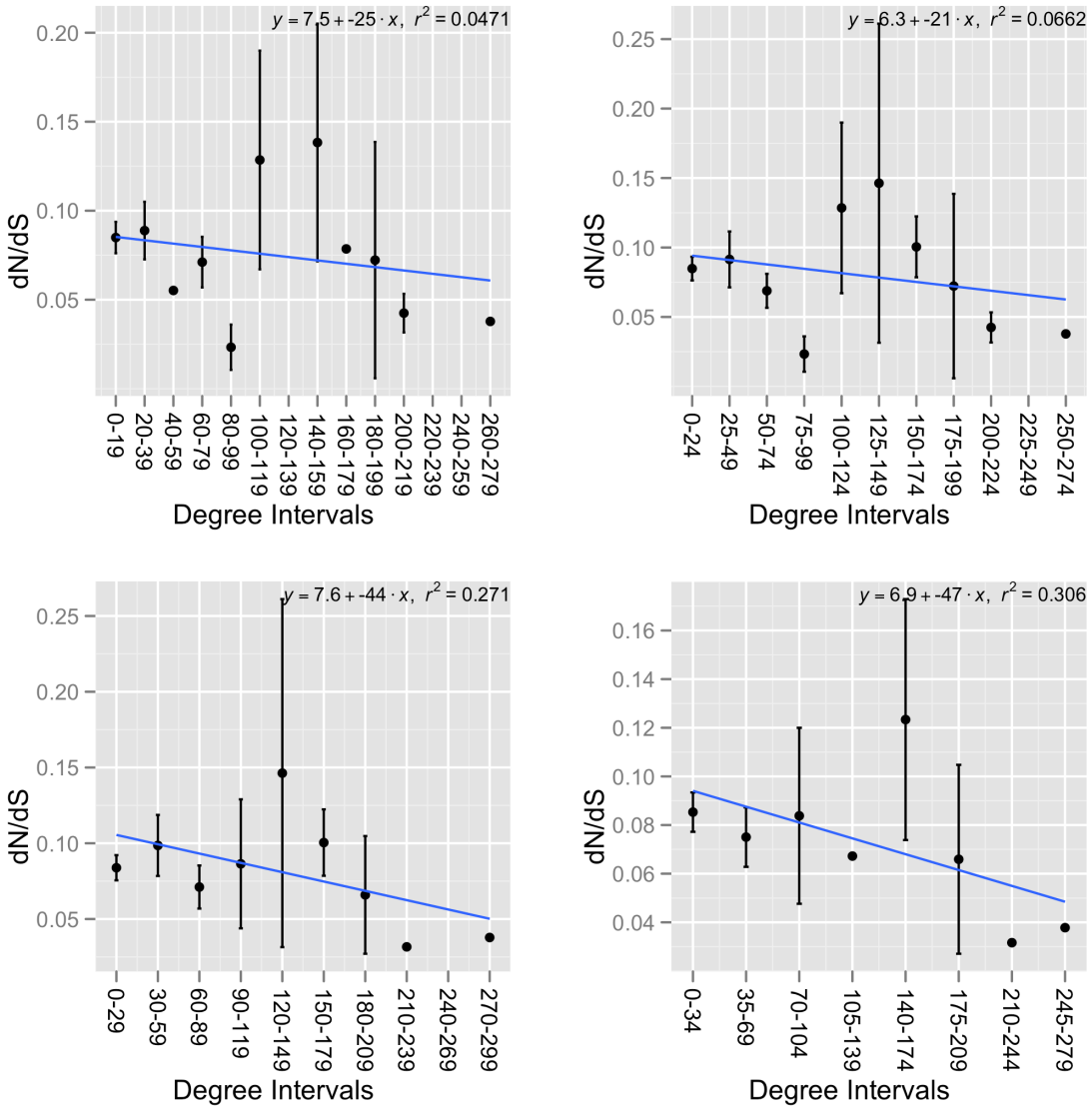
| Gene Name | Homologue | Cluster | Degree | Fold Change | Annotation |
|---|---|---|---|---|---|
| eugene3.10810001 | AT5G03170.1 | purple | 1 | 30.27 | Encodes FLA11, a member of fasciclin-like arabinogalactan proteins (FLAs) containing a cell adhesion fasciclin (FAS) domain. |
| grail3.0194001601 | AT4G16380.1 | black | 36 | 29.24 | Heavy metal transport/detoxification superfamily protein . |
| gw1.86.114.1 | AT1G09610.1 | black | 70 | 28.44 | Protein of unknown function (DUF579). |
| eugene3.00021566 | AT4G22950.1 | black | 48 | 28.25 | MADS-box protein AGL19 |
| eugene3.00100834 | AT3G05990.1 | black | 16 | 26.54 | Leucine-rich repeat (LRR) family protein. |
| eugene3.00050013 | AT3G18660.3 | black | 59 | 26.35 | Plants expressing an RNAi construct specifically targeting PGSIP1 was shown to have a dramatically reduced amount of starch. |
| gw1.XVII.817.1 | AT5G40020.1 | black | 53 | 25.28 | Pathogenesis-related thaumatin superfamily protein. |
| estExt_Genewise1_v1.C_LG_X0543 | AT3G16920.1 | black | 23 | 24.25 | Encodes a chitinase-like protein expressed predominantly in stems. |
| gw1.117.146.1 | AT2G03200.1 | black | 15 | 23.75 | Eukaryotic aspartyl protease family protein. |
| grail3.0008017001 | AT1G27440.1 | black | 37 | 23.26 | GUT2. |
| estExt_Genewise1_v1.C_LG_XIX2288 | AT1G20090.1 | black | 1 | 22.47 | Member of the Rho GTPase family. |
| eugene3.00070393 | AT3G18660.3 | black | 65 | 21.71 | Plants expressing an RNAi construct specifically targeting PGSIP1 was shown to have a dramatically reduced amount of starch. |
| estExt_Genewise1_v1.C_LG_I4958 | AT3G15050.1 | black | 44 | 21.41 | IQ-domain 10 (IQD10). |
| eugene3.00061618 | AT5G11890.1 | black | 12 | 21.26 | FUNCTIONS IN: molecular_function unknown. |
| gw1.IX.4902.1 | AT5G60720.1 | black | 23 | 20.97 | Protein of unknown function, DUF547. |
| gw1.X.6160.1 | AT1G24030.2 | black | 9 | 20.11 | Protein kinase superfamily protein. |
| estExt_Genewise1_v1.C_LG_XVI2679 | AT2G37090.1 | black | 53 | 19.97 | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein. |
| gw1.II.3755.1 | AT2G46710.1 | black | 26 | 19.84 | Rho GTPase activating protein with PAK-box/P21-Rho-binding domain. |
| estExt_fgenesh4_pm.C_LG_VIII0291 | AT2G40370.1 | black | 27 | 19.16 | putative laccase, a member of laccase family of genes (17 members in Arabidopsis). |
| grail3.0025018302 | AT5G01360.2 | black | 81 | 19.03 | Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 domain. |

**Table_Apx 6: Genes that are upregulated in xylem (unpublished data) (contd.)**

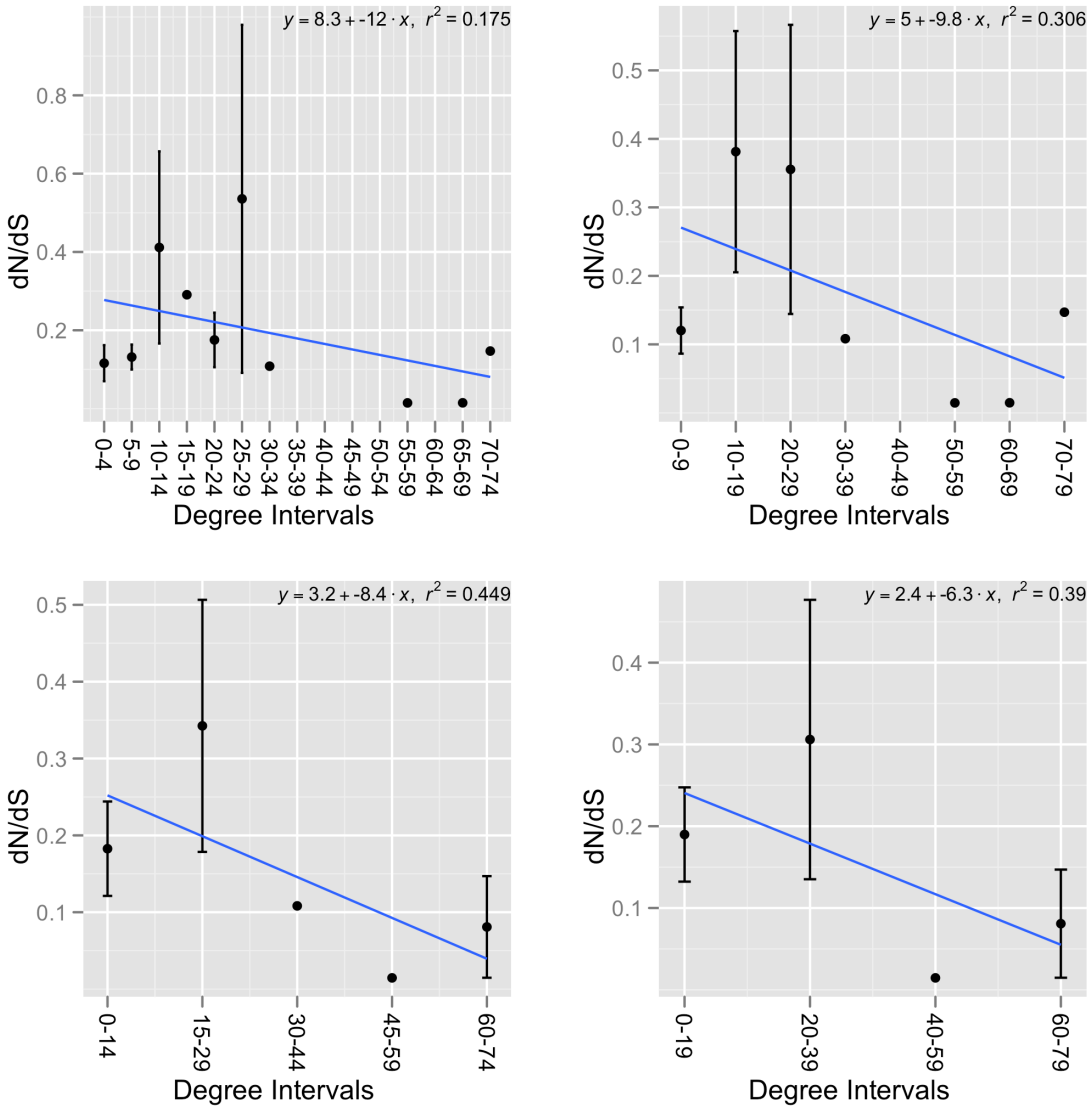
| Gene Name | Homologue | Cluster | Degree | Fold Change | Annotation |
|---|---|---|---|---|---|
| gw1.X.3201.1 | AT4G34200.1 | black | 2 | 18.9 | embryo sac development arrest 9 (EDA9). |
| grail3.0001059501 | AT4G34050.2 | black | 9 | 18.77 | caffeoyl coenzyme A O-methyltransferase 1 (CCoAOMT1). |
| eugene3.00040363 | AT5G17420.1 | black | 39 | 18.38 | Encodes a xylem-specific cellulose synthase that is phosphorylated on one or more serine residues (on either S185 or one of S180 or S181). |
| eugene3.00110990 | AT5G54240.1 | black | 22 | 18.38 | Protein of unknown function (DUF1223). |
| gw1.XI.3218.1 | AT4G18780.1 | black | 50 | 17.88 | Encodes a member of the cellulose synthase family involved in secondary cell wall biosynthesis. |
| gw1.I.9208.1 | AT1G62990.1 | black | 22 | 16.8 | Encodes a homeodomain transcription factor of the Knotted family. |
| gw1.X.5904.1 | AT5G43150.1 | black | 7 | 16.68 | unknown protein. |
| estExt_Genewise1_v1.C_LG_XI2954 | AT5G55970.2 | black | 8 | 16.68 | RING/U-box superfamily protein. |
| estExt_fgenesh4_pg.C_LG_XI1000 | AT4G27430.2 | black | 9 | 16.34 | Positive regulator of light-regulated genes. |
| eugene3.00101062 | AT1G23040.1 | black | 20 | 16.11 | hydroxyproline-rich glycoprotein family protein. |
| estExt_fgenesh4_pg.C_440200 | AT5G37478.1 | black | 41 | 15.78 | TPX2 (targeting protein for Xklp2) protein family. |
| eugene3.00002636 | AT5G44030.1 | black | 69 | 15.67 | Encodes a cellulose synthase involved in secondary cell wall biosynthesis. |
| estExt_fgenesh4_pg.C_LG_VI0684 | AT2G29130.1 | black | 52 | 15.03 | putative laccase, knockout mutant had reduced root elongation under PEG-induced dehydration |
| gw1.VIII.1397.1 | AT2G40320.1 | black | 17 | 15.03 | putative laccase, a member of laccase family of genes (17 members in Arabidopsis). |
| estExt_fgenesh4_pg.C_1870020 | AT5G45970.1 | black | 3 | 15.03 | Encodes a Rac-like protein ARAC2. |
| gw1.XIV.1704.1 | AT3G61750.1 | black | 8 | 14.83 | Cytochrome b561/ferric reductase transmembrane with DOMON related domain. |
| eugene3.00120284 | AT5G53588.1 | black | 1 | 14.83 | Upstream open reading frames (uORFs) are small open reading frames found in the 5' |
| eugene3.00100938 | AT2G37870.1 | turquoise | 220 | 14.22 | LAC4 appears to have laccase activity based on enzyme assays performed using lac4 mutants. |
| fgenesh4_pg.C_LG_VI000783 | AT2G38080.1 | black | 41 | 13.74 | Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. |

**Table_Apx 6: Genes that are upregulated in xylem (unpublished data) (contd.)**

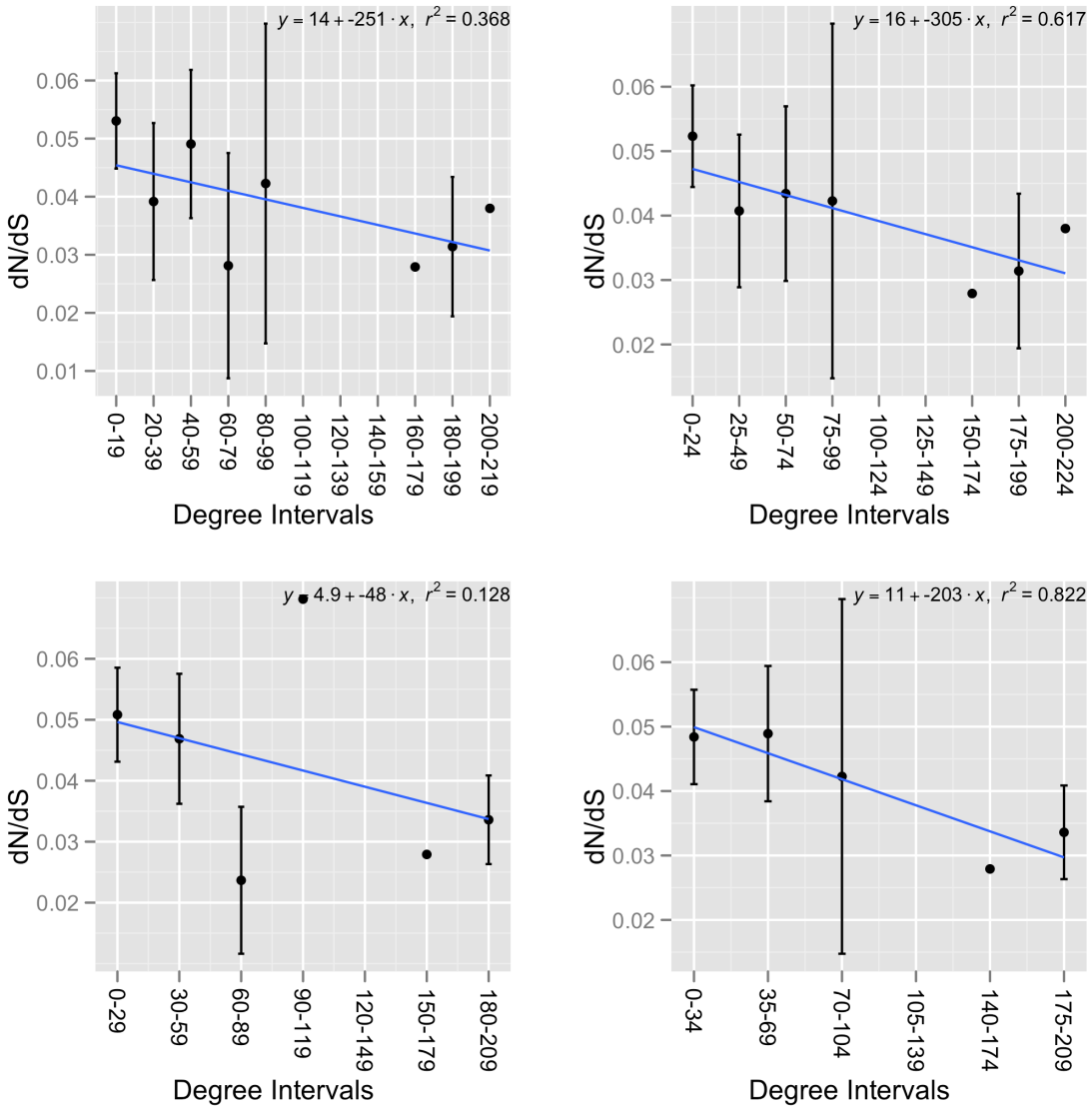
| Gene Name | Homologue | Cluster | Degree | Fold Change | Annotation |
|---|---|---|---|---|---|
| estExt_Genewise1_v1.C_LG_XII1403 | AT4G28690.1 | purple | 1 | 13.74 | BEST Arabidopsis thaliana protein match is: RPM1 interacting protein 13 (TAIR:AT2G20310. |
| estExt_fgenesh4_pm.C_LG_VIII0382 | AT3G29360.2 | black | 20 | 12.47 | Encodes one of four UDP-glucose dehydrogenase UGD) genes. |
| fgenesh4_pg.C_LG_XIV000877 | AT5G16490.1 | black | 18 | 12.47 | encodes a member of a novel protein family that contains contain a CRIB (for Cdc42/Rac-interactive binding) motif required for their specific interaction with GTP-bound Rop1 (plant-specific Rho GTPase). |
| estExt_fgenesh4_pm.C_LG_VI0045 | AT5G47530.1 | black | 1 | 12.3 | Auxin-responsive family protein. |
| fgenesh4_pg.C_LG_IX001228 | AT2G29130.1 | black | 35 | 12.04 | Encodes a homolog of the protein Cas1p known to be involved in polysaccharide O-acetylation in Cryptococcus neoformans. |
| estExt_Genewise1_v1.C_660785 | AT2G28315.1 | black | 13 | 11.96 | Nucleotide/sugar transporter family protein |
| estExt_Genewise1_v1.C_LG_XII0499 | AT5G60490.1 | black | 19 | 11.79 | Encodes a member of fasciclin-like arabinogalactan proteins (FLAs) containing a cell adhesion fasciclin (FAS) domain. |
| estExt_Genewise1_v1.C_LG_I1353 | AT5G59290.2 | black | 15 | 11.47 | Encodes an isoform of UDP-glucuronic acid decarboxylase, which is predicted to be cytosolic by PSORT. |
| grail3.0014024601 | AT1G50890.1 | black | 4 | 11.39 | ARM repeat superfamily protein. |
| gw1.I.247.1 | AT5G60020.1 | black | 33 | 11.39 | LAC17 appears to have laccase activity based on enzyme assays performed using lac17 mutants. |
| grail3.0054000101 | AT2G34410.3 | black | 9 | 11.08 | The IRX9 gene encodes a putative family 43 glycosyl transferase. |
| estExt_fgenesh4_pm.C_LG_I1023 | AT4G34050.1 | black | 4 | 11.08 | caffeoyl coenzyme A O-methyltransferase 1 (CCoAOMT1). |
| grail3.0140003901 | AT3G13275.1 | black | 32 | 11 | unknown protein. |




**Figure_Apx 3: Relationship between gene connectivity and evolutionary rate in *Oryza Sativa***




**Figure_Apx 4: Relationship between gene connectivity and evolutionary rate in *Pinus taeda***




**Figure_Apx 5: Relationship between gene connectivity and evolutionary rate in *Populus tremula***