POPULUS PHENYLPROPANOID METABOLISM ACROSS BIOLOGICAL SCALES: PHYLOGENOMICS, MOLECULAR PHYSIOLOGY, AND ECOSYSTEM MODELING

by

LINDSEY KAY TUOMINEN

(Under the Direction of Chung-Jui Tsai)

ABSTRACT

Phenylpropanoid metabolism is a major contributor to plant biotic and abiotic stress responses while also influencing ecosystem-level processes such as nitrogen cycling. *Populus* in particular is characterized by its rich diversity of phenylpropanoids, varying both qualitatively and quantitatively within and across species. This dissertation develops a cross-scale view of *Populus* secondary metabolism to consider (1) the genetic basis of metabolic diversity within the genus, (2) influences of phenylpropanoid homeostasis on cellular metabolism, and (3) theoretical ecosystem-level effects of altered *Populus* phenylpropanoid levels. I first characterized the *Populus* BAHD acyltransferase family, due to its likely contributions to phenylpropanoid diversity in the taxon, using a phylogenomic approach. The one hundred putative full-length BAHD genes in *Populus* arose through genome-wide and local duplication events, and possibly through retrotransposition. Correlation of phylogenetic and gene expression data suggests that some recent duplicates have undergone functional divergence. To study the cellular-level effects of phenylpropanoid metabolism in *Populus*, I analyzed metabolite and gene expression profiles of cell cultures fed phenylpropanoid enzyme inhibitors and/or the elicitor methyl jasmonate. Results suggest that the effects of enzyme inhibitors manifest primarily at the metabolic level, in

contrast to the transcriptionally-driven changes under methyl jasmonate elicitation. Links between core phenylpropanoid metabolism and phenylpropanoid derivatives, glycosylation, amino acid metabolism, and the Krebs cycle became evident under metabolic perturbation. To consider possible ecological effects of manipulating phenylpropanoid metabolism in *Populus*, I developed ecosystem models for probing indirect effects of *Populus* phenotypes exhibiting increased growth and reduced secondary metabolism. Such phenotypes would be consistent with metabolic engineering goals for trees to be used as biofuels. Initial simulations indicated shifts in biomass of ecosystem components not directly interacting with *Populus* and generated hypotheses for future field research. The results support the potential of ecological modeling as a research and decision-making tool prior to design and field release of novel tree genotypes. This research advances knowledge of *Populus* phenylpropanoid metabolism in a coordinated manner across biological scales and subdisciplines, that should be complementary to more traditional, single-discipline oriented research.

INDEX WORDS: *Populus*, secondary metabolism, phenylpropanoids, phylogenomics, metabolic profiling, transcriptomics, plant cell culture, systems ecology, transgenic risk assessment, indirect effects, cross-scale approach.

POPULUS PHENYLPROPANOID METABOLISM ACROSS BIOLOGICAL SCALES: PHYLOGENOMICS, MOLECULAR PHYSIOLOGY, AND ECOSYSTEM MODELING

by

LINDSEY KAY TUOMINEN

B.A., Macalester College, 2001

M.S., University of Massachusetts Amherst, 2003

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

POPULUS PHENYLPROPANOID METABOLISM ACROSS BIOLOGICAL SCALES: PHYLOGENOMICS, MOLECULAR PHYSIOLOGY, AND ECOSYSTEM MODELING

by

LINDSEY KAY TUOMINEN

Major Professor:	Chung-Jui Tsai
Committee:	Jeffrey F.D. Dean
	Robert O. Teskey
	James H. Leebens-Mack

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2012

DEDICATION

For my parents, who raised me to see that the natural world matters, have allowed me the freedom to choose my own way, and *still* help me dust myself off every time I realize I've dug myself into a hole.

For the Blasiaks, who provided me with more encouragement, inspiration, and advice than any doctoral student should ever require, particularly given the circumstances under which I asked for them. Mary will be greatly missed, and her words will not be forgotten.

For Gwen, who stuck with me tenaciously from the snowiest winters to the stickiest summers. I know you made sacrifices for my ambition. I still hope to help your own dreams become real in some small way, even if my ability to do so is more limited than we imagined.

Last, but not least, for the plants, from whom I have taken much and to whom I have given precious little, except perhaps some alternative pronouns in the places where I can indulge such extravagance. Our species is forever beholden to their kingdom.

ACKNOWLEDGEMENTS

I owe a particular debt of gratitude to my advisor, Dr. Chung-Jui Tsai, who invited me to join her laboratory nine months earlier than I had originally planned to start my doctoral studies. She has been an incredible teacher, providing regular and reliable opportunities to check in, give me feedback, and allow time for questions. She has displayed infinite patience when faced with my own diverse and wandering interests, and she held me back at just the right moment when I was about to walk off a metaphorical cliff, planning to defend a dissertation two-thirds finished. Her generous financial support, supplemented by my own efforts to secure assistantship and travel funding through the Graduate School, helped to sustain my home life through the worst economic downturn in my lifetime.

My dissertation committee members, Drs. Leebens-Mack, Dean, and Teskey have each provided helpful research advice along the way. Dr. Leebens-Mack's comments on an earlier draft of the work presented in Chapter 2 were extremely beneficial and gave me some much-needed confidence that the work as it stood was reasonably competent from the perspective of a molecular phylogeneticist. Dr. Teskey acted as a Capstone Reader for the version of Chapter 4 originally turned in for the Environmental Ethics Graduate Certificate Program, and he has provided me with a wealth of advice for future work to grow that seed into something sturdy enough to withstand the winds of peer review. Dr. Dean provided an array of ideas for new experiments to build on the work in Chapter 3, many of which I failed to fully appreciate until months afterwards. He can rest assured that I'll be tapping his shoulder for more detailed

editorial advice as the work presented in Chapter 3 is advanced to peer-reviewed publication in the near future.

Dr. Scott Harding, the Senior Research Scientist and Co-PI in the Tsai lab, has greatly assisted me by applying his depth of technical expertise and metabolic knowledge both in the laboratory and in the form of substantial editorial advice on the work presented in Chapter 3. Although not officially a member of my advisory committee, he has easily provided me enough help to qualify as an honorary member. I have also benefitted from the advice, collaboration, and expertise of our Assistant Research Scientist, Dr. Chris Frost, and of our postdoctoral researchers, most prominently Drs. Hong-Qiang (Horace) Wang, Batbayar (Baggi) Nyamdari, Ben Babst, Keming Luo, and Yinan Yuan. Our lab managers Hwee Chi (Kate) Tay and Vanessa Michelizzi have been impressively helpful in their detailed knowledge of the logistics of the Tsai lab, and their assistance in scheduling and training undergraduate student workers – along with the direct assistance of the undergraduates themselves -- greatly enhanced my workflow. The Tsai lab computational ninja Ed Johnson has provided cheerful assistance in a wide range of scientific and technical areas of my work relating to bioinformatics and computation, particularly in Chapter 2. Finally, my fellow graduate students have always helped me keep moving when the work was difficult. I particularly thank Raja Payyavula for laying the experimental groundwork for the analyses I carried out in Chapter 3 and for assisting with some early RNAi construct development, Mark Wilson for his ongoing work developing the lab's custom metabolic profiling database and heroic attempts to translate chemical and biological technical needs into a computational framework, Lei (Eric) Du for applying the SLIM analysis to the microarray data in Chapter 3, Han-Yi Chen for providing refresher training in HPLC-MS and

introducing me to the peculiarities of data extraction from the Tsai lab instrument, and Michael Bordeaux in the Dean lab for a healthy sense of perspective whenever I was mired in frustration.

In a "second life" I have lived outside of the Tsai lab, Dr. Bernie Patten introduced me to the fascinating world of systems ecology. He has been an additional source of inspiration to me throughout my time at UGA, finally confirming for me a suspicion I have long held about the importance of mathematics for biology as a whole and for ecology in particular. Dr. Patten was a Capstone Reader for the work presented in Chapter 4, but his contributions to my thinking in Chapter 3 should not be underestimated. My understanding of metabolic pathways was deepened substantially by my interactions with him and the rest of the faculty, postdoctoral researchers, and graduate students in the Systems Ecology and Engineering group.

My doctoral work actually started well over a year before my arrival at UGA, so I also wish to acknowledge the help, inspiration, and advice I received from my committee members at Michigan Technological University prior to my move to Georgia with the Tsai lab in May, 2008. Along with my advisor, Drs. Jackie Grant, Don Leuking, and the late Dave Karnosky ran me through what turned out to be a trial run for my comprehensive exams at UGA. I consider them equal to my committee members at UGA regarding their role in judging my worthiness to advance within the doctoral program, even though the later stages were entirely in the hands of Drs. Tsai, Dean, Teskey, and Leebens-Mack. Other faculty at MTU, as well as the postdocs, graduate students, and lab technicians there, were uniformly welcoming, helpful, and inspirational, making my decision to move to UGA an incredibly difficult one.

TABLE OF CONTENTS

CHAPTER 1.

INTRODUCTION AND LITERATURE REVIEW

Literature Review

Phenylpropanoids are among the most extensively studied secondary metabolites in nature. They are broadly classified into various subclasses encompassing both structural (e.g., lignin) and non-structural (e.g., flavonoids, hydroxycinnamate derivatives and various phenolic conjugates) components (Douglas 1996; Yamamoto et al. 1989). Phenylpropanoids are important in plant development and ecological physiology in that they provide structural support (reviewed by Douglas 1996), function as chemical defense against generalist herbivores (e.g., Coley 1986; Osier et al. 2000; Tahvanainen et al. 1985) and pathogens (e.g., Maher et al. 1994; Miranda et al. 2007), act as protective agents against abiotic stressors (e.g., Babu et al. 2003; Rittinger et al. 1987), and attract pollinators (e.g., Saito and Harborne 1992; Tan et al. 2006). Some compounds in this class of natural products are also among the oldest known herbal medicines. The usefulness of *Salix spp*. (willow) bark for reducing pain and fevers was independently discovered by several cultures. Research to isolate the active component led the German scientist Johann Andreas Buchner to the identification of salicin, and the eventual invention, by a French chemist Charles Frederic Gerhardt, of aspirin (acetylsalicylic acid) - one of the most widely-used pharmaceuticals on the planet (Raskin 1992).

Phenylpropanoids are biosynthetically derived from the amino acid phenylalanine, and as such, the phenylpropanoid pathway interconnects primary and secondary metabolism. From a basic research standpoint, an improved understanding of phenylpropanoid metabolism could help us decipher the biological basis for growth-defense tradeoffs. This is of particular relevance for members of the Salicaceae, which accumulate large amounts of phenylpropanoids while retaining relatively rapid growth compared to other tree species (Julkunen-Tiitto 1986; Lindroth and Hwang 1996a). Such work could also help us understand the consequences of carbon partitioning among various phenylpropanoid pools, or between phenylpropanoids and carbohydrates, in response to developmental or environmental cues. From a practical standpoint, the ability to manipulate phenylpropanoids in economically important taxa such as *Populus spp*. could lead to the generation of trees that are easier to delignify for pulping or saccharify for biofuels (reviewed by Chen et al. 2001; Chen and Dixon 2007), more tolerant to biotic and abiotic stress (e.g., Bhuiyan et al. 2009; Solecka and Kacperska 2003), or better able to perform ecosystem services such as carbon sequestration (e.g., Druart et al. 2006). It may also become possible to engineer novel, synthetic phenylpropanoids for medicinal use or to introduce a taxon-specific portion of the pathway into non-salicaceous plants. With genomic tools now available for investigating the molecular genetics of *Populus spp*., these basic and applied research goals are within reach.

Previous work in the Tsai lab has laid a strong foundation of research on the molecular genetics of phenylpropanoid metabolism in *Populus spp.* (Harding et al. 2002; Kao et al. 2002). The availability of the *Populus* genome (Tuskan et al. 2006) has

allowed lab members to transition from a more conventional gene-by-gene research perspective towards a more holistic approach by surveying the genome for genes with homology to gene families of known function in other species (Oakley et al. 2007; Payyavula et al. 2011; Rajinikanth et al. 2007; Tsai et al. 2006). This genome-enabled approach can provide information about the evolutionary history of phenylpropanoid biosynthetic gene families, especially when employed in conjunction with comparative analysis across multiple taxa. As an example, the initial phylogenetic analysis of the *Populus* hydroxycinnamoyltransfersase (HCT) genes was conducted as part of the annotation of genes involved in *Populus* phenylpropanoid metabolism (Tsai et al. 2006). HCTs are involved in biosynthesis of lignin (e.g., Wagner et al. 2007) and chlorogenic acid (e.g., Comino et al. 2009), both of which are abundant in *Populus.* Interestingly, HCTs also belong to a much larger gene family, the BAHD acyltransferases, many members of which are known to be involved in acylation of various flavonoids, benzenoids, and hydroxycinnamates (e.g., Boatright et al. 2004; Luo et al. 2007; Sullivan 2009). The BAHD superfamily therefore participates in phenylpropanoid metabolism at multiple points and plays a notable role in chemical modifications of these metabolites, thereby contributing to the phenylpropanoid diversity within and among *Populus* species (Greenaway et al. 1991; Greenaway et al. 1992; Tsai et al. 2006). However, genes involved in the chemical elaboration of phenylpropanoids were often overlooked in studies focusing more closely on the core pathways leading to lignin and/or condensed tannin biosynthesis (Hamberger et al. 2007; Tsai et al. 2006). We therefore saw the phylogenomic characterization of the BAHD acyltransferase superfamily in *Populus* as a natural opportunity to fill a gap in our understanding of phenylpropanoid metabolism. In

conjunction with this effort, a functional genomics approach was undertaken to identify additional candidate genes from the BAHD family for functional characterization based on phenylpropanoid coregulation.

A related line of research in the lab has explored the use of *Populus* cell suspension cultures as a tool for understanding phenylpropanoid metabolism. Although cell cultures lack the differentiated cell types that characterize intact plants, their metabolism can be easily manipulated through administration of metabolic intermediates (e.g., Mavandad et al. 1990; Payyavula et al. 2009), elicitors (e.g., Farag et al. 2008; Shinde et al. 2009), or metabolic inhibitors (e.g., Holländer-Czytko and Amrhein 1983; Sircar and Mitra 2009), and the resulting perturbations can be assessed in a matter of hours or days. Cell culture systems can therefore serve as tools for investigation of pathway steps, metabolic dynamics, or cross-talk via chemical perturbation, setting the stage for longer-term studies in whole plants. For example, a previous *Populus* cell culture study in the Tsai lab provided evidence for metabolic competition between phenolic glycoside and condensed tannin pathways, via feeding of the phenolic glycoside precursor salicyl alcohol (Payyavula et al. 2009). The findings support earlier reports of a possible tradeoff between the two phenylpropanoid pools (Lindroth and Hwang 1996b; Orians and Fritz 1995).

A follow-up study in the lab has involved feeding metabolic inhibitors in conjunction with an elicitor to *Populus* cell cultures in order to perturb phenylpropanoid metabolism. Metabolic inhibitors for the first three core phenylpropanoid pathway enzymes were employed to provide targeted perturbations to the pathway: α-aminooxy-β-propionic acid (AOPP) has been demonstrated to inhibit the activity of phenylalanine

ammonia lyase (PAL; Amrhein and Gödeke 1977); piperonylic acid (PIP) inhibits the activity of cinnamate-4-hydroxylase (C4H; Schalk et al. 1998); methylenedioxycinnamic acid inhibits the activity of 4-coumarate:CoA ligase (4CL; Funk and Brodelius 1990). These compounds have previously been used in studies elucidating the role of cinnamate levels in regulating flux into the phenylpropanoid pathway (Bolwell et al. 1988; Mavandad et al. 1990; Orr et al. 1993) and clarifying routes of biosynthesis for hydroxybenzoates (Funk and Brodelius 1990; Sircar and Mitra 2009). In contrast, the elicitor methyl jasmonate (MeJA) is known to stimulate a wide variety of defense responses, including increased accumulation of phenylpropanoids and other defense secondary metabolites (e.g., Arnold et al. 2004; Gundlach et al. 1992), allowing for simultaneous transcriptional activation of multiple phenylpropanoid biosynthetic genes (e.g., Pauwels et al. 2008). Thus, the combined use of MeJA with metabolic inhibitors acting at sequential points in the pathway is expected to shed light on carbon allocation between primary and secondary metabolism, and among various phenylpropanoid pools. While some data was collected for preliminary trials to optimize the doses of the fed compounds based on condensed tannin accumulation, a full characterization of the samples at the transcriptomic and metabolomic levels has not been completed. Therefore, the latter became the focus of the second project in this dissertation. Results from such work should be complementary to other ongoing research in the lab investigating phenylpropanoid metabolism using transgenic *Populus*.

Biotechnology holds great promise for bettering human lives on short time scales and is, by and large, regulated in a manner aimed to provide a balance between safety, environmental soundness, and innovation (7 CFR 340, USGPO 2010). However,

unintended ecological and social consequences of genetic manipulation, particularly as the area of land devoted to cultivating genetically modified organisms continues to increase (Marshall 2009), are still reasonable concerns for ecologists and environmental ethicists. Of relevance to the other research projects presented here, *Populus spp.* are considered by some as keystone species, which by definition are expected to have greater impacts on their ecological communities than the majority of species in the community (Soulé et al. 2005). Therefore, novel genotypes in this genus are more likely to display altered 'extended phenotypes' that propagate beyond the individual organism or population (Whitham et al. 2003). In particular, the manipulation of plant secondary metabolism could have impacts on nitrogen cycling, insect pests, and other species, such as detritivores, which depend on *Populus spp.* for survival without directly impacting tree health (reviewed by Schweitzer et al. 2008).

While many scientists consider the regulatory practices currently in place in the U.S. to ensure the basic safety of field releases as inhibitory to advances in research, in extreme cases even limiting the capacity to conduct biosafety field trials (Strauss et al. 2010; Strauss et al. 2009), the current scheme is essentially a deregulatory approach driven by commodity crops that, in practice, tends to focus more on gene flow than on ecological assessments (National Research Council 2008). This is likely due in part to the logistical complexity of comprehensive ecological studies. However, our limited understanding of possible effects of novel genotypes at the ecological level is often indicated as a source of concern by individuals who oppose the use of genetically modified foods or transgenic field releases (Wagner et al. 2001). The limited ability to "re-regulate" or track longer-term ecological patterns involving deregulated plants,

combined with a lack of scientific will to do the latter (Strauss et al. 2009), is more likely to exacerbate the issue with opponents of biotechnology than to mitigate it. In a democratic society, public acceptance of new technologies must ultimately still be earned. Therefore, developing biosafety assessment methods that can help address this major area of public and scientific concern without requiring time-consuming field studies at the outset would be beneficial to biotechnologists.

Biosafety is a growing area of research in which empirical field trials studying gene flow (e.g., Lu et al. 2006; Zapiola et al. 2008), fitness (e.g., Pennington et al. 2010; Stewart et al. 1997; Sundström et al. 2004), and ecological interactions (e.g., Andreote et al. 2009; Lövei and Arpaia 2005) of transgenic organisms are supplemented with modeling studies and tools allowing predictive forecasting (e.g., Andow and Hilbeck 2004; Muir and Howard 2001; Todd et al. 2008). However, models assessing risk on the basis of specific, trait-based engineered goals while incorporating a holistic, systems-based consideration of ecosystem-level effects have not, to my knowledge, previously been attempted. Such an approach would help provide greater assurance to the public that particular transgenic organisms are likely to have low environmental risks even prior to initial field trials. From a scientific perspective, such tools can help ecologists identify field trial targets for meaningful environmental risk assessment and help biotechnologists define ecologically-relevant molecular design criteria for developing novel genotypes.

Thus, the overarching goal of my dissertation research has been to explore *Populus* phenylpropanoid metabolism at multiple biological scales. In particular, I have aimed to (a) establish genome-scale knowledge in *Populus* and four other genome-enabled plant taxa of a gene family previously shown to be involved in the modification

of phenylpropanoids across a range of angiosperms and gymnosperms and (b) assess the transcriptomic and metabolomic impacts of perturbing phenylpropanoid metabolism in *Populus* cell cultures while (c) considering the broader ethical and ecological implications of the use of transgenic trees in forestry applications.

Dissertation Objectives and Structure

The specific objectives of this dissertation are threefold:

- Identify the BAHD acyltransferase genes in the *Populus trichocarpa* genome and in the genomes of four other sequenced angiosperms, and assess evolutionary and functional patterns of *Populus* BAHD acyltransferases in a phylogenomic context in conjunction with transcriptomics analysis.
- Characterize the metabolic and transcriptomic changes in *Populus* cell cultures under chemical perturbations via a defense elicitor known to stimulate phenylpropanoid metabolism and/or inhibitors of three phenylpropanoid core pathway enzymes.
- Develop ethically responsive, trait-based ecosystem modeling as a scientific tool to help address concerns surrounding potential ecological effects of transgenic tree field releases, with a focus on *Populus spp*. metabolically engineered to favor growth over phenylpropanoid metabolism.

Work towards each of these objectives has established new understandings of *Populus* biology relating to phenylpropanoid metabolism, but each objective focuses on a different biological scale: genome (**CHAPTER 2**), undifferentiated cells (**CHAPTER 3**), or ecosystem (**CHAPTER 4**). I attempt to integrate results from the three scales in a brief concluding chapter. This work has also involved laying paths for future research, especially on questions arising out of my research findings at the genome level. Such projects, including their status and any initial results, are addressed in the appendices.

As a final note, Chapter 2 has previously been published in the peer-reviewed journal *BMC Genomics*. The structure of the chapter has been modified from the original manuscript format for consistency of presentation within this dissertation, but all material submitted for publication is presented here. In addition, Chapter 4 was submitted to the Environmental Ethics Certificate Program in partial fulfillment of graduate certificate requirements. Again, all material in the original version is presented here, with a modified format for consistency of presentation.

References

Amrhein N, Gödeke K-H (1977) α-aminooxy-β-phenylpropionic acid -- a potent inhibitor of L-phenylalanine ammonia-lyase *in vitro* and *in vivo*. *Plant Science Letters* 8: 313-317

Andow DA, Hilbeck A (2004) Science-based risk assessment for nontarget effects of transgenic crops. *BioScience* 54: 637-649

Andreote F, Carneiro R, Salles J, Marcon J, Labate C, Azevedo J, Araújo W (2009) Culture-independent assessment of Rhizobiales-related Alphaproteobacteria and the diversity of Methylobacterium in the rhizosphere and rhizoplane of transgenic Eucalyptus. *Microbial Ecology* 57: 82-93

Arnold T, Appel H, Patel V, Stocum E, Kavalier A, Schultz J (2004) Carbohydrate translocation determines the phenolic content of *Populus* foliage: a test of the sink–source model of plant defense. *New Phytologist* 164: 157-164

Babu TS, Akhtar TA, Lampi MA, Tripuranthakam S, Dixon DG, Greenberg BM (2003) Similar stress responses are elicited by copper and ultraviolet radiation in the aquatic plant *Lemna gibba*: Implication of reactive oxygen species as common signals. *Plant and Cell Physiology* 44: 1320-1329

Bhuiyan NH, Selvaraj G, Wei Y, King J (2009) Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. *Journal of Experimental Botany* 60: 509-521

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding *in vivo* benzenoid metabolism in Petunia petal tissue. *Plant Physiology* 135: 1993-2011

Bolwell GP, Mavandad M, Millar DJ, Edwards KJ, Schuch W, Dixon RA (1988) Inhibition of mRNA levels and activities by *trans*-cinnamic acid in elicitor-induced bean cells. *Phytochemistry* 27: 2109-2117

Chen C, Baucher M, Holst Christensen J, Boerjan W (2001) Biotechnology in trees: Towards improved paper pulping by lignin engineering. *Euphytica* 118: 185-195

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology* 25: 759-761

Coley PD (1986) Costs and benefits of defense by tannins in a neotropical tree. *Oecologia* 70: 238-241

Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E (2009) The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. *BMC Plant Biology* 9: 30

Douglas CJ (1996) Phenylpropanoid metabolism and lignin biosynthesis: From weeds to trees. *Trends in Plant Science* 1: 171-178

Druart N, Rodriguez-Buey M, Barron-Gafford G, Sjodin A, Bhalerao R, Hurry V (2006) Molecular targets of elevated [CO_2] in leaves and stems of *Populus deltoides*: Implications for future tree growth and carbon sequestration. *Functional Plant Biology* 33: 121-131

Farag MA, Huhman DV, Dixon RA, Sumner LW (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. *Plant Physiology* 146: 387-402

Funk C, Brodelius PE (1990) Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia* Andr.: II. Effects of precursor feeding and metabolic inhibitors. *Plant Physiology* 94: 95-101

Greenaway W, English S, May J, Whatley FR (1991) Chemotaxonomy of section *Leuce* poplars by GC-MS of bud exudate. *Biochemical Systematics and Ecology* 19: 507-518

Greenaway W, English S, Whatley FR (1992) Relationships of *Populus* x *acuminata* and *Populus* x *generosa* with their parental species examined by gas chromatography - mass spectrometry of bud exudates. *Canadian Journal of Botany* 70: 212-221

Gundlach H, Müller MJ, Kutchan TM, Zenk MH (1992) Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Proceedings of the National Academy of Sciences* 89: 2389-2393

Hamberger B, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ (2007) Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Canadian Journal of Botany* 85: 1182-1201

Harding SA, Leshkevich J, Chiang VL, Tsai C-J (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiology* 128: 428-438

Holländer-Czytko H, Amrhein N (1983) Subcellular compartment of shikimic acid and phenylalanine in buckwheat cell suspension cultures grown in the presence of shikimate pathway inhibitors. *Plant Science Letters* 29: 89-96

Julkunen-Tiitto R (1986) A chemotaxonomic survey of phenolics in leaves of northern Salicaceae species. *Phytochemistry* 25: 663-667

Kao Y-Y, Harding SA, Tsai C-J (2002) Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. *Plant Physiology* 130: 796-807

Lindroth RL, Hwang S-Y (1996a) Clonal variation in foliar chemistry of quaking aspen (*Populus tremuloides* Michx.). *Biochemical Systematics and Ecology* 24: 357-364

Lindroth RL, Hwang S-Y (1996b) Diversity, redundancy, and multiplicity in chemical defense systems of aspen. *Recent Advances in Phytochemistry* 30: 25-54

Lövei GL, Arpaia S (2005) The impact of transgenic plants on natural enemies: A critical review of laboratory studies. *Entomologia Experimentalis et Applicata* 114: 1-14

Lu M-Z, Chen X-L, Hu J-J (2006) Empirical assessment of gene flow from transgenic poplar plantation. *Ninth International Symposium on the Biosafety of Genetically Modified Organisms, September 24-29, 2006*. International Society for Biosafety Research, Jeju Island, Korea, pp 131-136

Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y, Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C (2007) Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *The Plant Journal* 50: 678-695

Maher EA, Bate NJ, Ni W, Elkind Y, Dixon RA, Lamb CJ (1994) Increased disease susceptibility of transgenic tobacco plants with suppressed levels of preformed phenylpropanoid products. *Proceedings of the National Academy of Sciences* 91: 7802-7806

Marshall A (2009) 13.3 million farmers cultivate GM crops. *Nature Biotechnology* 27: 221

Mavandad M, Edwards R, Liang X, Lamb CJ, Dixon RA (1990) Effects of *trans*-cinnamic acid on expression of the bean phenylalanine ammonia-lyase gene family. *Plant Physiology* 94: 671-680

Miranda M, Ralph SG, Mellway R, White R, Heath MC, Bohlmann J, Constabel CP (2007) The transcriptional response of hybrid Poplar (*Populus trichocarpa* x *P. deltoides*) to infection by *Melampsora medusae* leaf rust involves induction of flavonoid pathway genes leading to the accumulation of proanthocyanidins. *Molecular Plant-Microbe Interactions* 20: 816-831

Muir WM, Howard RD (2001) Fitness components and ecological risk of transgenic release: A model using Japanese medaka (*Oryzias latipes*). *The American Naturalist* 158: 1-16

National Research Council (2008) *Genetically Engineered Organisms, Wildlife, and Habitat: A Workshop Summary*. The National Academies Press, Washington, D.C.

Oakley RV, Wang Y-S, Ramakrishna W, Harding SA, Tsai C-J (2007) Differential expansion and expression of α- and β-tubulin gene families in *Populus*. *Plant Physiology* 145: 961-973

Orians CM, Fritz RS (1995) Secondary chemistry of hybrid and parental willows: Phenolic glycosides and condensed tannins in *Salix sericea*, *S. eriocephala*, and their hybrids. *Journal of Chemical Ecology* 21: 1245-1253

Orr JD, Edwards R, Dixon RA (1993) Stress responses in Alfalfa (*Medicago sativa* L.) (XIV. Changes in the levels of phenylpropanoid pathway intermediates in relation to regulation of L-phenylalanine ammonia-lyase in elicitor-treated cell-suspension cultures). *Plant Physiology* 101: 847-856

Osier TL, Hwang S-Y, Lindroth RL (2000) Effects of phytochemical variation in quaking aspen *Populus tremuloides* clones on gypsy moth *Lymantria dispar* performance in the field and laboratory. *Ecological Entomology* 25: 197-207

Pauwels L, Morreel K, De Witte E, Lammertyn F, Van Montagu M, Boerjan W, Inzé D, Goossens A (2008) Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured *Arabidopsis* cells. *Proceedings of the National Academy of Sciences* 105: 1380-1385

Payyavula RS, Babst BA, Nelsen MP, Harding SA, Tsai CJ (2009) Glycosylation-mediated phenylpropanoid partitioning in *Populus tremuloides* cell cultures. *BMC Plant Biology* 9: 151

Payyavula RS, Tay KHC, Tsai C-J, Harding SA (2011) The sucrose transporter family in *Populus*: The importance of a tonoplast PtaSUT4 to biomass and carbon partitioning. *The Plant Journal* 65: 757-770

Pennington KM, Kapuscinski AR, Morton MS, Cooper AM, Miller LM (2010) Full life-cycle assessment of gene flow consistent with fitness differences in transgenic and wild-type Japanese medaka fish (*Oryzias latipes*). *Environmental Biosafety Research* 9: 41-57

Rajinikanth M, Harding SA, Tsai C-J (2007) The glycine decarboxylase complex multienzyme family in *Populus*. *Journal of Experimental Botany* 58: 1761-1770

Raskin I (1992) Role of salicylic acid in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 43: 439-463

Rittinger PA, Biggs AR, Peirson DR (1987) Histochemistry of lignin and suberin deposition in boundary layers formed after wounding in various plant species and organs. *Canadian Journal of Botany* 65: 1886-1892

Saito N, Harborne JB (1992) Correlations between anthocyanin type, pollinator and flower colour in the Labiatae. *Phytochemistry* 31: 3009-3015

Schalk M, Cabello-Hurtado F, Pierrel MA, Atanossova R, Saindrenan P, Werck-Reichhart D (1998) Piperonylic acid, a selective, mechanism-based inactivator of the *trans*-cinnamate 4-hydroxylase: A new tool to control the flux of metabolites in the phenylpropanoid pathway. *Plant Physiology* 118: 209-218

Schweitzer J, Madritch M, Bailey J, LeRoy C, Fischer D, Rehill B, Lindroth R, Hagerman A, Wooley S, Hart S, Whitham T (2008) From genes to ecosystems: The genetic basis of condensed tannins and their role in nutrient regulation in a *Populus* model system. *Ecosystems* 11: 1005-1020

Shinde A, Malpathak N, Fulzele D (2009) Optimized production of isoflavones in cell cultures of *Psoralea corylifolia* L. using elicitation and precursor feeding. *Biotechnology and Bioprocess Engineering* 14: 612-618

Sircar D, Mitra A (2009) Accumulation of *p*-hydroxybenzoic acid in hairy roots of *Daucus carota* 2: Confirming biosynthetic steps through feeding of inhibitors and precursors. *Journal of Plant Physiology* 166: 1370-1380

Solecka D, Kacperska A (2003) Phenylpropanoid deficiency affects the course of plant acclimation to cold. *Physiologia Plantarum* 119: 253-262

Soulé ME, Estes JA, Miller B, Honnold DL (2005) Strongly interacting species: Conservation policy, management, and ethics. *BioScience* 55: 168-176

Stewart CN, Jr., All JN, Raymer PL, Ramachandran S (1997) Increased fitness of transgenic insecticidal rapeseed under insect selection pressure. *Molecular Ecology* 6: 773-779

Strauss SH, Kershen DL, Bouton JH, Redick TP, Tan H, Sedjo RA (2010) Far-reaching deleterious impacts of regulations on research and environmental studies of recombinant DNA-modified perennial biofuel crops in the United States. *BioScience* 60: 729-741

Strauss SH, Schmitt M, Sedjo RA (2009) Forest scientist views of regulatory obstacles to research and development of transgenic forest biotechnology. *Journal of Forestry* 107: 350-357

Sullivan ML (2009) A novel red clover hydroxycinnamoyl transferase has enzymatic activities consistent with a role in phaselic acid (2-*O*-[caffeoyl]-L-malate) biosynthesis. *Plant Physiology*: 1866-1879

Sundström LF, Lôhmus M, Johnsson JI, Devlin RH (2004) Growth hormone transgenic salmon pay for growth potential with increased predation mortality. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271: S350-S352

Tahvanainen J, Helle E, Julkunen-Tiitto R, Lavola A (1985) Phenolic compounds of willow bark as deterrents against feeding by mountain hare. *Oecologia* 65: 319-323

Tan K, Tan L, Nishida R (2006) Floral phenylpropanoid cocktail and architecture of *Bulbophyllum vinaceum* orchid in attracting fruit flies for pollination. *Journal of Chemical Ecology* 32: 2429-2441

Todd JH, Ramankutty P, Barraclough EI, Malone LA (2008) A screening method for prioritizing non-target invertebrates for improved biosafety testing of transgenic crops. *Environmental Biosafety Research* 7: 35-56

Tsai C-J, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytologist* 172: 47-62

Tuskan GA, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604

USGPO (2010) Introduction of Organisms and Products Altered or Produced Through Genetic Engineering Which are Plant Pests or Which There is Reason to Beleive are Plant Pests. 7 CFR 340. U.S. Government Printing Office, pp 458-475

Wagner A, Ralph J, Akiyama T, Flint H, Phillips L, Torr K, Nanayakkara B, Te Kiri L (2007) Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase in *Pinus radiata*. *Proceedings of the National Academy of Sciences* 104: 11856-11861

Wagner W, Kronberger N, Gaskell G, Allansdottir A, Allum N, de Cheveigne S, Dahinden U, Diego C, Montali L, Mortensen AT, Pfenning U, Rusanen T, Seger N (2001) Nature in disorder: The troubled public of biotechnology. In: Gaskell G, Bauer MW (eds) *Biotechnology 1996-2000: The Years of Controversy*. NMSI Trading Ltd, London

Whitham TG, Young WP, Martinsen GD, Gehring CA, Schweitzer JA, Shuster SM, Wimp GM, Fischer DG, Bailey JK, Lindroth RL, Woolbright S, Kuske CR (2003) Community and ecosystem genetics: A consequece of the extended phenotype. *Ecology* 84: 559-573

Yamamoto E, Bokelman Gordon H, Lewis Norman G (1989) Phenylpropanoid metabolism in cell walls. *Plant Cell Wall Polymers*. American Chemical Society, pp 68-88

Zapiola ML, Campbell CK, Butler MD, Mallory-Smith CA (2008) Escape and establishment of transgenic glyphosate-resistant creeping bentgrass *Agrostis stolonifera* in Oregon, USA: a 4-year study. *Journal of Applied Ecology* 45: 486-494

CHAPTER 2.

DIFFERENTIAL PHYLOGENETIC EXPANSIONS IN BAHD ACYLTRANSFERASES ACROSS FIVE ANGIOSPERM TAXA AND EVIDENCE OF DIVERGENT EXPRESSION AMONG *POPULUS* PARALOGUES[1]

[1]Tuominen LK, Johnson VE, Tsai CJ (2011) *BMC Genomics* 12: 286. Reprinted here with permission of publisher.

Abstract

BAHD acyltransferases are involved in the synthesis and elaboration of a wide variety of secondary metabolites. Previous research has shown that characterized proteins from this family fall broadly into five major clades and contain two conserved protein motifs. Here, we aimed to expand the understanding of BAHD acyltransferase diversity in plants through genome-wide analysis across five angiosperm taxa. We focus particularly on *Populus*, a woody perennial known to produce an abundance of secondary metabolites. Phylogenetic analysis of putative BAHD acyltransferase sequences from *Arabidopsis*, *Medicago*, *Oryza*, *Populus*, and *Vitis*, along with previously characterized proteins, supported a refined grouping of eight major clades for this family. Taxon-specific clustering of many BAHD family members appears pervasive in angiosperms. We identified two new multi-clade motifs and numerous clade-specific motifs, several of which have been implicated in BAHD function by previous structural and mutagenesis research. Gene duplication and expression data for *Populus*-dominated subclades revealed that several paralogous BAHD members in this genus might have already undergone functional divergence. Differential, taxon-specific BAHD family expansion via gene duplication could be an evolutionary process contributing to metabolic diversity across plant taxa. Gene expression divergence among some *Populus* paralogues highlights possible distinctions between their biochemical and physiological functions. The newly discovered motifs, especially the clade-specific motifs, should facilitate future functional study of substrate and donor specificity among BAHD enzymes.

Background

BAHD acyltransferases make up a large family of enzymes responsible for acyl-CoA dependent acylation of secondary metabolites, typically resulting in the formation of esters and amides. In a foundational paper, St. Pierre and De Luca (2000) named the family after the first four characterized members (**B**EAT or benzylalcohol *O*-acetyltransferase from *Clarkia breweri*; **A**HCTs or anthocyanin *O*-hydroxycinnamoyltransferases from *Petunia*, *Senecio*, *Gentiana*, *Perilla*, and *Lavandula*; **H**CBT or anthranilate *N*-hydroxycinnamoyl/benzoyltransferase from *Dianthus caryophyllus*; **D**AT or deacetylvindoline 4-*O*-acetyltransferase from *Catharanthus roseus*). Currently, the BAHD family encompasses over sixty biochemically characterized members in plant taxa ranging from gymnosperms to monocots to legumes. Previous work has shown that these enzymes may be involved in synthesis or modification of such diverse metabolites as alkaloids, terpenoids and phenolics, with ecophysiological roles in minimizing cuticular water loss, defending against herbivory, and attracting pollinators (reviewed in D'Auria 2006).

The BAHD family has been previously organized into five major phylogenetic clades, using 46 biochemically or genetically characterized members (D'Auria 2006). This classification revealed both clade-specific and clade-independent biochemical activities among family members. For example, benzoyl-CoA donor utilization so far appears to be limited to Clade V, while hydroxycinnamoyl-CoA has been reported as a donor for members in multiple clades (D'Auria 2006). Substrate specificity typically varies among clades, and sometimes within clade as well. For example, Clade I members act mainly upon flavonoids, while Clade V members utilize substrates ranging from

terpenoids to medium-chain alcohols to quinic acid, in association with major phylogenetic branches within this clade (D'Auria 2006). Similar diversity of function was also noted for Clade III members, which are involved in formation of alkaloids, esters, and flavonoids, but functional association was less clear due to the smaller size of subclades in this branch. This highlights both the diversity of the BAHD family and the potential challenge of phylogeny-based functional inference with limited sequence and/or species representation.

Most functionally characterized BAHD acyltransferases share two conserved motifs, HXXXD and DFGWG (D'Auria 2006). The conservation of these motifs has facilitated *in silico* identification of BAHD acyltransferases from available genome sequences (Luo et al. 2007; Yu et al. 2009). The HXXXD motif is also found in other thioester CoA-utilizing acyltransferase families (St. Pierre and De Luca 2000) and is absolutely conserved among BAHD acyltransferases. Its importance for catalysis was first established by site-directed mutagenesis (Bayer et al. 2004; Suzuki et al. 2003). Crystallographic analysis of the chrysanthemum (*Dendranthema* x *morifolium*) malonyltransferase Dm3MaT3 provided the structural basis for the catalytic role of the His residue in malonyl-CoA binding (Unno et al. 2007). The importance of the DFGWG motif, which is highly but not absolutely conserved, for enzyme activity was first shown in a *Salvia* malonyltransferase (Suzuki et al. 2003) and a *Rauvolfia* vinorine synthase (Bayer et al. 2004) based on mutagenesis studies of the Asp residue. However, structural analysis of Dm3MaT3 suggested that this Asp residue most likely plays a structural, rather than catalytic, role in enzyme function (Unno et al. 2007). Coupling the structural analysis with mutagenesis studies of two other malonyltransferases from the same species

also revealed a greater structural diversity of acyl acceptor binding sites relative to the acyl-CoA donor binding sites (Unno et al. 2007). This is consistent with the known broad range of acceptor molecules and relatively narrow range of acyl-donors utilized by different BAHD acyltransferases (D'Auria 2006).

Despite the prevalence of BAHD acyltransferases in plants, cross-genome analysis of this family is lacking. Genome-wide analyses of this family have recently been reported for *Arabidopsis* and *Populus* (Luo et al. 2007; Yu et al. 2009), but only in a single-taxon context. We sought to explore BAHD acyltransferase diversity from an evolutionary perspective, with a primary focus in *Populus* due to its ability to synthesize a broad array of secondary metabolites. The most abundant of these metabolites are the phenylpropanoid-derived non-structural phenolics known to play significant roles in biotic and abiotic stress responses in this genus (Constabel and Lindroth 2010; Tsai et al. 2006). The diversity of *Populus* phenylpropanoids (e.g., hydroxycinnamate derivatives, flavonoids, condensed tannins and salicylate-containing phenolic glycosides) can be attributed in large part to side-chain modifications, such as glycosylation, methylation, and acylation (Tsai et al. 2006). We therefore used a phylogenomic approach to develop an updated phylogeny of the *Populus* BAHD acyltransferase family in reference to four other angiosperm taxa. Together with gene duplication and expression analyses, our data suggest that lineage-specific gene duplication is a key process in BAHD family evolution. The results are consistent with a role of the BAHD acyltransferases in diversifying the secondary metabolite repertoire in plants.

Methods

Identification of Putative BAHD Family Members

Published BAHD acyltransferase sequences (D'Auria 2006) were used in initial BLASTP searches against the JGI *Populus trichocarpa* genome v1.1 (Tuskan et al. 2006). Because multiple *in silico* gene prediction programs were used in v1.1, gene models were manually examined for possible structural annotation errors, and to select alternate models if necessary. Gene models located in unanchored short (< 20 kb) scaffolds often represent redundant sequences due to sequence quality or assembly artefacts (Tsai et al. 2011). When used in BLASTN searches against the *Populus* genome, all putative BAHD sequences from the short scaffolds had high similarity to at least one gene model placed in the 19 linkage groups or larger scaffolds, and were removed from further analysis. The remaining putative BAHD gene models were cross-referenced with the recently released *Populus* genome v2.0 available from the Phytozome website (JGI), followed again by manual curation (Table 2.1). Manually curated sequences for erroneous gene models are provided in Table 2.2.

Protein sequences of the putative *Populus* BAHD acyltransferases were aligned with previously characterized BAHD proteins (D'Auria 2006 and Table 2.3) using ClustalW (Larkin et al. 2007), then imported into MEGA v.4.0.2 for motif inspection (Tamura et al. 2007). Sequences which exhibited no HXXXD motif were removed from consideration. Sequences were further screened for a DFGWG-like motif containing at least three of the five amino acids; strict conservation was not required due to known polymorphisms in biochemically characterized BAHD proteins. Exceptions were made for loci highly similar to Clade II members ZmGlossy2 and AtCER, which contain no

DFGWG motif. Finally, sequences less than 300 amino acids in length were removed from the list as likely pseudogenes; these sequences either lack the conserved motifs or represent obsolete gene models from the previous genome releases (Table 2.1). BLASTN searches against the NCBI *Populus* EST database revealed no expression support for any of these suspected pseudogenes.

Similar BLASTP searches were conducted against the *Arabidopsis* TAIR9 database (Swarbreck et al. 2008), the Rice Genome Annotation Database release 5 (Ouyang et al. 2007; Yuan et al. 2005), the *Medicago truncatula* genome database MtDB v2.0 (Retzel et al. 2008), and the *Vitis vinifera* genome database (8X) at Genoscope (Jaillon and et 2007). The sequences were aligned for motif inspection as described above, yielding 55 putative BAHD members in *Arabidopsis*, 84 in *Oryza*, 50 in *Medicago*, and 52 in *Vitis*. For *Oryza*, annotation and final number of genes were determined partially through comparison with the Rice Genome Annotation Database release 6.1. Manual sequence curation revealed two full-length *Arabidopsis* genes previously considered partial sequences (Yu et al. 2009), and the remaining partial sequences along with several BAHD-like members lacking either of the conserved motifs (Table 2.4) were excluded from our analysis. Protein length was not used as a criterion for further curation in *Medicago*, *Oryza*, or *Vitis*, but unusually long or short models were noted in Table 2.4.

Phylogenetic Analysis

Putative BAHD protein sequences from *Populus*, *Arabidopsis*, *Oryza*, *Medicago*, and *Vitis* were aligned along with 69 biochemically characterized BAHD members

(Table 2.3) using the MAFFT v6.717 online server (Katoh et al. 2002; Katoh and Toh 2008). The FFT-NS-i iterative refinement method was run twice, once with default settings using the BLOSUM62 substitution matrix, and once using the JTT200 substitution matrix. The resulting alignments were imported into BioEdit v7.0.9.0 (Hall 1999), where any positions containing less than five sequences were designated as gaps and deleted from the alignment. The data were submitted to the CIPRES portal v2.2 (Miller et al. 2009) for phylogenetic tree construction using RAxML-HPC v7.2.6 (Stamatakis 2006; Stamatakis et al. 2008). Trees were obtained using empirical base frequencies and a maximum likelihood search. The resulting RAxML_bipartitionsBranchLabels.result file was converted to Newick format in Dendroscope v2.3 (Huson et al. 2007) and imported into MEGA v4.0.2 (Tamura et al. 2007) for visualization. Because the topologies of the maximum likelihood trees resulting from use of the two substitution matrices were broadly consistent with each other, only the BLOSUM62-based tree is shown.

In Silico Characterization of Conserved Protein Motifs

Aligned protein sequences from MAFFT were split into separate FASTA files by clade using BioEdit, and subjected to motif analysis using the MINER v2.0 web interface with default settings (La and Livesay 2005a; La and Livesay 2005b; La et al. 2005). Because a minimum of 25 sequences is recommended by the program to achieve good statistical support, Clade IV (five members) and Clade II (18 members) were excluded from the analysis. Putative motifs were identified based on a phylogenetic similarity z-score threshold automatically determined by the MINER. Previous work suggests

thresholds of -1.5 to -2.2 are typical (La et al. 2005); actual thresholds in our study ranged from -2.05 to -2.28. The corresponding sequence alignments for multi-clade motifs were manually trimmed to 25 amino acids, including bordering residues, and submitted to Weblogo v2.8.2 (Crooks et al. 2004) for visualization. Sequence alignments corresponding to representative clade-specific motifs were also trimmed to 10 amino acids and submitted to Weblogo for visualization. Any motifs lacking at least one amino acid conserved at a rate >1.5 bits were not reported.

Putative subcellular localization for all BAHD proteins by clade was examined using WoLF PSORT (Horton et al. 2007; Horton et al. 2006), Predotar (Small et al. 2004), and TargetP (Emanuelsson et al. 2007; Emanuelsson et al. 2000), assigning "plant" as the organism type. The predicted subcellular localization site (mitochondrial, chloroplast, secretory organelles, or any others) for each protein was noted, and overall patterns were summarized for each clade.

Visualization of Putative BAHD Genes on *Populus* Linkage Groups and Identification of Gene Duplication Events

The chromosomal locations of the 100 *Populus* BAHD genes were visualized in ideograms using the software package from Böhringer et al. (2002), based on the *Populus trichocarpa* genome v2.0. Syntenous segments of the genome derived from the "salicoid" genome-wide duplication event (Tuskan et al. 2006) were color-coded according to the position information provided in the SalicaceaeDup.seg file downloaded from Phytozome (JGI). Two types of duplication events were noted: genome-wide duplications originating from the salicoid event, and local duplications. Salicoid

duplications were identified according to Tuskan et al. (2006) based on the SalicaceaeDup.ort.txt file from Phytozome (JGI). Because many of the *in silico* gene model predictions have not been validated (*e.g.*, some represent partial gene models or transposons), the "local duplications" category is used here to include tandem or tandem array duplications with no intervening predicted gene models (Table 2.5). Neither partial BAHD acyltransferase sequences nor transposons were counted as intervening gene models. Three cases deserve special mention. One appears to be a two-gene tandem duplication, involving POPTR_0011s12480 + POPTR_0011s12490 (*AATL16*) and POPTR_0011s12500 + POPTR_0011s12510 (*AATL17*). *AATL16* and *17* were therefore retained as a local duplication pair in our analysis. Another involves *AATL12-13* vs. *AATL14* with an intervening partial BAHD gene model (POPTR_0010s06400). *AATL14* is a salicoid duplicate of *AATL11*, and shares less than 40% protein sequence similarity with the highly homologous *AATL12* and *AATL13* (98% similarity). *AATL14* was thus excluded as part of the tandem array. The other case involves a six-gene tandem array (*HMTL1-6*), separated by a non-BAHD gene model POPTR_0001s45170. Several discrepancies were noted for this region between the two genome assembly versions. The intervening gene model prediction corresponded to a full-length disease resistance protein in v1.1 (eugene3.00012870) but to a partial one in v2.0 (POPTR_0001s45170). *HMTL3* was predicted in an opposite orientation relative to other genes within this region in v2.0, but the corresponding *HMTL2* (eugene3.0012871), *HMTL3* (eugene3.0012869) and the intervening gene models in v1.1 were in the same orientation. The predicted tandem copies also varied between the two versions, presumably due to the difficulty in assembling highly similar sequences. For all these reasons, we tentatively assigned

HMTL1-2 and *HTML3-6* (including the inverted *HMTL3* locus) to two separate tandem duplication blocks in our analysis (Table 2.5).

To search for retrotransposons, BioPerl SeqIO was used to extract the 10-kb sequences immediately upstream and downstream of each of the 100 putative *Populus* BAHD acyltransferases from the v2.0 genome. Sequences were subjected to BLASTX searches against the GenBank non-redundant protein database with an *E*-value cutoff of $1e^{-10}$. The output file was processed with the BioPerl SearchIO scripts, and the results were manually inspected to determine whether the regions of interest were likely to contain retrotransposons based on the descriptions of matches. Only sequences with multiple hits to retrotransposon elements were documented (Table 2.5).

Microarray Data Mining

Affymetrix *Populus* microarray datasets generated in our laboratory (Yuan et al. 2009) were used to investigate BAHD gene expression across genotypes, tissues, and stress treatments. These arrays corresponded to nine experimental groups, including 1) nitrogen-stressed young and expanding leaves of two *Populus fremontii* x *angustifolia* genotypes (1979 and 3200), 2) systemic young and expanding leaves of *Populus fremontii* x *angustifolia* genotype RM5 one week after lower leaf wounding, or systemic expanding leaves and root tips 90 h post-wounding, 3) expanding leaves of *P. tremuloides* genotype 271 following detopping, and 4) methyl jasmonate-elicited suspension cell cultures of *P. tremuloides* genotype L4. All experiments contained respective non-stressed controls and two biological replicates. The arrays were pre-processed by the GC-RMA algorithm using GeneSpring GX 11.0.2 (Agilent

Technologies Inc.). *Populus* probes exhibiting mean raw hybridization intensities of at least 50 in any experimental group were flagged as "present", yielding a list of 24,871 probes, and the rest designated as "absent" and excluded from analysis. Hierarchical clustering was performed using several distance metrics to evaluate the sample clustering patterns. All control and treatment samples from the same experimental group clustered together, except for the expanding leaves from the one week wounding experiment. These arrays were excluded from further analysis. Based on the POParray database (Tsai et al. 2011) and the v2.0 poplar genome (JGI), the filtered list contained a total of 60 probes annotated as BAHD acyltransferases, representing 48 unique BAHD genes. Because the Affymetrix array was designed based on the v1.0 genome release and a large collection of ESTs from several *Populus* species, redundancy is a known issue (Tsai et al. 2011). To minimize redundant representation, we further reduced the list of 60 probes to those that have unique gene matches, and in cases of multi-probe representation, to those that exhibited the highest hybridization signals consistently across multiple samples. The final list included 36 probes with unique gene representation, and 5 probes matching to multiple highly similar genes. The list of BAHD acyltransferase gene-to-probe correspondences can be found in Table 2.6. The BAHD probe expression values from all control samples across genotypes and tissues were grouped by clade and $\log_{10}$-transformed for visualization using the Heatmapper *Plus* tool at the Bio-Array Resource for Plant Functional Genomics (Toufighi et al. 2005). Stress responses of BAHD genes were also visualized in heatmaps using $\log_2$-transformed expression ratios of experimental treatments relative to control samples.

Gene Expression Correlation Analysis

Log-transformed microarray data was imported in to JMP v8.0 (SAS Institute, Inc.) and distribution of expression values for each gene probe was analyzed using histogram plots. The majority of probes did not generate curves similar to a normal distribution. Therefore, we used Spearman's ρ as a non-parametric measure of pairwise correlation for gene expression among genes within each clade. We then organized gene pairs by duplication type (local, salicoid or other) according to Table 2.5, generating box plots for each using SigmaStat v3.5 (Systat Software Inc). For the salicoid duplicates that have also been associated with more recent local duplications, all possible pairwise comparisons between the lone salicoid member and the local duplicates (e.g., *CHATL6* vs. *CHATL1-3*, and *HMTL7* vs. *HMTL1-6*) were included. Kruskall-Wallis one-way ANOVA on Ranks was used to test for differences among any duplication categories, followed by a *post-hoc* Dunn's Method test for pairwise differences between categories.

Quantitative Real-Time PCR Analysis

Apices, leaves at leaf plastochron index (LPI) 0-1 and LPI 8, internodes corresponding to LPI 1-4 and LPI 7-10, and root tips of *P. tremuloides* genotype 271 were flash frozen and ground under liquid nitrogen for RNA extraction. Male and female flowers were collected from wild *P. tremuloides* at field sites near Houghton, Michigan. RNA was extracted from three biological replicates of all samples using the CTAB method (Tsai et al. 2003), quantified via Nanodrop spectrophotometry and quality-checked on a 1% agarose gel. cDNA was synthesized with 5.0 µg of RNA using dT20-VN primers and SuperScript II reverse transcriptase (Invitrogen). RNA samples from the

nitrogen stress microarray experiments detailed above were also used to generate cDNA samples with two biological replicates per condition.

QPCR reactions were carried out in a 12.5 µl reaction volume using cDNA equivalent to 2.5 ng of total RNA, 100 nM each of forward and reverse primers, and the ABsolute^{TM} SYBR Green Master Mix (ABgene) with 0.003% ROX reference dye. Two technical replicates were included for each sample, and sample plates were run on the Mx3005P^{TM} (Stratagene). Relative expression was calculated by the ΔCt method using the geometric mean of three housekeeping genes (elongation factor 1β, cyclophilin, and ubiquitin-conjugating enzyme E2), except for the nitrogen experiment where the last housekeeping gene was excluded due to missing data for some samples. PCR amplification efficiency was calculated using the LinRegPCR program (Ramakers et al. 2003). Primers were designed based on the predicted transcript sequences of the target *P. trichocarpa* gene models and the corresponding GenBank *Populus* ESTs, and wobbles were introduced wherever variation exists. The primer sequences are: *CHATL1/2* forward AGTTWCWTGCAGACACCGAGCGTA, and reverse AGGGCAATGGYMCGACATATCCAA; *CHATL3/6* forward TGGCCCTTCAGARATRTCTGCTCT, and reverse AGTCACGTCAGCCTTRGCCTTTCT; *CHATL4/5* forward ACACCACTGACAACGTTCCGCTTA, and reverse TGTTGCCATTGCCACTGAGTATGC; *elongation factor 1β* forward AAGAGGACAAGAAGGCAGCA, and reverse CTAACCGCCTTCTCCAACAC; *cyclophilin* forward ATGGCTTGATGGGAAACAT, and reverse AATCTCATTAGGATCATTAAAGGACAG; and *ubiquitin-conjugating enzyme E2*

forward CTGAAGAAGGAGATGACARCMCCA, and reverse GCATCCCTTCAACACAGTTTCAMG.

Results

Populus Has More BAHD Acyltransferase Genes Than *Arabidopsis*, *Medicago*, *Oryza*, and *Vitis*

BLASTP searches against the JGI *Populus trichocarpa* genome release v1.1 revealed 149 unique loci with high similarity to biochemically characterized BAHD acyltransferases from a previous review (D'Auria 2006). Manual curation and referencing against the recently released genome v2.0 were conducted to exclude loci lacking a conserved motif (HXXXD or DFGWG), loci that represented redundant, possibly allelic copies, and loci resembling spurious gene models (see **Methods**). The final list of 100 putative *Populus* BAHD acyltransferases was used for all subsequent analyses and annotation (Table 2.1). In the course of our work, another group also annotated the BAHD family in *Populus* (Yu et al. 2009) and reported 94 putative gene models. These models correspond to 74 putative BAHD genes on our list, with one model that matched two v2.0 gene models on our list; the 21 remaining models were either redundant or rejected based on our manual curation criteria (Table 2.1). Similar BLAST search and quality control measures were also performed for the genomes of *Arabidopsis*, *Medicago, Oryza,* and *Vitis*, producing final lists of 55, 50, 84, and 52 putative BAHD genes, respectively (Table 2.4). These lists include ten biochemically

characterized *Arabidopsis* members and one biochemically characterized *Medicago* member (see Table 2.4 for references).

Phylogenetic Analysis Supports Eight Major Clades of Plant BAHD Acyltransferases

Phylogenetic relationships among the BAHD acyltransferases were reconstructed using a maximum-likelihood algorithm, for a collection of 69 biochemically characterized plant BAHD acyltransferases and the putative members from *Populus*, *Arabidopsis*, *Oryza*, *Medicago*, and *Vitis* (Figure 2.1A). The resulting phylogenetic tree is broadly consistent with that of D'Auria (2006), who sorted biochemically characterized BAHD acyltransferases into five major groups. Our expanded analysis suggests that a grouping of eight major clades is now warranted, a finding consistent with previous, single-genome-based, neighbour-joining analyses (D'Auria et al. 2007b; Luo et al. 2007; Yu et al. 2009). In particular, a strongly-supported clade comprised entirely of BAHD acyltransferases lacking biochemical characterization data was sister to the group of proteins previously designated as Clade I by D'Auria (2006). To maintain consistency, we adopted a similar clade nomenclature, and name the previous and the "new" groups as Clades Ia and Ib, respectively. Clades Ia and Ib correspond respectively to the *Populus* clades Vb+Vc and Va, and to the *Arabidopsis* clades IIb and IIa reported by Yu et al. (2009). Another strongly supported clade containing the *Petunia* acetyl CoA:coniferyl alcohol acetyltransferase (CFAT, Dexter et al. 2007) was sister to the group classified by D'Auria as Clade III (D'Auria 2006). We name the previous and the "new" clades as IIIa and IIIb, respectively; these correspond to the *Populus* clades IV and II and *Arabidopsis* clades IV and IIIa in Yu et al. (2009). Members of the former Clade V (D'Auria 2006)

clustered into two well-supported groups in our analysis, renamed hereafter Clades Va and Vb. These clades correspond to Yu et al.'s (2009) clades Ia and Ib for both *Populus* and *Arabidopsis*. Characterized proteins in Clade Va tend to be involved in volatile ester formation, while those in Clade Vb are closely related to hydroxycinnamoyltransferases (HCTs) responsible for the synthesis of chlorogenic acid and monolignols. Our analysis also placed Clade IV basal to Clades Va and Vb, with good support. The remaining sequences clustered into one strongly supported group corresponding to D'Auria's Clade II (2006).

The distribution of sequences among the five species varied within each clade (Figure 2.1B). *Populus* and *Oryza* have the largest number of BAHD members overall, and collectively these made up the majority of Clades Ia, Va, and Vb. *Populus* also predominated in the dicot-specific Clade IIIa, while Clade IV was monocot-specific. Taxon bias was also evident in Clades Ib and IIIb, where *Medicago* and *Vitis*, respectively, were over-represented. When analyzed by species, Clade Va, the largest clade, remained the largest group in all taxa, except in *Medicago* where Clade Ib predominated (Figure 2.1C). Clades II, IIIb, and IV had the lowest representations overall, consistent with their small overall sizes. The major exception to this pattern was *Vitis*, which showed a relatively higher representation of Clade IIIb, coinciding with a much lower representation of Clade Ia. Other species-biased patterns included high (>20%) representation for Clade IIIa in *Populus*, Clade Ib in *Arabidopsis*, and Clade Ia in *Oryza*.

Closer examination of the phylogeny revealed that BAHD sequences from a single taxon tended to cluster together, especially within the larger clades. In Clade Ia,

all sequences from the five taxa formed lineage-specific groups with strong bootstrap support, except for one well-supported subgroup (Figure 2.2, bracket). *Oryza* sequences were basal to all eudicot sequences in this clade. Two strongly-supported subclades consisting of a combined total of sixteen *Populus* sequences comprised another large, but in this case weakly-supported, group, sister to a group of eight *Arabidopsis* sequences (Figure 2.2), including a malonyl CoA:cyanidin 3,5-diglucoside transferase (At5MaT, D'Auria et al. 2007b; Luo et al. 2007). Similar, but less dramatic patterns were observed for Clade Ib (Figure 2.3). While the two most basal subgroups in this clade did not show strong taxon specificity, the two remaining subgroups each comprised five taxon-specific branches with strong support (Figure 2.3). In accordance with its overrepresentation overall in Clade Ib, *Medicago* exhibited substantial taxon-specific expansions within these two branches.

Taxon-specific clustering appeared more scattered in Clade IIIa, perhaps because the larger of the two major branches was poorly resolved (Figure 2.4). Ten *Populus* sequences formed a well-supported subclade together with a *Clarkia breweri* acetyltransferase involved in benzyl acetate formation (CbBEAT, Dudareva et al. 1998), and with an uncharacterized *Vitis* sequence. A smaller subclade contained five *Populus* sequences, and a third taxon-specific subclade containing seven of the nine *Arabidopsis* sequences in Clade IIIa also had high bootstrap support.

As the largest phylogenetic group, Clade Va contained a number of highly-derived branches, some specific to gymnosperms, monocots, or dicots (Figure 2.5). The largest well-supported branch in this clade contained four taxon-specific clusters of at least seven members (Figure 2.5, boxed), one each for *Vitis* (eight members), *Populus*

(seven), *Medicago* (nine), and *Oryza* (eleven). *Oryza* sequences were over-represented in this clade and fell mainly into two large branches with moderate bootstrap support. One was *Oryza*-specific as mentioned above, and the other contained three eudicot sequences (Figure 2.5). Taxon-specific clustering was not as evident in Clade Vb, except for a well-supported branch of seven *Oryza* sequences, sister to a group of hydroxycinnamoyltransferases (HCT/HQT) involved in biosynthesis of lignin, chlorogenic acid, and other phytoalexins (Figure 2.3).

Clade II lacked species-specific clustering patterns, as members were more evenly distributed among species (Figure 2.3). Clade IIIb was relatively small, and exhibited some degree of taxon-specific clustering. The largest such grouping comprised nine *Vitis* sequences, consistent with their overrepresentation in this clade (Figure 2.3). A four-member subclade of *Oryza* sequences and a three-member subclade each for *Arabidopsis* and *Medicago* were also evident. Clade IV was the smallest clade and was restricted to monocots, as mentioned previously.

With regard to *Populus*, species-specific expansion was evidenced within Clades Ia, IIIa, and Va. Because the *Populus*-specific subgroup in Clade Ia is most closely related to several biochemically characterized malonyltransferases from *Arabidopsis*, *Medicago*, and *Glycine*, we have named members of this clade as malonyltransferase-like (MATLs). The sequences in the *Populus*-specific branch are MATL1-14 and 16-17. We designated all *Populus* sequences in Clade IIIa as alcohol acyltransferase-like (AATLs), after the numerous characterized alcohol acyltransferases within that clade. The *Populus*-specific branch includes AATL1, 3, 7-9, 18-19, and 22-24. We refer to the three *Populus* clusters within the largest branch of Group Va by three names. First, we named the set of

four *Populus* sequences clustering with two *Malus* sequences and a set of *Vitis* sequences, including an anthraniloyl-CoA:methanol acyltransferase from *Vitis labrusca* (VlAMAT, Wang and De Luca 2005), as AMAT-like (AMATLs). Next, we refer to the six *Populus* proteins most closely related to the *Arabidopsis* acetyl CoA:*cis*-3-hexen-1-ol acetyl transferase (D'Auria et al. 2002) as CHAT-like (CHATLs). Finally, the subgroup of seven *Populus* sequences that fell into a poorly-resolved region of Clade Va, most closely to a tigloyl-CoA:(–)-13α-hydroxymultiflorine/(+)-13α-hydroxylupanine *O*-tigloyltransferase from *Lupinus* (LaHMT/HLT, Okada et al. 2005), were named HMT-like (HMTLs).

New Family-wide and Clade-Specific Motifs are Present in BAHD Acyltransferases

The large number of BAHD genes available from sequenced plant genomes presents an opportunity to expand the analysis of conserved motifs in this family beyond the two known functional domains, HXXXD and DFGWG. We subjected sequences from each clade to motif analysis using MINER v2.0 (La and Livesay 2005a; La and Livesay 2005b; La et al. 2005). Clades II and IV were excluded from the analysis due to their small sizes. Using a sequence window of five amino acids and the default z-score threshold, four to nine motifs were predicted for each clade (Figure 2.6, Figure 2.7). MINER identified the DFGWG motif in four of the six tested clades (Ia, Ib, IIIa, and Va). Although it did not meet the MINER threshold, visual inspection revealed high conservation of this motif in Clades IIIb and Vb as well (Figure 2.6). This supports the validity of our approach towards the identification of conserved motifs. The HXXXD

motif escaped detection by MINER, but this was expected since the motif contains a variable core.

Two new motifs were identified with multi-clade conservation. The first motif had a consensus of YPLAGR beginning around position 71-78, and was predicted in Clades IIIa, Va, and Vb. Manual inspection of the other clades identified a similar motif in this region, but with notable variability from the consensus, especially for the two flanking residues (Figure 2.6). The second motif had a consensus of QVTX(F/L)XCGG around position 136-156 and was predicted in Clades Ib, IIIa, and Va. Manual inspection revealed that QVT was highly conserved in the other three clades, but CGG was poorly conserved in Clades Ia and Ib (Figure 2.6). Clade-specific motifs were also observed, several of which were located near the N-terminus of the protein: the LTFFD motif from Clade Ia was located at positions 33-37, the IKPSSPTP motif of Clade IIIa at positions 11-18, and SNLDL from Clade Vb at positions 25-29 (Figure 2.6). Because the N-terminus often contains targeting peptide sequences, we examined the predicted protein subcellular localization patterns by clade using three different prediction programs. However, we found no evidence for a link between the observed clade-specific N-terminal motifs and the predicted subcellular targeting of the BAHD proteins (Figure 2.8).

Although Clade II was too small for motif analysis, we note that none of its members would have been accepted using our initial search criteria (both HXXXD and DFGWG present). The two original clade members, ZmGlossy2 and AtCER2, are known to participate in cuticular wax biosynthesis based exclusively on genetic characterization studies (Negruk et al. 1996; Tacke et al. 1995; Xia et al. 1996). In the

absence of biochemical data, it remains debatable as to whether Clade II members should be considered true BAHD acyltransferases.

Multiple Gene Duplication Types Have Contributed To BAHD Family Expansion in the *Populus* Genome

Populus has experienced at least two genome-wide duplication events, the salicoid event approximately 60-65 MYA and the older eudicot triplication event, as well as numerous segmental and tandem duplication events (Tang et al. 2008; Tuskan et al. 2006). We sought to determine whether the various types of gene duplications contributed towards the expansion of the *Populus* BAHD family, especially with regard to *Populus*-specific subclades (HMTLs, CHATLs, and subgroups of MATLs and AATLs). Overall, we found sixty BAHD genes were associated with recent (salicoid or local) duplications (Table 2.5), accounting for more than half of the BAHD acyltransferases in *Populus* (Table 2.7). This is broadly consistent with previous analysis of chromosomal location of BAHD acyltransferases in *Populus*, which mapped 25 of 58 genes to homeologous chromosome segments or tandem duplication blocks based on the v1.1 genome release (Yu et al. 2009). Events were spread approximately evenly across the two duplication types, with a greater number of local (e.g., tandem) duplications overall. Duplications were found in all but the two smallest clades (II and IIIb). Salicoid and local duplications were overrepresented in Clades Ib, Va, and Vb relative to the genome overall. Such duplications impacted every member of Clade Ib (three salicoid pairs, one local pair and one local triplet), all but two genes in the largest subclade of Va (Figure 2.5, boxed; including two salicoid duplications, three local pairs, one local triplet,

and one local quadruplet), and all but one member of Clade Vb (including two local pairs and two salicoid pairs; Figure 2.9, Table 2.5). For two subclades within the large, poorly resolved region in Clade Va, multiple local duplications appear to have followed genome-wide duplication events in one of the two salicoid paralogues (Figure 2.9, Table 2.5). The first instance is the relationship between *HMTL7* on linkage group (LG) XI and the *HMTL1-6* cluster on LG I. The second is the relationship between *CHATL6* on LG XIX and the *CHATL1-3* triplet on LG XIII.

Although Clade IIIa exhibited several duplications, the *Populus*-dominated AATL subclade had just one tandem pair (*AATL23* and *AATL24*). Clade Ia had the lowest rate of duplications among the larger clades, with two local triplets within the *Populus*-dominated MATL subclade (Table 2.7). The relatively low numbers of local and salicoid duplications in the *Populus*-dominated AATL and MATL subclades raises the possibility that some of these genes might have originated through other mechanisms, such as transposable elements. We therefore searched for the presence of retrotransposons within the two 10-kb windows flanking either side of each *Populus* BAHD gene. We found retrotransposon associations in each clade, covering over one third of the family as a whole, although the majority of associated genes were flanked on only one end (Table 2.7). Retrotransposon associations were frequently observed for recently duplicated genes (Table 2.5, Table 2.7). Retrotransposon associations were overrepresented in Clade Va, noted for all AMATLs and the majority of CHATLs and HMTLs (Table 2.5, Table 2.7). However, all of these gene models contained at least one intron (Table 2.1), suggesting that retrotransposition is unlikely to be a direct cause of duplication. Retrotransposon associations were underrepresented in Clade IIIa and absent from the

AATL *Populus*-dominated subclade (Table 2.5, Table 2.7). Despite its average representation of retrotransposon associations, Clade Ia had the greatest number of genes with retrotransposons flanking both sides (Table 2.7). Two such genes, *MATL12* and *13*, formed a strongly supported branch with *MATL10*. All three are located on LG IV (Figure 2.9), lack predicted introns (Table 2.1), and share a high degree of nucleotide identity with one another (98%). Although preliminary, our analysis suggests that retrotransposons have contributed to the duplications of some BAHD genes.

Some Recently Duplicated BAHD Acyltransferases are Differentially Expressed

To investigate expression of *Populus* BAHD genes, we mined a set of nine Affymetrix microarray datasets encompassing five different genotypes and four different tissue types generated in our laboratory (Yuan et al. 2009). After excluding probes that had consistently low expression across all samples (see **Methods**) and annotating probes based on the POParray database (Tsai et al. 2011), we obtained expression data for 41 probes corresponding to 48 BAHD genes (some probe sequences match multiple gene targets, and some gene targets are represented by multiple probes). Pairwise correlations of BAHD gene expression across all microarray experiments were computed and the results organized by duplication type (Figure 2.10). Median Spearman rank correlations were significantly different among the duplication categories according to one-way ANOVA ($p<0.001$). Not surprisingly, median correlations for gene pairs derived from local or salicoid duplications were significantly higher than for other types of (all possible) gene pairs (Figure 2.10).

When the log-transformed microarray data were visualized as a heatmap, expression across the BAHD family as a whole was biased towards leaves, and we did not observe clear differences in expression patterns among the major clades (Figure 2.11A). Within the major clades, genotype- and/or tissue-dependent expression patterns were evident. For example, root-specific expression dominated in the HMTL subclade, while the majority of other Clade Va genes showed the more typical leaf-biased expression (Figure 2.11A). In another case, *HCT1* and *HCT6* were relatively uniformly expressed in all three *P. fremontii x angustifolia* hybrid genotypes examined, while *HCT5* and *HCT7* were detected only in genotype 1979 (Figure 2.11A). *HCT2*, on the other hand, was most abundant in roots. Expression patterns diverged for closely related genes in several cases, including genes within the *Populus*-dominated subclades. For example, *MATL4* was biased towards *P. fremontii* x *angustifolia* genotype 1979 relative to *MATL1-3*, which were more evenly expressed across genotypes and tissues. The *Populus*-dominated AATL subclade includes *AATL3*, which was preferentially expressed in cell suspension cultures, as well as *AATL7*, *23,* and *24,* which exhibited different expression patterns by leaf age and genotype. The CHATL cluster includes two members (*CHATL3* and *6*) that were fairly evenly expressed across sampled tissues, and two (*CHATL1* and *2*) that were detected only in leaves. The more divergent *CHATL4/5* were most strongly expressed in non-photosynthetic tissues, yielding an overall pattern that resembled the HMTLs more than the other CHATLs (Figure 2.11A).

QPCR was performed to verify the expression patterns of closely related *CHATL* transcripts observed by microarray analysis, using an independent set of *P. tremuloides* tissues (Figure 2.12). Specific primers were designed to distinguish among the three

paralogous pairs with different duplication history (Table 2.5): *CHATL1/2*, *CHATL3/6* and *CHATL4/5*. *CHATL1/2* were expressed relatively consistently across all leaf and stem internode tissues sampled, but were lower in root and flower tissues (near or below the corresponding microarray threshold marked by a dotted line in Figure 2.12, Panel A). *CHATL3/6* were most strongly expressed in young leaves and roots, followed by apices and mature leaves, and were much lower in stem and flower tissues. The transcript levels of *CHATL4/5* were very low overall, with the highest levels detected in roots, similar to the microarray data of *P. fremontii* x *angustifolia*. Overall, the QPCR data were broadly consistent with the microarray results, and support the idea that the three pairs of *CHATL* genes have diverged in their expression patterns despite their high homology.

We next analyzed the microarray data to examine the responses of BAHD gene expression to four different stress treatments, including nitrogen limitation, wounding, detopping, or methyl jasmonate feeding, across several tissues and/or genotypes. Again, no clear overall patterns by clade were observed, and the differential gene expression patterns observed among some paralogous genes described above also held for the stress treatments (Figure 2.11B). Additional evidence of functional divergence was observed. For example, Clade IIIa member *AATL7* showed its strongest upregulation in young leaves one week after wounding, while the responses of *AATL23* and *AATL24* were most drastically affected via down-regulation in expanding leaves following detopping. The leaf-expressing *CHATL* genes were generally up-regulated by nitrogen stress in *P. fremontii x angustifolia* genotype 1979, except *CHATL6* in expanding leaves. However, the trend was more variable in genotype 3200 (Figure 2.11B), despite similar baseline expression of these genes between the two genotypes (Figure 2.11A). QPCR analysis of

the same suite of samples confirmed this general discrepancy between the two genotypes (Figure 2.12B), although the degree of expression changes varied between the two analytical (microarray vs. QPCR) methods. The data hint at differential expression among closely related BAHD genes in response to nitrogen stress between different *Populus* hybrid genotypes. Future investigation would help determine how widespread this pattern is across the BAHD family and a broader range of genotypes in this genus.

Discussion

BAHD Family Expansion as a Factor Enabling Metabolic Diversification

Across the five angiosperm genomes investigated here, we observed numerous differential lineage expansions within the BAHD acyltransferase phylogeny. Examination of retained gene copies following duplications in *Populus* revealed that the majority of BAHD genes, at least in this genus, are associated with recent genome-wide as well as local duplication events. An estimated 32% of all v2.0 *Populus* genes (6655 pairs or 13268 unique gene models) were derived from the salicoid duplication event (JGI). However, only 26% of the *Populus* BAHD acyltransferases were associated with the salicoid duplication. Tandem or local duplications, on the other hand, accounted for over one-third (36%) of the *Populus* BAHD genes, much higher than the genome average estimated at 16% (Tuskan et al. 2006). It thus appeared that local duplications were over-represented and genome-wide duplications were under-represented in the *Populus* BAHD family relative to the genome average. We speculate that this pattern may be generally applicable to the other angiosperm genomes surveyed in this study. Local duplications

might be more likely than polyploidization events to account for the observed taxon-specific expansions of BAHD acyltransferases. This was indeed the case for the *Populus*-dominated HMTL and CHATL subclades, where the majority of the genes were derived from local duplications, and to a much lesser extent, for the MATL subclade. In contrast, only two of the ten members in the *Populus*-dominated subclade among the AATLs were implicated in any duplication event. Preliminary molecular clock analysis suggested that the divergence times among members of the *Populus*-dominated MATL and AATL subclades were similar and predated the salicoid duplication event. This suggests that other duplications, prior to the salicoid duplication event but after the eudicot triplication event, probably contributed to the *Populus*-specific expansion as well.

Previous work has shown that genes involved in stress responses, including secondary metabolic genes, are more likely than average to experience lineage-specific diversification via tandem duplication (Hanada et al. 2008). When placed in a metabolic pathway context, we suggest that taxon-specific, local duplication-derived expansion of gene (sub)families may be characteristic of enzymes that occupy a terminal or tangential position in a metabolic pathway. Conversely, enzymes with an intermediate position in a core pathway would likely retain a more constant number of gene copies across taxa due to evolutionary constraint for balanced stoichiometry between enzymes acting within the same pathway. In support of this idea, HCTs known to be involved in intermediate steps of monolignol biosynthesis formed a multi-taxon cluster within Clade Vb, encoded by 1-2 genes in all sequenced genomes. In *Populus*, Clade Vb diversified about equally via salicoid and local duplication events. In contrast, the sister Clade Va exhibited extensive taxon-specific clustering (boxed region, Figure 2.5); *Populus* genes in the major subclade

were associated with more than three times as many local duplications as retained salicoid duplicates. The biochemically characterized enzymes within this branch are all involved in the final step of various volatile ester and alkaloid ester biosynthetic pathways (Boatright et al. 2004; D'Auria et al. 2002; El-Sharkawy et al. 2005; Li et al. 2006; Okada et al. 2005; Souleyre et al. 2005; Wang and De Luca 2005).

Taxon-specific phylogenetic expansions have also been observed within the *O*-methyltransferase (OMT, Constabel and Lindroth 2010; Han et al. 2007; Lam et al. 2007) and glycosyltransferase (GT, especially group 1, Tsai and Johnson unpublished; Cao et al. 2008; Yin et al. 2010) families. Like BAHD acyltransferases, OMTs and GTs form large families, and collectively the three are responsible for the elaboration (acylation, methylation, and glycosylation) of a wide range of secondary metabolites (D'Auria 2006; Tsai et al. 2006). These modifications increase the diversity of natural products, with regard to both their chemical structures and biological activities, and they hold chemotaxonomic value due to their taxon-specificity (Greenaway et al. 1991; Greenaway et al. 1992; Richardson and Young 1982). Studies from the OMT and GT1 families have shown that many of these enzymes possess promiscuous substrate specificity (Bowles et al. 2006; Kopycki et al. 2008). This, coupled with the tendency of taxon-specific diversification of some of the subfamily members, may be a means to afford metabolic plasticity. Consistent with this idea, multifunctional OMTs from *Thalictrum tuberosum* formed a taxon-specific subclade, while those with a limited substrate range formed multi-taxon subclades (Lam et al. 2007). In the case of BAHD acyltransferases, numerous biochemically characterized Clade Va members involved in volatile ester biosynthesis have been shown to accept multiple substrates and/or donors, at least *in vitro*

(Boatright et al. 2004; D'Auria et al. 2002; El-Sharkawy et al. 2005; Okada et al. 2005; Souleyre et al. 2005; Wang and De Luca 2005). Many of these were found in the same subclade where taxon-specific diversification was common (Figure 2.5, boxed region). These data further support the view that differential lineage expansions of the BAHD family may be linked to taxon-specific metabolic diversification.

Implications of Divergent Paralogue Expression and Newly Identified Conserved Motifs for BAHD Acyltransferase Function

Given the extent of gene duplications in the *Populus* BAHD family, it was not surprising that many of the closely related members retained similar expression patterns. This, however, was not universal, as we noted for the CHATLs, especially in stressed tissues. While genes with high sequence similarity are likely to have similar biochemical functions, the differential expression patterns observed here suggest that physiological functions might have already diverged for some recent duplicates. Our examination of BAHD acyltransferase expression in *P. fremontii* x *angustifolia* and *P. tremuloides* under various stress conditions builds upon the previous analysis in *P. trichocarpa* by Yu et al. (2009). Overall, these results are broadly consistent in that BAHD expression was generally much stronger in photosynthetic than non-photosynthetic tissues. In the present study, the expression bias towards leaves was generally consistent across genotypes. Relatively little difference was observed between genotypes under control conditions, but differences became evident under stress. Biotic and abiotic stresses are known to influence the relative proportions of secondary metabolites in *Populus* (reviewed by Chen et al. 2009; Constabel and Lindroth 2010), and different genotypes differ in both

secondary metabolite diversity and quantities (Greenaway et al. 1991; Greenaway et al. 1992). The variety of BAHD expression patterns in stressed tissues, either among recently duplicated genes, between closely related genotypes or during leaf development, is consistent with a role for BAHD genes in regulating secondary metabolite accumulation and diversity in *Populus*.

The two known conserved motifs in BAHD acyltransferases, HXXXD and DFGWG, have been implicated in the binding of the acyl-CoA donor and in the structural integrity of the enzyme-donor complex, respectively (Bayer et al. 2004; Suzuki et al. 2003; Unno et al. 2007). Our analysis revealed two new motifs, YPLAGR and QVTX(F/L)XCGG, that were conserved across multiple clades. The YPLAGR motif corresponds to the small α-helix-3 (α-3) on the crystal structure of vinorine synthase (Ma et al. 2005), but no mutagenesis analysis targeting this region has yet been conducted. The lower conservation of this motif in Clade Ia is correlated with a lack of this α-helix and an extra string of 9-14 residues (Unno et al. 2007), positioned between the corresponding Gly and Arg residues of the YPLAGR motif on the Clade IIIa enzyme vinorine synthase (Ma et al. 2005). The QVTX(F/L)XCGG motif is eight amino acids upstream of the HXXXD motif, spanning β-6 and β-7 of vinorine synthase (Ma et al. 2005), or β-9 and β-10 of Dm3MaT3 (Unno et al. 2007). Previous work in vinorine synthase (Bayer et al. 2004) has established the functional importance of the Cys residue in the QVTX(F/L)XCGG motif. A Cys-to-Ala point mutation reduced enzymatic activity by 90%, an impact only exceeded by mutation of the His or Asp residues in HXXXD (Bayer et al. 2004). Functional support for several clade-specific motifs can also be garnered. The LTFFD motif conserved in Clade Ia maps to α-1 on the Dm3MaT3 crystal

structure, near the acyl acceptor binding site (Unno et al. 2007). Site-directed mutagenesis of three adjacent α-1 residues on Dm3MaT1 reduced enzyme activity (Unno et al. 2007), supporting the importance of this motif. Another Clade Ia-specific motif, YFGNC (Figure 2.7), is thought to be involved in anthocyanin acyltransferase interactions with malonyl-CoA (Unno et al. 2007). Site-directed mutagenesis of vinorine synthase targeting the first Ser residue in the IKPSSPTP motif (Bayer et al. 2004) has also implicated its role in Clade IIIa enzyme function. Given the reported structural diversity of the acyl acceptor binding sites (Unno et al. 2007), they are more likely to exhibit sequence conservation by clade. In this regard, the suite of clade-specific motifs that we identified should be of value in future structural modelling and mutagenesis studies to understand the diverse enzyme functions in the large BAHD acyltransferase family.

Conclusions

Our phylogenomic analysis expanded and improved upon the previous BAHD family phylogeny, highlighting two major clades for which almost no biochemical data has yet been generated. Our analysis also identified striking patterns of differential expansion of the BAHD family across five angiosperm taxa, including numerous taxon-specific subclades. This finding may provide a basis for understanding the differentiation of secondary metabolism across taxa. Examining clusters of homologous genes within *Populus* demonstrated that tandem gene duplication has been an important evolutionary force for BAHD diversification within this genus, particularly with respect to two

lineage-specific expansions. The retention of salicoid duplicates and likely retrotransposition events have also contributed to the large number of BAHD genes in this taxon. Microarray analysis showed diversity of gene expression among some highly homologous genes in *Populus*, suggesting that some recently duplicated BAHD paralogues have undergone functional divergence in this genus. The discovery of two multi-clade conserved protein motifs as well as clade-specific motifs supports previous research on BAHD enzyme structure and biochemical function, while opening the door to future investigation on the structural basis of donor and substrate specificity within and between clades.

Authors' Contributions

LKT conducted BAHD sequence alignment and manual annotation, performed phylogenetic analysis, motif identification, microarray data analysis, QPCR, and drafted the manuscript. VEJ conducted all BLAST searches, handled large-scale data extraction from external databases, and developed ideograms. CJT conceived of and coordinated the study, participated in BAHD annotation and microarray data analysis, and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Jim Leebens-Mack for advice on the phylogenetic analysis, and Jim Leebens-Mack and Christopher Frost for critical reading of the manuscript. Funding for

this work was provided by the U.S. National Science Foundation Plant Genome Research Program (Nos. DBI-0421756 and DBI-0836433).

Tables

Table 2.1: Summary of Putative BAHD Acyltransferases in the *Populus trichocarpa* Genome

BAHD acyltransferase loci are listed by clade, then by corresponding JGI v2.0 and v1.1 gene models and previously assigned names (Yu et al. 2009). Protein length, exon number, and intron number are included along with manual curation notes.

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
MATL1	POPTR_0001s40570	eugene3.01460050	Ia	476	1	0	PtACT18	**MA**lonyl**T**ransferase-**L**ike
MATL2	POPTR_0001s40600	eugene3.01070001	Ia	478	1	0	PtACT21	
MATL3	POPTR_0001s40620	eugene3.01460042	Ia	479	1	0	PtACT20	
MATL4	POPTR_0001s45940	eugene3.00012958	Ia	442	2	1	PtACT3	
MATL5	POPTR_0004s09280	estExt_fgenesh4_pg.C_LG_IV0624	Ia	463	1	0	PtACT5	
MATL6	POPTR_0004s09520	eugene3.65260001	Ia	443	2	1	PtACT9	
MATL7	POPTR_0004s09530	eugene3.00040594	Ia	473	1	0	PtACT7	
MATL8	POPTR_0004s09550	eugene3.00040592	Ia	473	1	0	PtACT6	
MATL9	POPTR_0004s10330	gw1.IV.4081.1	Ia	471	1	0	PtACT2	
MATL10	POPTR_0004s10920	eugene3.02990009, eugene3.03090010	Ia	476	1	0	PtACT13, PtACT14	
MATL11	POPTR_0004s11990	eugene3.00041024	Ia	471	1	0	N/A	
MATL12	POPTR_0004s18920	eugene3.00040923	Ia	476	1	0	PtACT11	
MATL13	POPTR_0004s19020	eugene3.20150002	Ia	472	1	0	PtACT12	
MATL14	POPTR_0009s02480	eugene3.00091450	Ia	482	1	0	PtACT24	
MATL15	POPTR_0009s06800	eugene3.00091018	Ia	471	1	0	PtACT1	
MATL16	POPTR_0010s21520	eugene3.00102012	Ia	478	1	0	PtACT10	
MATL17	POPTR_0017s13040[b]	eugene3.00170502	Ia	473	1	0	PtACT16	v1.1 sequence used for phylogenetic analysis

Table 2.1, Continued:

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
MATL18	POPTR_0019s14060	eugene3.00191056, eugene3.00191057, fgenesh4_pg.C_LG_XIX001038, eugene3.48300001	Ia	466	1	0	PtACT26, PtACT28, PtACT29, PtACT27	
ATL1[a]	POPTR_0003s05570	fgenesh4_pg.C_LG_III000356, grail3.0037009501	Ib	329+	1	0	PtACT31	**A**cyl**T**ransferase-**L**ike
ATL2	POPTR_0003s05580	fgenesh4_pg.C_LG_III000356, grail3.0037009501	Ib	440	1	0	PtACT31	
ATL3	POPTR_0003s05590	eugene3.20970001, estExt_fgenesh4_pg.C_20970002	Ib	440	1	0	PtACT30, PtACT32	
ATL4	POPTR_0006s09870	eugene3.00060844	Ib	476	1	0	PtACT34	
ATL5	POPTR_0008s06520	gw1.VIII.1749.1	Ib	488	1	0	N/A	
ATL6	POPTR_0010s19980	gw1.X.1848.1	Ib	441	2	1	N/A	
ATL7	POPTR_0012s12890	eugene3.00121077,	Ib	451	2	1	PtACT37	
ATL8	POPTR_0014s02560	eugene3.00400329, fgenesh4_pg.C_scaffold_40000319	Ib	460	1	0	N/A	
ATL9	POPTR_0014s02570	eugene3.00400327, eugene3.00400328	Ib	455	1	0	N/A	
ATL10	POPTR_0015s12810	estExt_fgenesh4_pg.C_LG_XV1167	Ib	436	2	1	PtACT35	
ATL11	POPTR_0016s11990	estExt_Genewise1Plus.C_LG_XVI2952	Ib	478	1	0	PtACT33	
CERL1	POPTR_0001s32660[b]	eugene3.00012234	II	443	2	1	N/A	E**CERI**FERUM2-**L**ike (Negruk et al. 1996; Xia et al. 1996); manual sequence curation
CERL2	POPTR_0005s05380	eugene3.00700077	II	450	2	1	N/A	
CERL3	POPTR_0005s19690	eugene3.00050609	II	436	2	1	Pt-CER2.1	
CERL4	POPTR_0013s03730	match found, but no gene model predicted	II	450	2	1	N/A	manual sequence curation
CERL5	POPTR_0018s01250	gw1.XVIII.1452.1	II	445	1	0	PtACT66	manual sequence curation
AATL1	POPTR_0001s31750	grail3.0017017601	IIIa	460	1	0	PtACT78	**A**lcohol **A**cyl**T**ransferase-**L**ike
AATL2	POPTR_0002s01180[b]	eugene3.00020091	IIIa	449	1	0	PtACT90	v1.1 sequence used for phylogenetic analysis

Table 2.1, Continued:

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
AATL3	POPTR_0004s01720	fgenesh4_pg.C_LG_IV000056	IIIa	433	1	0	PtACT80	
AATL4	POPTR_0005s27190	eugene3.00051554	IIIa	435	2	1	N/A	
AATL5	POPTR_0006s01180[d]	gw1.VI.1340.1[d]	IIIa	403	3	2	N/A	
AATL6	POPTR_0006s01190	estExt_fgenesh4_pg.C_LG_VI0095	IIIa	432	1	0	PtACT91	
AATL7	POPTR_0006s03260[b]	fgenesh4_pg.C_LG_VI000275	IIIa	394	2	1	PtACT75	manual sequence curation
AATL8	POPTR_0006s03450	fgenesh4_pg.C_LG_VI000293, eugene3.00060321	IIIa	439	1	0	PtACT76, PtACT77	
AATL9	POPTR_0007s00580	eugene3.00070043	IIIa	340	3	2	N/A	
AATL10	POPTR_0007s00980	eugene3.00070082	IIIa	383	3	2	PtACT92	
AATL11	POPTR_0008s18070	grail3.0009029801	IIIa	439	1	0	PtACT82	
AATL12	POPTR_0010s06380[b]	e_gw1.X.4991.1[b]	IIIa	430	1	0	PtACT83	manual sequence curation
AATL13	POPTR_0010s06390	eugene3.00100492, eugene3.12430001	IIIa	430	1	0	PtACT84, PtACT85	
AATL14	POPTR_0010s06410	grail3.0036015001	IIIa	445	1	0	N/A	
AATL15	POPTR_0010s06640	fgenesh4_pg.C_LG_X000480	IIIa	441	1	0	PtACT81	
AATL16	POPTR_0011s12490	eugene3.07910001, eugene3.07910002, eugene3.07910004	IIIa	451	1	0	PtACT88, PtACT89	
AATL17	POPTR_0011s12510	eugene3.00110968	IIIa	451	1	0	PtACT87	
AATL18	POPTR_0015s14800	fgenesh4_pm.C_LG_XV000381	IIIa	434	1	0	PtACT73	
AATL19	POPTR_0015s14850	eugene3.00151049	IIIa	430	1	0	PtACT72	
AATL20	POPTR_0017s04320	fgenesh4_pg.C_LG_XVII000003	IIIa	430	1	0	PtACT93	
AATL21	POPTR_0017s05330	fgenesh4_pg.C_scaffold_64000145	IIIa	426	2	1	PtACT94	
AATL22	POPTR_0017s05810	fgenesh4_pg.C_scaffold_64000105	IIIa	438	1	0	PtACT69	
AATL23	POPTR_0019s01520	eugene3.01170047	IIIa	440	1	0	PtACT70	
AATL24	POPTR_0019s01540	grail3.0117003001	IIIa	435	1	0	PtACT71	

Table 2.1, Continued:

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
CFATL1	POPTR_0001s01020	fgenesh4_pg.C_scaffold_29000346	IIIb	448	1	0	PtACT68	Acetyl-CoA:**C**oni**F**eryl Alcohol **A**cetyl**T**ransferase-**L**ike (Dexter et al. 2007)
CFATL2	POPTR_0007s15050	estExt_fgenesh4_pg.C_LG_VII1284	IIIb	494	2	1	PtACT67	
ABTL1	POPTR_0001s33390	fgenesh4_pm.C_LG_I000960, gw1.I.3198.1	Va	442	2	1	PtACT41, PtACT42	**A**cetyl/**B**enzoyl**T**ransferase-**L**ike (after sequences described in D'Auria 2006)
ABTL2	POPTR_0002s03350	fgenesh4_pm.C_LG_II000154	Va	444	3	2	N/A	
ABTL3	POPTR_0002s24610	eugene3.00002504	Va	441	2	1	N/A	
ABTL4	POPTR_0004s05280	eugene3.00040425	Va	448	2	1	N/A	
ABTL5	POPTR_0005s25250	gw1.V.1448.1	Va	454	2	1	N/A	
ABTL6	POPTR_0006s16070	eugene3.01570012	Va	470	1	0	PtACT43	
ABTL7	POPTR_0008s07130	fgenesh4_pm.C_LG_VIII000282	Va	435	1	0	N/A	
ABTL8	POPTR_0010s18720 plus POPTR_0010s18710	estExt_Genewise1Plus.C_LG_X4616	Va	449	2	1	PtACT55	
ABTL9	POPTR_0010s19360	grail3.0022010901	Va	435	1	0	N/A	
ABTL10	POPTR_0014s10890	eugene3.00140581	Va	423	2	1	N/A	
ABTL11	POPTR_0014s16460	fgenesh4_pm.C_LG_XIV000486	Va	441	2	1	PtACT38	
ABTL12	POPTR_0015s11290	grail3.0005031901	Va	440	3	2 (1 in 5' UTR)	PtACT39	
ABTL13	POPTR_0017s09970	gw1.295.3.1	Va	400	2	1	PtACT40	
ABTL14	POPTR_0019s14700	eugene3.00190987	Va	486	1	0	PtACT44	
AMATL1	POPTR_0003s01420	eugene3.00030070	Va	454	2	1	PtACT52	**A**nthraniloyl-CoA:**M**ethanol **A**cyl**T**ransferase-**L**ike (Wang and De Luca 2005)
AMATL2	POPTR_0013s11650[b]	gw1.41.263.1	Va	427	2	1	PtACT51	v1.1 sequence used for phylogenetic analysis

Table 2.1, Continued:

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
AMATL3	POPTR_0001s00980 plus POPTR_0001s00970	gw1.29.221.1[b]	Va	459	2	1	N/A	manual sequence curation
AMATL4	POPTR_0001s00950	gw1.29.19.1	Va	462	3	2	PtACT50	
HMTL1	POPTR_0001s45150[b,c]	gw1.I.5422.1[b]	Va	465	2	1	PtACT53	Tigloyl-CoA:(-)-13α-**H**ydroxy**M**ultiflorine-*O*-**T**igloyltransferase-**L**ike (Okada et al. 2005); manual sequence curation
HMTL2	POPTR_0001s45160[b,c]	gw1.I.5436.1[b]	Va	465	2	1	N/A	manual sequence curation
HMTL3	POPTR_0001s45180	gw1.I.5447.1	Va	465	2	1	N/A	
HMTL4	POPTR_0001s45190	gw1.I.5452.1	Va	449	2	1	N/A	
HMTL5	POPTR_0001s45200	fgenesh4_pg.C_LG_I003141	Va	345	2	1	N/A	
HMTL6	POPTR_0001s45210	grail3.0051000601	Va	469	2	1	PtACT54	
HMTL7	POPTR_0011s15660	gw1.XI.960.1	Va	465	2	1	N/A	
CHATL1	POPTR_0013s07220	fgenesh4_pm.C_LG_XIII000317	Va	440	2	1	PtACT46	Acetyl-CoA:***C****is*-3-**H**exen-1-ol **A**cetyl**T**ransferase-**L**ike (D'Auria et al. 2002; D'Auria et al. 2007a)
CHATL2	POPTR_0013s07230	eugene3.00130737	Va	456	2	1	PtACT45	
CHATL3	POPTR_0013s07240	eugene3.00130740	Va	459	2	1	PtACT47	
CHATL4	POPTR_0019s01680	eugene3.01170065	Va	452	2	1	PtACT48	
CHATL5	POPTR_0019s01700 plus POPTR_0019s01690	eugene3.01170066	Va	451	2	1	PtACT65	
CHATL6	POPTR_0019s06040	grail3.0065010701	Va	460	2	1	PtACT49	
HCT1	POPTR_0003s18210[b]	eugene3.00031532	Vb	457	3	2	PtACT58	**H**ydroxy**C**innamoyl**T**ransferase; v1.1 sequence used for phylogenetic analysis

Table 2.1, Continued:

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
HCT2	POPTR_0018s11440 plus POPTR_0018s11450	estExt_fgenesh4_pm.C_LG_XVIII0344	Vb	430	2	1	PtACT60	
HCT3	POPTR_0018s11380	estExt_fgenesh4_pg.C_LG_XVIII0910	Vb	440	2	1	PtACT64	
HCT4	POPTR_0018s11370	eugene3.00180947	Vb	440	2	1	PtACT63	
HCT5	POPTR_0005s02820	fgenesh4_pg.C_scaffold_133000007	Vb	444	2	1	PtACT62	
HCT6	POPTR_0001s03440	eugene3.02080010	Vb	431	3	2 (1 in 5' UTR)	PtACT59	
HCT7	POPTR_0005s02810	eugene3.18780002	Vb	443	2	1	PtACT61	
SHTL1	POPTR_0006s17880	fgenesh4_pm.C_scaffold_121000022	Vb	451	1	0	PtACT57	**S**permidine **H**ydroxycinnamoyl**T**ransferase-**L**ike (Grienenberger et al. 2009)
SHTL2	POPTR_0018s11840	fgenesh4_pg.C_LG_XVIII000954	Vb	455	1	0	PtACT56	
n/a	n/a	fgenesh1_pg.C_scaffold_2984000001					PtACT4	obsolete model from v1.0; unanchored short scaffold
n/a	POPTR_0004s09540	eugene3.00040593		237			PtACT8	missing HXXXD domain; partial sequence
n/a	n/a	fgenesh1_pg.C_scaffold_165000045					PtACT15	obsolete model from v1.0; unanchored short scaffold; partial sequence
n/a	n/a	fgenesh1_pg.C_scaffold_2754000001					PtACT17	obsolete model from v1.0; unanchored short scaffold
n/a	n/a	eugene3.16300001					PtACT19	obsolete model from v1.1; unanchored short scaffold; partial sequence
n/a	n/a	fgenesh1_pg.C_scaffold_145000050					PtACT22	obsolete model from v1.0; unanchored short scaffold; partial sequence
n/a	n/a	fgenesh1_pg.C_scaffold_1532000001					PtACT23	obsolete model from v1.0; unanchored short scaffold; partial sequence

Table 2.1, Continued:

Name	Gene Model v2.0	Gene Model v1.1	Clade	Protein Length	Exons	Introns	Previous Name	Notes[e]
n/a	n/a	eugene3.00111166					PtACT25	obsolete model from v1.1; unanchored short scaffold; partial sequence
n/a	n/a	eugene3.00121073					PtACT36	missing HXXXD domain; partial sequence
n/a	POPTR_0660s00200	eugene3.00170007		73			PtACT74	short scaffold; partial sequence
n/a	POPTR_0005s02830	grail3.0133000501		335			PtACT79	missing HXXXD domain; BAHD-like
n/a	n/a	eugene3.00031450					PtACT86	missing HXXXD domain; partial sequence

[a]Only a partial sequence is available due to sequence gaps. [b]Gene model contains structural annotation error. See notes for sequences used in phylogenetic analysis. [c]Model correspondence for these two tandem duplicates here differs from the assignment in Phytozome. [d]Validity of the predicted gene model is in question. The predicted v2.0 model contains two unusual short introns, while the v1.1 model contains a premature stop codon. [e]See Table 2.2 for manually curated sequences.

Table 2.2: Manually Curated *Populus* BAHD Acyltransferase Protein and CDS Sequences

Data provided for sequences noted in Table 2.1.

>CERL1_pmanual [POPTR_0001s32660 eugene3.00012234 both contained extra 2nd intron]
MVSCKEDQNLVYDIKLSSAGPGRITGSDVIHEPNGMDLAMKLPYLKGVYFFNSQACQGLTIMQIKGSMFYWLNDYYTVCGRFQRTEAGRPYMKCNDCGVRIVEARCS
KTVDEWLETRDCSLDNLLIYHSPIGPELFFSPSLYMQVTKFKCGGMSLGISWAHIIGDVYSASECLNSWGQFLAGLKSYGPLKLTKSPTGLEDSKSPSVGTQEPVSL
KQVDPVGDLWVTANNCKMETFSFHLSASQVSQLHSRIWGPSGIAKIPFFESLCAIMWQCIAKAKDGLEPKVVTLCKKDPNNPKDGILSNSQIISSVKADSSVVDADL
QELATLLVDQATEENSQIEEVVEKDNGVFDYIVYGANLTFVDLEETNFYGLEWNGHKPEAVHYSIQGVGDEGAVMVLPWPKDSGTDGNIGRIVVVTLPENEVVKLRF
ELQKNGLMLEDDIDS

>CERL1_CDSmanual
ATGGTTTCTTGCAAGGAGGATCAAAATTTGGTCTATGACATTAAACTATCATCTGCTGGACCGGGTCGTATTACTGGTTCGGACGTGATTCACGAGCCCAATGGCAT
GGACTTGGCCATGAAGCTTCCCTATCTAAAAGGTGTCTACTTTTTTAACAGCCAAGCATGCCAAGGATTGACCATCATGCAAATTAAGGGAAGCATGTTTTACTGGC
TTAACGACTACTATACAGTTTGTGGCCGCTTCCAGCGAACGGAGGCTGGACGGCCATACATGAAATGTAATGACTGTGGTGTGAGGATTGTAGAGGCTCGGTGCAGC
AAGACAGTCGATGAATGGCTAGAAACAAGGGATTGTTCTCTTGACAATCTTCTTATTTACCATTCACCTATTGGTCCTGAATTATTTTTCTCTCCTTCGCTTTACAT
GCAGGTAACCAAGTTCAAATGTGGAGGAATGTCCTTGGGCATTAGCTGGGCCCATATAATTGGAGATGTGTATTCAGCTTCGGAGTGCCTCAACTCCTGGGGCCAAT
TCCTGGCTGGTCTTAAGTCGTATGGGCCTTTGAAGCTCACAAAATCACCCACTGGGCTTGAAGATTCCAAGAGCCCATCTGTGGGTACCCAGGAACCCGTTTCTCTG
AAACAGGTCGACCCGGTTGGTGACCTCTGGGTAACTGCCAATAACTGCAAAATGGAAACTTTTTCATTCCATTTATCTGCTTCGCAAGTATCTCAATTACACTCAAG
AATTTGGGGTCCGAGTGGAATTGCCAAGATCCCATTTTTTGAGTCGCTGTGTGCTATAATGTGGCAATGCATAGCCAAAGCTAAAGATGGGCTTGAGCCTAAAGTTG
TCACCCTCTGCAAGAAAGATCCCAACAACCCTAAAGATGGGATTTTGAGCAACAGTCAAATTATAAGTTCAGTCAAGGCAGATTCCTCAGTTGTGGACGCAGATTTG
CAAGAGCTGGCCACGTTGCTGGTTGATCAAGCCACAGAGGAGAATAGCCAAATTGAAGAGGTGGTGGAAAAGGATAATGGAGTATTTGATTACATTGTGTACGGTGC
AAATTTGACATTTGTGGACTTAGAAGAGACTAATTTTTATGGACTGGAATGGAATGGACATAAACCAGAAGCTGTGCACTATAGCATCCAAGGTGTTGGAGATGAAG
GAGCTGTTATGGTACTTCCATGGCCAAAAGATTCAGGCACGGATGGCAACATTGGAAGGATCGTGGTGGTAACTCTGCCTGAAAATGAGGTGGTGAAGCTCAGATTT
GAGCTGCAAAAAAATGGTCTGATGCTTGAAGATGACATCGATTCCTGA

>CERL4_pmanual [POPTR_0013s03730 missing sequence data]
MVKKNSRVSVHSMLTAVSSQPVGSGKTHPLSVLDHAMGLHTVHVVXXXXKNPFGIFDVDPLRIALSEVLCLYPQVTGRLTRGESGNWLVKCNDAGVRVLRAKVEATM
DEWLRSADSSEEKDLTFWEEIPEEPSTWSPFRIQVNEFEGGGVAFGLSCTHMNADPTSITILFKSWIESHRQEPIEHPPLFSSTTLHHQQVPNTSTKSDNYCATKGN
AETPSVKMVTATFKFSDSAIKKWLDEVHDQCAKATPFELLAALFWTRVAHLKAPKNDNKHSLSICLDFRRLVQPPISLGYFGNALHFSLLTLDEEEMDYGKLGHVVE
LVHRHISDVREEEVWSVVDWFESQKEEGGKYAEPFRMYGPELTCVSMEHIIIGHKSLMFSASFKSDEKPVHVSCHVGNVRGEGLIVVLPSVEEGLARTVMVTLPEEE
MPKLCEDQAIQCLQPTMLIGGR

Table 2.2, Continued:

>CERL4_CDSmanual
ATGGTGAAAAAAAATAGCAGGGTTAGTGTCCATTCAATGTTAACGGCAGTATCCAGTCAGCCAGTCGGGTCGGGTAAAACTCATCCATTATCAGTACTTGATCATGC
AATGGGTCTCCATACAGTACATGTAGTTTTNNNNNNNNNGAAAAACCCATTTGGGATTTTTGACGTAGATCCTTTAAGGATTGCTCTGTCAGAGGTTCTTTGTTTGT
ACCCACAAGTTACGGGTCGGTTGACCCGAGGGGAGTCGGGTAATTGGTTAGTGAAGTGTAATGATGCTGGTGTTAGAGTTCTGAGAGCAAAAGTTGAGGCCACCATG
GATGAATGGCTGAGATCAGCTGATAGTTCAGAGGAGAAAGATTTAACGTTTTGGGAGGAAATTCCCGAGGAACCTAGTACATGGTCACCCTTCCGAATTCAGGTAAA
TGAATTTGAAGGAGGAGGTGTAGCTTTTGGGCTAAGTTGTACACACATGAATGCAGACCCAACTTCCATAACTATACTCTTCAAATCCTGGATTGAGAGTCACCGCC
AGGAGCCCATTGAGCACCCACCCCTGTTCAGCTCAACCACCCTCCATCACCAACAAGTTCCTAATACTAGCACTAAATCAGACAATTACTGTGCAACTAAGGGCAAT
GCAGAAACTCCCTCGGTGAAAATGGTCACAGCCACGTTCAAGTTCTCTGATTCAGCAATCAAGAAATGGCTTGACGAAGTGCATGATCAATGTGCTAAAGCTACTCC
TTTTGAATTGCTAGCTGCACTCTTTTGGACACGTGTTGCACATCTAAAGGCTCCAAAAAAATGACAACAAACACTCCCTCTCAATTTGCTTGGACTTTAGAAGGCTAG
TGCAGCCACCAATTTCTCTTGGTTACTTTGGCAATGCATTGCATTTTTCACTGCTGACACTAGATGAGGAAGAAATGGACTACGGTAAGTTGGGACATGTGGTGGAG
TTGGTGCATCGCCATATTTCAGATGTACGGGAAGAGGAGGTTTGGTCTGTTGTAGATTGGTTTGAATCACAGAAGGAAGAGGGAGGGAAGTATGCAGAACCTTTCAG
AATGTATGGTCCTGAGCTAACTTGTGTCAGCATGGAACACATAATAATAGGGCATAAATCGTTGATGTTCTCAGCAAGTTTCAAGAGTGATGAAAAACCAGTTCACG
TTTCATGTCATGTTGGAAATGTGAGGGGTGAAGGTCTGATTGTGGTGCTGCCTTCAGTAGAAGAAGGGCTTGCAAGGACAGTGATGGTAACCTTGCCAGAGGAGGAG
ATGCCTAAATTATGCGAGGATCAAGCTATCCAGTGTCTACAACCAACAATGCTGATAGGTGGAAGATAA

>CERL5_pmanual [POPTR_0018s01250 with edited exons]
MADITYICKRTVVSTKPVQPGKHCSLSVLDRLMEQNHLRSVYYFRTPGGREPGELTKKLRESLSEMLTCFPIVTGRLLKDPKGHWLIKCNDAGVRMVEARIKGSVED
WLKSVDREKELMLVHWEEMYHKPYFWSTFYVQITEFGEGGLAIGLSCFHLLADPTCATMFVKAWADVTLTGKMLNPPLFHQLPPRRPGRKNPNHEPYMELINCYKPI
ADKTNLVSDTKHATIALAFSDPMVRACMANGQAMNAFDQSSPSPFEALAGLFWVCISKLKGAGDGLIDMSICLDMRNVLHLDNGFFGNCMVYNKVNSKSLKEHKLSD
VAKAIGEVMAKMDNDGITDLIEWLEHNDYQSPPPMNGCELMCASLEAVDPYLAVFEEGFVPIRVSSYVEPVVGAGHVLVLPSPPCEGPLSRTVMVTLPEDEAARLCE
DDLILHFSPTILMGVNN

Table 2.2, Continued:

>CERL5_CDSmanual
ATGGCTGATATCACCTACATTTGCAAACGCACTGTTGTTAGCACAAAACCAGTGCAACCAGGAAAACACTGCTCGCTTTCAGTTTTAGATCGCCTTATGGAACAAAA
CCACCTAAGATCTGTGTACTATTTTCGAACCCCAGGAGGGAGGGAGCCTGGAGAATTAACCAAGAAGCTAAGAGAGTCTCTGTCTGAAATGCTTACATGTTTTCCTA
TAGTGACAGGCAGGCTGTTGAAGGACCCGAAAGGTCATTGGTTGATCAAGTGCAATGATGCTGGCGTAAGAATGGTGGAGGCTAGAATAAAAGGAAGTGTTGAAGAC
TGGTTAAAGAGTGTAGATAGAGAGAAGGAGCTTATGCTTGTTCACTGGGAAGAAATGTACCATAAGCCTTATTTTTGGTCTACCTTCTATGTTCAGATAACTGAATT
TGGCGAAGGTGGACTAGCAATTGGCTTGAGCTGCTTTCACCTGCTAGCTGATCCCACTTGTGCCACCATGTTCGTTAAGGCCTGGGCCGACGTGACACTCACCGGGA
AAATGCTCAACCCTCCTCTTTTCCATCAGCTGCCGCCTCGAAGACCTGGCAGGAAGAATCCCAACCACGAGCCTTACATGGAGTTGATCAATTGCTATAAACCTATT
GCTGATAAAACAAATTTGGTATCTGACACGAAGCATGCGACTATTGCTCTTGCATTTTCGGACCCTATGGTCCGAGCTTGCATGGCAAATGGTCAAGCCATGAACGC
ATTCGATCAGTCTAGCCCGTCACCATTCGAGGCACTGGCCGGGTTATTTTGGGTTTGTATAAGCAAATTGAAAGGAGCAGGAGACGGTCTTATAGACATGTCTATAT
GTTTAGACATGAGAAATGTGCTGCACCTGGATAATGGATTTTTCGGAAACTGCATGGTATATAACAAAGTTAACTCAAAGTCCTTAAAAGAACACAAGTTATCGGAT
GTTGCTAAGGCAATCGGAGAAGTCATGGCAAAAATGGACAATGATGGGATCACTGACTTGATCGAATGGCTTGAACATAATGATTATCAATCTCCTCCTCCGATGAA
TGGCTGTGAACTCATGTGTGCTAGCTTGGAAGCCGTGGACCCCTATTTAGCTGTGTTTGAAGAAGGATTTGTCCCAATTCGTGTGTCCTCCTATGTGGAACCAGTTG
TGGGGGCAGGACATGTTTTGGTCCTTCCATCGCCGCCGTGCGAGGGTCCATTGAGCAGGACAGTCATGGTTACACTCCCGGAAGATGAGGCGGCTAGACTATGTGAA
GATGATCTCATTCTGCATTTCTCTCCAACTATTTTAATGGGGGTGAATAATTAA

>AATL7_pmanual [POPTR_0006s03260 (extra 5’-intron) fgenesh4_pg.C_LG_VI000275]
MEVQILSRKLIAPSSPTPPHLQNLKVSCFDQLAPSIYLPCIFYYPADGENNGKRSKEMEKSLAETLSLFYPLGGRYIKDEFSVECNDMGAEFLEAKVGGFLSQLLER
EERESEMASHLVAPLFQTENSPLVIVQFNMFECGGLAIGISIAHRIADAFTIGTFINAWATACRIGSEKVHCRPSFQLGSLFPPKEMPSSSATAPGTDIKIIRRRFV
FDGHTLSKLKAIARGGPSLLGLTFNMRGKTAMTTPDYSCGNFVNWANAQFMPDDEIKMELHHFVNRVHDAISTTTHDCAKASNSDDIYSMVSSKAREVGEALGEGNV
DTYMFSCWCRFPWYEADFGWGKPSWVSSVDVPTGIVMLMDTKDGDGIEVFLALDESSMLTLQQNLDKTISFTG

>AATL7_CDSmanual
ATGGAGGTTCAAATCTTATCTAGAAAACTGATAGCCCCTTCATCACCAACTCCACCGCACCTTCAAAACTTGAAAGTATCATGTTTTGACCAGCTTGCTCCTTCCAT
TTACTTACCATGCATTTTCTACTATCCAGCCGATGGTGAAAACAACGGAAAACGAAGCAAGGAAATGGAAAAATCATTAGCTGAAACCTTAAGCCTCTTTTACCCAC
TCGGGGGAAGATACATCAAGGACGAATTCTCAGTTGAGTGTAATGACATGGGAGCAGAGTTTTTGGAAGCCAAAGTTGGTGGTTTCTTATCTCAGCTTCTCGAAAGA
GAAGAGCGTGAGTCTGAAATGGCGAGTCATCTGGTTGCACCACTATTCCAAACGGAGAACAGCCCTCTTGTGATAGTTCAATTCAACATGTTTGAATGTGGTGGACT
GGCTATTGGTATATCTATCGCACACAGGATAGCTGATGCATTCACTATAGGTACATTCATCAATGCTTGGGCCACTGCTTGTCGAATTGGGAGTGAAAAGGTTCATT
GTCGTCCAAGCTTCCAATTGGGTTCTCTCTTTCCACCCAAAGAAATGCCTTCGTCCAGTGCCACGGCTCCAGGAACGGACATCAAGATTATCAGGAGAAGGTTCGTG
TTTGATGGCCATACTTTATCAAAACTGAAAGCAATTGCTAGAGGTGGCCCTTCCCTCCTCGGACTTACATTTAACATGCGAGGGAAGACAGCTATGACTACTCCGGA
CTATTCCTGTGGAAACTTTGTTAATTGGGCGAATGCGCAGTTTATGCCAGATGATGAGATCAAGATGGAGCTTCATCACTTCGTGAATCGGGTGCACGACGCCATAA
GCACCACCACCCATGATTGTGCAAAAGCTTCAAATAGTGATGACATTTACTCTATGGTGTCCAGTAAGGCGAGGGAGGTGGGTGAAGCACTTGGGGAAGGCAATGTA
GATACCTATATGTTCAGTTGCTGGTGTCGGTTCCCATGGTATGAAGCAGATTTTGGATGGGGAAAACCCTCTTGGGTGAGTAGCGTGGACGTACCAACTGGCATTGT
TATGCTTATGGACACTAAAGATGGTGATGGAATTGAAGTATTTCTAGCCTTGGATGAAAGCAGCATGCTTACTCTTCAGCAAAATCTGGACAAAACTATTTCGTTCA
CTGGCTAG

Table 2.2, Continued:

>AATL12_pmanual [POPTR_0010s06380(extra intron) e_gw1.X.4991.1 (N's and introns)]
MKIEIEVISNEIIKPSSPTPDHLRHYQLSFLDQISPPTYNPLLLFYPADGDVKINNIEKPNQLKQSLSEVLNLYYPLAGRIKDNLFVECNDEGIPFFQAEVKCRLPQ
VVENPEPSELNKLIPFALDDAEELPLGIQYNIFECGGIVIGLCISHKVGDASSLFTFIKYWAATARGEADHISRPEFISATLFPPINISGFKPATGITKEDVVTKRF
VFRSSSIELLKEKCSPASGSLENQRPPSRVEALSVFIWQRFTAATKVESRPERIYSMVHAVNLRSRMEPPLPEYSFGNYYRIAFTIPSIDTGEENYNLVSQIRDSIG
KVDKEYVKKLQKGSEHLGFMKEQAARFLRGEVVTLNFTSLCRFPLYEADFGWAKPIWVGSPSLTFKNLVVFMDTASGDGIEALVHLKEEDMAKFEEDEELLQYIVPT
KC

>AATL12_CDSmanual
ATGAAGATTGAAATTGAAGTAATCTCCAACGAGATCATCAAGCCATCTTCTCCAACCCCAGATCACCTTCGCCATTACCAGCTCTCCTTTCTTGATCAAATCTCTCC
CCCAACCTATAACCCTTTGCTCCTCTTCTATCCAGCAGACGGTGATGTCAAGATCAACAACATAGAGAAACCTAACCAGCTCAAGCAATCCTTGTCTGAGGTCTTAA
ACCTTTACTATCCCTTAGCCGGACGTATTAAGGACAACCTTTTCGTAGAGTGCAACGATGAGGGCATTCCATTTTTCCAGGCAGAAGTCAAGTGCCGACTTCCACAA
GTTGTTGAGAATCCAGAACCTAGTGAACTCAACAAGTTGATCCCATTTGCACTAGATGATGCTGAAGAACTGCCTCTAGGCATCCAGTACAACATCTTTGAGTGTGG
TGGAATTGTTATTGGTCTGTGCATCTCACACAAAGTTGGAGATGCATCATCACTGTTCACGTTTATCAAATATTGGGCTGCCACTGCTCGTGGAGAAGCAGATCACA
TATCAAGACCAGAGTTTATCTCTGCAACTCTCTTCCCACCTATCAACATATCAGGGTTTAAACCAGCCACTGGTATCACTAAAGAAGATGTTGTGACAAAAAGGTTC
GTGTTCCGCTCCTCTTCAATAGAGCTGCTAAAAGAAAAATGCAGTCCTGCAAGTGGAAGCTTGGAAAATCAGCGACCACCATCACGTGTTGAGGCCTTGTCAGTATT
CATATGGCAACGCTTCACAGCTGCCACTAAAGTAGAATCAAGACCTGAAAGAATTTACTCCATGGTTCATGCAGTGAACCTGCGCTCGAGGATGGAACCTCCGCTCC
CAGAATACTCCTTCGGAAACTACTATCGGATTGCATTCACAATTCCATCCATTGATACTGGCGAGGAAAACTACAATCTTGTTAGTCAGATCAGAGACTCAATTGGT
AAAGTTGACAAAGAATACGTGAAGAAACTTCAAAAGGGCAGTGAGCACTTGGGCTTCATGAAAGAACAAGCTGCAAGATTTCTCAGAGGTGAGGTGGTTACTTTGAA
CTTCACAAGCTTGTGCAGGTTTCCTTTGTATGAAGCTGATTTTGGGTGGGCGAAACCTATATGGGTAGGCTCTCCAAGTCTCACCTTCAAGAACCTAGTTGTTTTCA
TGGACACTGCATCAGGTGATGGAATAGAAGCACTTGTGCACTTGAAGGAGGAAGACATGGCCAAATTCGAAGAGGATGAGGAGTTGCTTCAGTATATTGTACCAACC
AAATGTTAA

>AMATL3_pmanual [POPTR_0001s00980+POPTR_0001s00970 edited from gw1.29.221.1]
MAPSSSLVSFKVRHRDPELVVPAKPVPYEQKQLSDVDDQEALRYQIPFIMFYDSNSNPCMEGEDQVKIIRAALAEALVYYYPLAGRLKEGPDGKLLVDCTGEGVLFL
EADADTTLELLEDTIQPPCPYLDQLLYNVPGSTGIVGCPLLLIQVTRLMCGGFVFAIRWSHIIADAVGMSKFLNTIAEMVPGATKPSFLPVWQRELLNARDPPRATY
EHHEFDEVNDTDFGTMNDDTIIVHKSFFFGPREMSSIRKHLPPHLRASSSFLVLTACLWKCRTIATQLDPNEIVRVSYMVTASGKEGLKLPAGYYGNAFTFPVALSE
AGLLCKNPLEYGLELVKEIKNRLSEEYTRSAIDLLVIKGKKQYRTVRDFVIADTTRVPFGEIDLGWGKPVYGGPAGAIKDVSFFAKFKNGKGEDGIVVQVSLPWQIM
ERFQKELAKMAGNSSNDQCCRNATEVARSKL

Table 2.2, Continued:

>AMATL3_CDSmanual
ATGGCACCGTCTTCCTCTCTGGTATCATTCAAGGTGAGACATAGGGATCCTGAACTGGTTGTGCCGGCAAAACCAGTCCCATATGAGCAAAAGCAGCTGTCAGATGT
AGATGACCAAGAAGCCCTGCGTTATCAAATTCCATTCATCATGTTCTATGACAGTAATAGCAATCCTTGTATGGAAGGAGAAGACCAGGTGAAAATCATTCGAGCTG
CCCTAGCAGAAGCACTAGTGTATTACTATCCACTTGCTGGCAGGCTTAAAGAAGGGCCTGATGGCAAGCTTCTTGTGGATTGCACAGGTGAAGGTGTCTTGTTTCTT
GAGGCTGATGCTGACACCACGCTTGAACTACTCGAGGACACTATTCAACCACCGTGCCCATATCTTGATCAGCTTCTTTATAATGTTCCTGGCTCTACAGGAATCGT
AGGATGCCCTTTGTTGCTGATCCAGGTGACGCGATTGATGTGTGGGGGATTTGTCTTTGCAATACGTTGGAGCCACATCATTGCTGATGCAGTCGGTATGTCCAAGT
TCTTGAACACAATTGCAGAGATGGTACCAGGTGCTACTAAACCATCCTTTCTCCCTGTGTGGCAAAGAGAACTATTAAATGCAAGAGACCCTCCACGGGCGACTTAT
GAACATCACGAATTCGATGAGGTCAATGACACCGATTTTGGCACTATGAATGATGATACAATTATTGTTCACAAGTCTTTCTTTTTCGGCCCCAGAGAGATGAGTTC
GATTCGGAAACATCTTCCACCACACCTTCGTGCGAGTTCTTCGTTTCTAGTATTAACTGCTTGTTTATGGAAATGCAGAACGATCGCAACGCAACTTGATCCTAATG
AGATCGTTCGCGTATCGTACATGGTCACTGCCAGTGGCAAGGAAGGCTTAAAACTGCCTGCTGGTTACTATGGGAATGCATTCACCTTCCCGGTCGCCCTTTCAGAG
GCTGGATTGCTATGTAAAAATCCACTAGAGTATGGACTAGAGTTAGTGAAGGAGATAAAGAACAGATTGAGTGAAGAGTACACAAGGTCTGCTATAGACCTTTTGGT
AATTAAGGGAAAGAAACAATATAGGACAGTCAGAGATTTTGTAATTGCTGATACAACGCGTGTGCCGTTTGGAGAGATTGATTTAGGCTGGGGAAAGCCAGTATATG
GTGGTCCTGCAGGAGCCATTAAAGATGTTAGTTTCTTTGCTAAGTTTAAGAATGGTAAAGGAGAGGATGGGATTGTAGTACAAGTTTCATTGCCATGGCAAATCATG
GAAAGGTTTCAAAAGGAGTTAGCAAAGATGGCAGGGAACTCTTCGAATGATCAGTGCTGCAGAAATGCCACAGAAGTCGCACGTTCCAAGCTCTAG

>HMTL1_pmanual [POPTR_0001s45150.1 edited from v1.1 gw1.1.5422.1]
MATPTSLSFAVRRCEPELVAPAKATPHEFRQLSDIDRQLYLQFQSPHYNLYAHNPSMQGKDPVKVIKEAIAQALVYYYPFAGRIRQGPDNKLIVDCTGEGVLFIEAD
ADATVEQFGDPIPSPFPCFQELLYNVPGSEGILNTPLLIFQVTRLKCGGFVLGLRLNHPMTDAFGMLQVLNAIGEIARGAQAPSILPVWRRELLCARNPPRVTCRHN
EYGNDAPVAVDPTAKVPEFHGQVHAVAHRSFVLNRKELSNIRRWIPSHLHPCSNFEVITACLWRCYAIASQANPNEEMRMQMLVNARSKFNPPLPKGYYGNVLALPA
AVTNARKLCLNSLGYALEMIRNAKNRITEEYMRSLADLMEITKGQPIGLQSYVVSDLTGFGFDQVDYGWGNTIYTGPPKAMPDEISMAGTYFLPYRFKNGERGVMLL
VSLRAPVMERFAILLEELARHDPERSQEQQEMIPSSL

Table 2.2, Continued:

>HMTL1_CDSmanual
ATGGCAACACCAACTTCCTTATCGTTCGCCGTCCGAAGGTGCGAACCAGAATTGGTTGCGCCAGCTAAGGCCACACCTCATGAATTCAGACAGCTTTCTGATATTGA
TCGCCAACTATACCTCCAATTTCAATCACCACATTACAACTTGTATGCACACAATCCATCGATGCAAGGGAAAGATCCTGTGAAGGTAATAAAGGAGGCAATTGCGC
AGGCACTTGTGTATTATTACCCTTTTGCTGGTAGGATTAGACAAGGGCCAGACAATAAGCTTATAGTTGATTGTACTGGTGAGGGTGTCTTGTTCATCGAAGCCGAT
GCCGATGCCACGGTGGAGCAGTTTGGTGATCCAATTCCATCTCCATTCCCATGCTTTCAGGAACTTCTTTACAACGTCCCAGGATCAGAAGGGATCCTCAATACCCC
ATTATTGATTTTTCAGGTGACACGCTTGAAGTGTGGTGGTTTTGTACTTGGGCTCCGTCTTAATCACCCAATGACTGATGCATTCGGCATGCTTCAGGTATTGAATG
CCATAGGTGAGATTGCACGAGGTGCTCAAGCCCCTTCAATTCTACCTGTGTGGCGAAGGGAACTCCTCTGTGCTAGGAATCCGCCACGAGTTACTTGCAGACACAAT
GAATATGGTAATGATGCTCCTGTTGCTGTTGATCCTACAGCCAAGGTGCCTGAATTCCACGGCCAGGTTCACGCTGTAGCCCACCGTAGTTTTGTTCTCAACCGCAA
GGAATTATCCAACATTCGTAGATGGATTCCTTCTCATTTACACCCATGTTCAAATTTTGAGGTAATAACTGCATGCTTATGGAGATGCTATGCCATAGCATCTCAAG
CTAACCCTAATGAGGAGATGCGCATGCAAATGCTTGTCAACGCACGTTCCAAATTTAACCCTCCATTACCGAAAGGATATTATGGTAACGTGCTAGCTTTGCCAGCA
GCTGTAACAAATGCTAGGAAGCTTTGCTTAAACTCTTTAGGGTATGCATTGGAAATGATAAGAAATGCCAAGAATAGAATAACTGAGGAGTACATGAGATCATTGGC
TGATCTAATGGAGATAACCAAAGGGCAGCCTATAGGGTTACAATCATATGTCGTGTCAGACTTAACAGGTTTTGGGTTCGATCAGGTGGACTATGGATGGGGCAACA
CAATTTATACTGGGCCACCCAAGGCTATGCCTGATGAAATTTCTATGGCAGGAACCTATTTCCTGCCGTATCGATTCAAGAACGGAGAGCGTGGGGTTATGCTTTTG
GTTTCCTTACGTGCACCAGTTATGGAGAGATTTGCAATACTATTAGAGGAATTGGCAAGGCATGACCCAGAAAGAAGCCAAGAACAACAAGAAATGATACCAAGCTC
CCTATAA

>HMTL2_pmanual [POPTR_0001s45160 edited from v1.1 gw1.I.5436.1]
MATPPSLSFAVRRCEPELIAPAKATPHEFRQLSDIDRQLYLQFQSPHYNLYAHNPSMQGKDPVKVIKEAIAQALVYYYPFAGRIRQGPDNKLIVDCTGEGVLFIEAD
ADATVEQFGDPIPSPFPCFQELLYNVPGSEGILNTPLLLFQVTRLKCGGFVLGFRFNHPMTDGLGMLQLLNAIGEMARGAQAPSILPVWQRELLCARNPPRVTCRHN
EYGNDAPVAVDPTAKVPEFRGEVHAVAHRSFVLNRKELSNIRRWIPSHLHPCSNFEVISACLWRCYAIASQANPNEEMRMQMLVNARSKFNPPLPKGYYGNVLALPA
AVTNARKLCLNSLGYALEMIRNAKNRITEEYMRSLADLMEITKGQPIGLQSYVVSDLTSIGFDQVDYGWGNTIYTGPPKAMPDEISIAGTYFLPYRFKNGERGVMLL
VSLRAPVMERFAILLEELARHDPERSQEQQEMIPSSL

Table 2.2, Continued:

>HMTL2_CDSmanual
ATGGCAACACCACCTTCCTTATCGTTCGCCGTCCGAAGGTGCGAACCAGAATTGATTGCTCCAGCTAAGGCCACACCTCATGAATTCAGACAGCTTTCTGATATTGA
TCGACAACTATACCTCCAATTTCAATCACCACATTACAACTTGTATGCACACAATCCATCGATGCAAGGGAAAGATCCTGTGAAGGTAATAAAGGAGGCAATTGCGC
AGGCACTTGTGTATTATTACCCTTTTGCTGGTAGGATTAGACAAGGGCCAGACAATAAGCTTATAGTTGATTGTACTGGTGAGGGTGTCTTGTTCATCGAAGCCGAT
GCCGATGCCACGGTCGAGCAGTTTGGTGATCCAATTCCATCTCCATTCCCATGTTTTCAGGAACTTCTTTACAACGTCCCAGGATCAGAAGGGATCCTCAATACCCC
ATTATTGCTTTTTCAGGTGACACGCTTGAAGTGTGGTGGTTTTGTACTTGGGTTCCGTTTTAATCACCCAATGACCGATGGACTCGGCATGCTTCAGTTATTGAATG
CCATAGGTGAGATGGCACGAGGTGCTCAAGCCCCTTCAATTCTACCTGTGTGGCAAAGGGAACTCCTCTGTGCTAGGAATCCGCCACGAGTTACATGCAGACACAAT
GAATATGGTAATGATGCTCCTGTTGCTGTTGATCCTACAGCCAAGGTGCCTGAATTCCGCGGCGAGGTTCACGCTGTAGCCCACCGTAGTTTTGTTCTTAACCGCAA
GGAATTATCCAACATTCGTAGATGGATTCCTTCTCATTTACACCCATGTTCAAATTTTGAGGTAATAAGTGCATGCTTATGGAGATGCTATGCCATAGCATCTCAAG
CTAACCCTAATGAGGAGATGCGCATGCAAATGCTTGTCAACGCACGTTCCAAATTTAACCCTCCATTACCGAAAGGATATTATGGTAACGTGCTAGCTTTGCCAGCA
GCTGTAACAAATGCTAGGAAGCTTTGCTTAAACTCTTTAGGGTATGCATTGGAAATGATAAGAAATGCCAAGAATAGAATAACTGAGGAGTACATGAGATCATTGGC
TGATCTAATGGAGATAACCAAAGGGCAGCCTATAGGGTTACAATCATATGTCGTGTCAGACTTAACAAGTATTGGGTTCGATCAGGTGGACTATGGATGGGGCAACA
CAATTTACACTGGGCCACCCAAGGCCATGCCTGATGAAATTTCTATTGCAGGAACCTATTTCCTGCCGTATCGATTCAAGAACGGAGAGCGTGGGGTTATGCTTTTG
GTTTCCTTACGTGCACCAGTTATGGAGAGATTTGCAATACTATTAGAGGAATTGGCAAGGCATGACCCAGAAAGAAGCCAAGAACAACAAGAAATGATACCAAGCTC
CCTATAA

Table 2.3: Biochemically Characterized BAHD Acyltransferases Included in the Phylogenetic Analysis

All biochemically characterized BAHD proteins included in our analysis are listed by clade and by their order of appearance (from top to bottom) in the detailed phylogenies. Proteins with no listed reference were reviewed by D'Auria (2006).

Name in Phylogeny	Clade	Taxon	Genbank ID	Reference
GmIF7MaT	Ia	*Glycine max*	BAF73620	(Suzuki et al. 2007)
GM7MAT	Ia	*Glycine max*	ABY59019	(Dhaubhadel et al. 2008)
MtMAT3	Ia	*Medicago truncatula*	ABY91221	(Yu et al. 2008)
MtMAT1	Ia	*Medicago truncatula*	ABY91220	(Yu et al. 2008)
MtMAT2	Ia	*Medicago truncatula*	ABY91222	(Yu et al. 2008)
At5MaT	Ia	*Arabidopsis thaliana*	NP_189600	(D'Auria et al. 2007b; Luo et al. 2007)
Sc3MaT	Ia	*Pericallis cruenta*	AAO38058	
Dm3MAT1	Ia	*Chrysanthemum* x *morifolium*	AAQ63615	
Dm3MAT2	Ia	*Chrysanthemum* x *morifolium*	AAQ63616	
Dv3MAT	Ia	*Dahlia variabilis*	AAO12206	
Gt5AT	Ia	*Gentiana triflora*	BAA74428	
Pf3AT	Ia	*Perilla frutescens*	BAA93475	
Ss5MaT1	Ia	*Salvia splendens*	AAL50566	(Suzuki et al. 2003)
Pf5MaT	Ia	*Perilla frutescens*	AAL50565	
NtMAT1	Ia	*Nicotiana tabacum*	BAD93691	
Lp3MAT1	Ia	*Lamium purpureum*	AAS77404	
Vh3MAT1	Ia	*Verbena* x *hybrida*	AAS77403	
At3AT1	Ia	*Arabidopsis thaliana*	NP_171890	(Luo et al. 2007)
At3AT2	Ia	*Arabidopsis thaliana*	NP_171849	(Luo et al. 2007)
ZmGlossy2	II	*Zea mays*	CAA61258	(Tacke et al. 1995)

Table 2.3, Continued:

Name in Phylogeny	Clade	Taxon	Genbank ID	Reference
AtCER2	II	*Arabidopsis thaliana*	AAM64817	(Negruk et al. 1996; Xia et al. 1996)
CaPun1	IIIa	*Capsicum annuum*	AAV66311	
CrDAT	IIIa	*Catharanthus roseus*	AAC99311	
CrMAT	IIIa	*Catharanthus roseus*	AAO13736	
CmAAT4	IIIa	*Cucumis melo*	AAW51126	
RhAAT1	IIIa	*Rosa* x *hybrida*	AAW31948	
FaSAAT	IIIa	*Fragaria* x *ananassa*	AAG13130	
FvVAAT	IIIa	*Fragaria vesca*	CAC09062	
Ss5MaT2	IIIa	*Salvia splendens*	AAR26385	
CbBEAT	IIIa	*Clarkia breweri*	AAC18062	
PsSalAT	IIIa	*Papaver somniferum*	AAK73661	
RsVISY	IIIa	*Rauvolfia serpentina*	CAD89104	(Bayer et al. 2004)
PhCFAT	IIIb	*Petunia x hybrida*	ABG75942	(Dexter et al. 2007)
HvACT	IV	*Hordeum vulgare*	AAO73071	
AtASFT	Va	*Arabidopsis thaliana*	Q94CD1	(Molina et al. 2009)
TcDBNTBT	Va	*Taxus cuspidata*	AAM75818	
TcDBAT	Va	*Taxus cuspidata*	AAF27621	
TcBAPT	Va	*Taxus cuspidata*	AAL92459	
TcDBBT	Va	*Taxus cuspidata*	Q9FPW3	
TcTAT	Va	*Taxus cuspidata*	AAF34254	
AtSCT	Va	*Arabidopsis thaliana*	Q8VZU3	(Luo et al. 2009)
AtSDT	Va	*Arabidopsis thaliana*	NP_179932	(Luo et al. 2009)
Ih3AT1	Va	*Iris hollandica*	BAE72676	(Yoshihara et al. 2006)
MsAAT	Va	*Musa sapientum*	CAC09063	

Table 2.3, Continued:

Name in Phylogeny	Clade	Taxon	Genbank ID	Reference
VlAMAT	Va	*Vitis labrusca*	AAW22989	(Wang and De Luca 2005)
MdAAT2	Va	*Malus domestica*	AAS79797	(Li et al. 2006)
MdAAT1	Va	*Malus pumila*	AAU14879	(Souleyre et al. 2005)
AtCHAT	Va	*Arabidopsis thaliana*	AAN09797	(D'Auria et al. 2002; D'Auria et al. 2007a)
CmAAT2	Va	*Cucumis melo*	AAL77060	(El-Sharkawy et al. 2005)
CmAAT1	Va	*Cucumis melo*	CAA94432	(El-Sharkawy et al. 2005)
CmAAT3	Va	*Cucumis melo*	AAW51125	(El-Sharkawy et al. 2005)
NtBEBT	Va	*Nicotiana tabacum*	AAN09798	(D'Auria et al. 2002)
PhBPBT	Va	*Petunia* x *hybrida*	AAU06226	(Boatright et al. 2004)
CbBEBT	Va	*Clarkia breweri*	AAN09796	(D'Auria et al. 2002)
LaHMT/HLT	Va	*Lupinus albus*	BAD89275	(Okada et al. 2005)
AtSHT	Vb	*Arabidopsis thaliana*	816424	(Grienenberger et al. 2009)
TpHCT2	Vb	*Trifolium pratense*	BB926056	(Sullivan 2009)
DcHCBT	Vb	*Dianthus caryophyllus*	CAB06430	
AsHHT1	Vb	*Avena sativa*	BAC78633	
PrHCT	Vb	*Pinus radiata*	ABO52899	(Wagner et al. 2007)
NtHCT	Vb	*Nicotiana tabacum*	CAD47830	
TpHCT1B	Vb	*Trifolium pratense*	BB903226	(Sullivan 2009)
TpHCT1A	Vb	*Trifolium pratense*	BB911266	(Sullivan 2009)
AtHCT	Vb	*Arabidopsis thaliana*	NP_199704	(Hoffman et al. 2005)
NtHQT	Vb	*Nicotiana tabacum*	CAE46932	
SlHQT	Vb	*Solanum lycopersicum*	CAE46933	(Niggeweg et al. 2004)
CcsHQT	Vb	*Cynara cardunculus var. scolymus*	DQ915589	(Comino et al. 2009)

Table 2.3, Continued:

Name in Phylogeny	Clade	Taxon	Genbank ID	Reference
CcaHQT	Vb	*Cynara cardunculus var. altilis*	DQ915590	(Comino et al. 2009)
SsHCT	Vb	*Solenostemon scutellarioides*	CAK55166	(Berger et al. 2006)

Table 2.4: Summary of Putative BAHD Acyltransferases in *Arabidopsis thaliana*, *Medicago truncatula*, *Oryza sativa*, and *Vitis vinifera* Genomes

BAHD acyltransferase loci are listed alphabetically by genus, then by clade, then by locus number. Databases used were TAIR9 for *Arabidopsis* (Swarbreck et al. 2008), MtDB v2.0 for *Medicago* (Retzel et al. 2008), Rice Genome Annotation Database releases 5 and 6.1 for *Oryza* (Ouyang et al. 2007; Yuan et al. 2005), and the Genoscope 8X browser for *Vitis* (Jaillon and et 2007).

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Arabidopsis thaliana						
Arabidopsis	At1g03495	Ia	465	1	0	included as At3AT2 (Luo et al. 2007)
Arabidopsis	At1g03940	Ia	469	1	0	included as At3AT1 (Luo et al. 2007)
Arabidopsis	At3g29590	Ia	449	1	0	included as At5MaT (D'Auria et al. 2007b; Luo et al. 2007)
Arabidopsis	At3g29635	Ia	458	1	0	
Arabidopsis	At3g29670	Ia	451	1	0	
Arabidopsis	At3g29680	Ia	451	1	0	
Arabidopsis	At5g39050	Ia	469	1	0	
Arabidopsis	At5g39080	Ia	463	1	0	
Arabidopsis	At5g39090	Ia	448	1	0	
Arabidopsis	At5g61160	Ia	452	1	0	
Arabidopsis	At2g39980	Ib	482	1	0	
Arabidopsis	At3g50270	Ib	450	1	0	
Arabidopsis	At3g50280	Ib	443	1	0	
Arabidopsis	At3g50300	Ib	448	1	0	
Arabidopsis	At5g01210	Ib	475	1	0	
Arabidopsis	At5g07850	Ib	456	2	1	
Arabidopsis	At5g07860	Ib	454	2	1	

Table 2.4, Continued:

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Arabidopsis thaliana						
Arabidopsis	At5g07870	Ib	464	2	1	
Arabidopsis	At5g23940	Ib	484	2	1	
Arabidopsis	At5g38130	Ib	462	1	0	
Arabidopsis	At5g42830	Ib	450	2	1	
Arabidopsis	At5g67150	Ib	448	1	0	
Arabidopsis	At3g23840	II	420	2	1	
Arabidopsis	At4g24510	II	421	2	1	included as AtCER2 (Negruk et al. 1996; Xia et al. 1996)
Arabidopsis	At4g29250	II	460	2	1	not included in Yu et al. (2009)
Arabidopsis	At1g24420	IIIa	436	1	0	
Arabidopsis	At1g24430	IIIa	435	1	0	
Arabidopsis	At3g26040	IIIa	442	1	0	
Arabidopsis	At3g30280	IIIa	443	1	0	
Arabidopsis	At4g15390	IIIa	446	1	0	
Arabidopsis	At4g15400	IIIa	435	1	0	
Arabidopsis	At5g23970	IIIa	428	1	0	
Arabidopsis	At5g47950	IIIa	426	1	0	
Arabidopsis	At5g47980	IIIa	443	1	0	
Arabidopsis	At1g31490	IIIb	444	1	0	
Arabidopsis	At1g32910	IIIb	464	2	1	
Arabidopsis	At1g78990	IIIb	455	2	1	
Arabidopsis	At5g16410	IIIb	480	2	1	
Arabidopsis	At1g03390	Va	461	1	0	
Arabidopsis	At1g27620	Va	442	2	1	
Arabidopsis	At1g28680	Va	451	3	2	

Table 2.4, Continued:

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Arabidopsis thaliana						
Arabidopsis	At2g23510	Va	451	3	2	included as AtSDT (Luo et al. 2009)
Arabidopsis	At2g25150	Va	461	3	2	included as AtSCT (Luo et al. 2009)
Arabidopsis	At2g40230	Va	433	1	0	
Arabidopsis	At3g03480	Va	454	2	1	included as AtCHAT (D'Auria et al. 2002)
Arabidopsis	At3g48720	Va	430	2	1	
Arabidopsis	At3g62160	Va	428	2	1	
Arabidopsis	At4g31910	Va	458	3	2	
Arabidopsis	At5g07080	Va	450	2	1	
Arabidopsis	At5g17540	Va	461	2	1	
Arabidopsis	At5g41040	Va	457	4	3	included as AtASFT (Molina et al. 2009)
Arabidopsis	At5g63560	Va	426	2	1	
Arabidopsis	At2g19070	Vb	451	2	1	included as AtSHT (Grienenberger et al. 2009)
Arabidopsis	At5g48930	Vb	433	2	1	included as AtHCT (Hoffman et al. 2005)
Arabidopsis	At5g57840	Vb	443	4	3	
Arabidopsis	At1g65445	pseudo	166	2	1	partial (Yu et al. 2009)
Arabidopsis	At1g65450	BAHD-like	286	1	0	partial (Yu et al. 2009); insufficient match to DFGWG
Arabidopsis	At3g29636	pseudo	199	4	3	partial (Yu et al. 2009)
Arabidopsis	At3g29690	pseudo	180	2	1	partial (Yu et al. 2009)
Arabidopsis	At3g47170	BAHD-like	468	2	1	insufficient match to DFGWG
Arabidopsis	At4g13840	BAHD-like	428	2	1	insufficient match to DFGWG
Arabidopsis	At5g67160	BAHD-like	434	1	0	AtEPS1 (Zheng et al. 2009), missing HXXXD domain

Table 2.4, Continued:

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Medicago truncatula						
Medicago	AC125389_1	Ia	477	1	0	
Medicago	AC146549_1	Ia	476	1	0	included as MtMAT2 (Yu et al. 2008)
Medicago	AC146549_16	Ia	478	1	0	
Medicago	AC146566_6	Ia	468	1	0	
Medicago	AC171168_6	Ia	468	1	0	
Medicago	AC122728_29	Ib	294	3	2	unusual length
Medicago	AC122728_6	Ib	457	1	0	
Medicago	AC122728_7	Ib	457	1	0	
Medicago	AC140916_25	Ib	434	1	0	
Medicago	AC153120_12	Ib	1159	5	4	unusual length; unusual number of introns
Medicago	AC153120_23	Ib	442	1	0	
Medicago	AC153120_5	Ib	443	1	0	
Medicago	AC153120_6	Ib	436	1	0	
Medicago	AC159185_15	Ib	443	1	0	
Medicago	AC174147_11	Ib	442	1	0	
Medicago	AC174329_42	Ib	443	1	0	
Medicago	AC203224_26	Ib	490	1	0	
Medicago	CT961056_4	Ib	473	1	0	
Medicago	CU179697_2	Ib	433	2	1	
Medicago	CU179697_3	Ib	456	1	0	
Medicago	CU179697_5	Ib	455	1	0	
Medicago	CU179697_7	Ib	500	2	1	
Medicago	CU179697_9	Ib	449	1	0	
Medicago	CU326389_10	Ib	489	1	0	
Medicago	AC144617_10	II	448	2	1	

Table 2.4, Continued:

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Medicago truncatula						
Medicago	AC155282_42	II	442	2	1	
Medicago	CU137640_9	II	463	3	2	
Medicago	AC123572_1	IIIa	416	1	0	
Medicago	AC202368_14	IIIa	425	2	1	
Medicago	CT573055_10	IIIa	439	1	0	
Medicago	CT573055_3	IIIa	440	1	0	
Medicago	CU137654_16	IIIb	473	2	1	
Medicago	CU137654_6	IIIb	471	2	1	
Medicago	CU137654_7	IIIb	471	2	1	
Medicago	AC130805_25	Va	465	1	0	
Medicago	AC148345_15	Va	462	2	1	
Medicago	AC148345_20	Va	462	2	1	
Medicago	AC148345_27	Va	459	2	1	likely same locus as AC171498_10; excluded
Medicago	AC148345_33	Va	445	2	1	
Medicago	AC148345_36	Va	174	1	0	unusual length
Medicago	AC148345_40	Va	345	1	0	
Medicago	AC148345_42	Va	174	1	0	unusual length; likely same locus as AC171498_24; excluded
Medicago	AC150566_3	Va	465	1	0	likely same locus as AC130805_25; excluded
Medicago	AC171498_10	Va	459	2	1	
Medicago	AC171498_24	Va	459	2	1	
Medicago	CT954252_19	Va	435	2	1	
Medicago	CU062659_17	Va	411	3	2	

Table 2.4, Continued:

Taxon	Gene Model		Clade	Protein Length	Exons	Introns	Notes
Medicago truncatula							
Medicago	AC121233_52		Vb	420	2	1	likely same locus as AC160096_30; excluded
Medicago	AC122170_18		Vb	456	1	0	
Medicago	AC125477_25		Vb	457	1	0	likely same locus as AC135313_21; excluded
Medicago	AC135313_21		Vb	457	1	0	
Medicago	AC147000_9		Vb	457	1	0	
Medicago	AC148816_10		Vb	466	1	0	
Medicago	AC148816_15		Vb	466	1	0	
Medicago	AC160096_30		Vb	420	2	1	
Medicago	CT954272_38		Vb	456	1	0	likely same locus as AC122170_18; excluded
Oryza sativa	**Previous Model[a]**	**New Model[b]**					
Oryza	12001.m08401	13101.m02017	Ia	623	3	2	unusual length
Oryza	12002.m07961	13102.m03071	Ia	489	1	0	
Oryza	12002.m07964	13102.m03074	Ia	506	1	0	
Oryza	12002.m07968	13102.m03078	Ia	396	2	1	
Oryza	12002.m07974	13102.m03084	Ia	475	1	0	
Oryza	12002.m07978	13102.m03089	Ia	468	1	0	
Oryza	12002.m07985	13102.m03096	Ia	461	1	0	
Oryza	12002.m07991	13102.m03103	Ia	335	2	1	
Oryza	12002.m10770	13102.m06635	Ia	454	1	0	
Oryza	12004.m10062	13104.m05276	Ia	467	1	0	
Oryza	12004.m10347	13104.m05644	Ia	445	2	1	
Oryza	12004.m10348	13104.m05646	Ia	413	2	1	
Oryza	12006.m05257	13106.m00509	Ia	491	2	1	

Table 2.4, Continued:

Taxon	Gene Model		Clade	Protein Length	Exons	Introns	Notes
Oryza sativa	Previous Model[a]	New Model[b]					
Oryza	12006.m05259	13106.m00510	Ia	490	2	1	
Oryza	12006.m05260	13106.m00511	Ia	449	4	3	
Oryza	12006.m05261	13106.m00512	Ia	486	2	1	
Oryza	12008.m04910	13108.m00793	Ia	483	2	1	
Oryza	12008.m04911	13108.m00794	Ia	470	2	1	
Oryza	12010.m05570	13110.m02237	Ia	457	2	1	
Oryza	12001.m12464	13101.m06813	Ib	484	1	0	
Oryza	12004.m08162	13104.m02902	Ib	446	2	1	
Oryza	12005.m07945	13105.m03925	Ib	481	1	0	
Oryza	12006.m04875	13106.m00037	Ib	462	1	0	
Oryza	12007.m04963	13107.m00442	Ib	476	1	0	
Oryza	12008.m04343	13108.m00128	Ib	480	2	1	
Oryza	12008.m04344	13108.m00129	Ib	534	1	0	
Oryza	12008.m04346	13108.m00131	Ib	465	1	0	
Oryza	12008.m04350	13108.m00135	Ib	485	2	1	
Oryza	12008.m26422	13108.m00136	Ib	469	1	0	
Oryza	12008.m08456	13108.m04880	Ib	446	2	1	
Oryza	12011.m05536	13111.m01428	Ib	479	1	0	
Oryza	12004.m10111	13104.m05340	II	438	2	1	
Oryza	12004.m10111	13104.m11949	II	438	2	1	likely same locus as 13104.m05340; excluded
Oryza	12003.m10307	13103.m05828	II	473	2	1	
Oryza	12001.m08165	13101.m01737	IIIb	461	2	1	
Oryza	12001.m08169	13101.m01742	IIIb	461	2	1	
Oryza	12004.m10093	13104.m05320	IIIb	467	2	1	
Oryza	12006.m05305	13106.m00573	IIIb	601	3	2	unusual length

Table 2.4, Continued:

Taxon	Gene Model		Clade	Protein Length	Exons	Introns	Notes
Oryza sativa	Previous Model[a]	New Model[b]					
Oryza	12011.m08084	13111.m04223	IIIb	443	2	1	
Oryza	12003.m09799	13103.m05161	IV	463	1	0	
Oryza	12009.m06667	13109.m03682	IV	453	1	0	
Oryza	12010.m05336	13110.m01980	IV	437	2	1	
Oryza	12010.m05338	13110.m01982	IV	435	2	1	
Oryza	12010.m065196	13110.m01983	IV	435	1	0	likely same locus as 12010.m05338; excluded
Oryza	12001.m06797	13101.m00077	Va	469	2	1	
Oryza	12001.m07463	13101.m00875	Va	438	1	0	
Oryza	12001.m07526	13101.m00955	Va	420	2	1	
Oryza	12001.m08413	13101.m02031	Va	440	2	1	
Oryza	12001.m08945	13101.m02626	Va	456	1	0	
Oryza	12001.m10551	13101.m04414	Va	447	2	1	
Oryza	12001.m10552	13101.m04415	Va	425	2	1	
Oryza	12003.m09598	13103.m04918	Va	457	2	1	
Oryza	12004.m06185	13104.m00818	Va	426	1	0	
Oryza	12004.m06213	13104.m00851	Va	432	1	0	
Oryza	12004.m06424	13104.m01075	Va	442	1	0	
Oryza	12004.m09248	13104.m04238	Va	450	2	1	
Oryza	12005.m04991	13105.m00466	Va	437	2	1	
Oryza	12005.m05022	13105.m00506	Va	445	1	0	
Oryza	12005.m05165	13105.m00682	Va	467	1	0	
Oryza	12005.m05379	13105.m00960	Va	440	2	1	
Oryza	12005.m06391	13105.m02102	Va	433	2	1	
Oryza	12005.m06529	13105.m02257	Va	465	2	1	
Oryza	12005.m06808	13105.m02576	Va	478	1	0	

Table 2.4, Continued:

Taxon	Gene Model		Clade	Protein Length	Exons	Introns	Notes
Oryza sativa	Previous Model[a]	New Model[b]					
Oryza	12006.m05344	13106.m00632	Va	464	2	1	
Oryza	12006.m08485	13106.m04091	Va	432	2	1	
Oryza	12006.m09382	13106.m05163	Va	405	2	1	
Oryza	12006.m09490	13106.m05324	Va	451	1	0	
Oryza	12007.m05853	13107.m01462	Va	452	2	1	
Oryza	12007.m07907	13107.m03753	Va	444	2	1	
Oryza	12010.m03590	13110.m00064	Va	328	2	1	
Oryza	12010.m03591	13110.m00065	Va	465	2	1	
Oryza	12010.m03593	13110.m00067	Va	497	4	3	
Oryza	12010.m03594	13110.m00068	Va	435	1	0	
Oryza	12010.m03753	13110.m00245	Va	422	1	0	
Oryza	12010.m03756	13110.m00248	Va	423	1	0	
Oryza	12010.m06411	13110.m03243	Va	554	2	1	
Oryza	12011.m07014	13111.m03011	Va	463	3	2	
Oryza	12002.m09071	13102.m04425	Vb	443	3	2	
Oryza	12004.m09197	13104.m04162	Vb	443	4	3	likely same locus as 12004.m101565; excluded
Oryza	12004.m101565	13104.m04160	Vb	443	3	2	
Oryza	12004.m35364	13104.m04163	Vb	443	3	2	likely same locus as 12004.m101565; excluded
Oryza	12006.m05577	13106.m00901	Vb	446	1	0	
Oryza	12006.m05583	13106.m00907	Vb	434	1	0	
Oryza	12008.m05175	13108.m01109	Vb	459	2	1	
Oryza	12008.m08279	13108.m04655	Vb	256	2	1	unusual length
Oryza	12008.m08281	13108.m04657	Vb	443	2	1	

Table 2.4, Continued:

Taxon	Gene Model		Clade	Protein Length	Exons	Introns	Notes
Oryza sativa	**Previous Model[a]**	**New Model[b]**					
Oryza	12008.m26535	13108.m04658	Vb	443	2	1	likely same locus as 12008.m08281; excluded
Oryza	12009.m05698	13109.m02436	Vb	441	2	1	
Oryza	12011.m04992	13111.m00836	Vb	448	2	1	
Vitis vinifera							
Vitis	GSVIVP00029205001		Ia	458	1	0	
Vitis	GSVIVP00029208001		Ia	457	1	0	
Vitis	GSVIVP00014853001		Ib	413	3	2	
Vitis	GSVIVP00014854001		Ib	457	1	0	
Vitis	GSVIVP00018284001		Ib	474	2	1	
Vitis	GSVIVP00019525001		Ib	347	3	2	
Vitis	GSVIVP00019527001		Ib	437	1	0	
Vitis	GSVIVP00019528001		Ib	448	1	0	
Vitis	GSVIVP00021474001		Ib	457	3	2	
Vitis	GSVIVP00037690001		Ib	486	2	1	
Vitis	GSVIVP00037693001		Ib	593	2	1	
Vitis	GSVIVP00015096001		II	429	2	1	
Vitis	GSVIVP00031637001		II	435	2	1	
Vitis	GSVIVP00032576001		II	449	2	1	
Vitis	GSVIVP00031153001		II	443	2	1	
Vitis	GSVIVP00015540001		IIIa	520	2	1	
Vitis	GSVIVP00031254001		IIIa	330	1	0	
Vitis	GSVIVP00035108001		IIIa	456	2	1	
Vitis	GSVIVP00001216001		IIIb	450	1	0	
Vitis	GSVIVP00006805001		IIIb	464	3	2	

Table 2.4, Continued:

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Vitis vinifera						
Vitis	GSVIVP00006807001	IIIb	454	3	2	
Vitis	GSVIVP00006814001	IIIb	461	2	1	
Vitis	GSVIVP00006817001	IIIb	461	2	1	
Vitis	GSVIVP00006825001	IIIb	413	4	3	
Vitis	GSVIVP00006837001	IIIb	461	2	1	
Vitis	GSVIVP00029716001	IIIb	470	2	1	
Vitis	GSVIVP00029719001	IIIb	466	3	2	
Vitis	GSVIVP00029720001	IIIb	471	2	1	
Vitis	GSVIVP00001202001	Va	445	2	1	
Vitis	GSVIVP00001204001	Va	325	3	2	
Vitis	GSVIVP00002990001	Va	446	2	1	
Vitis	GSVIVP00003212001	Va	451	3	2	
Vitis	GSVIVP00003215001	Va	451	2	1	
Vitis	GSVIVP00003220001	Va	451	2	1	
Vitis	GSVIVP00009436001	Va	451	2	1	
Vitis	GSVIVP00012869001	Va	448	3	2	
Vitis	GSVIVP00013875001	Va	459	2	1	
Vitis	GSVIVP00014562001	Va	450	2	1	
Vitis	GSVIVP00016799001	Va	464	1	0	
Vitis	GSVIVP00017980001	Va	433	2	1	
Vitis	GSVIVP00022386001	Va	456	2	1	
Vitis	GSVIVP00023980001	Va	463	1	0	
Vitis	GSVIVP00025220001	Va	449	3	2	high homology to VlAMAT (Wang and De Luca 2005)
Vitis	GSVIVP00028244001	Va	424	3	2	
Vitis	GSVIVP00029106001	Va	435	2	1	

Table 2.4, Continued:

Taxon	Gene Model	Clade	Protein Length	Exons	Introns	Notes
Vitis vinifera						
Vitis	GSVIVP00029327001	Va	577	2	1	
Vitis	GSVIVP00001546001	Vb	874	2	1	unusual length
Vitis	GSVIVP00017173001	Vb	503	3	2	
Vitis	GSVIVP00017176001	Vb	457	2	1	
Vitis	GSVIVP00017179001	Vb	473	2	1	
Vitis	GSVIVP00017191001	Vb	375	6	5	unusual number of introns
Vitis	GSVIVP00025189001	Vb	429	2	1	

[a]Output for original BLASTP searches came from release 5; this nomenclature was used in Figures 2.2-2.5. [b]Updated genome release 6.1 shown for correspondence purposes.

Table 2.5: Duplications and Retrotransposons Associated With *Populus* BAHD Acyltransferase Genes

Genes are organized as in Table 2.1.

Gene Model v2.0	Name[a]	Duplication Type[b]	Retrotransposon[c]	Gene Partner(s)	Notes
POPTR_0001s40570	MATL1	local (3)	5' and 3'	MATL2, MATL3	POPTR_0001s40580 is a transposon, POPTR_0001s40590 is a partial BAHD, POPTR_0001s40610 is a transposon
POPTR_0001s40600	MATL2	local (3)	5' and 3'	MATL1, MATL3	
POPTR_0001s40620	MATL3	local (3)	5'	MATL1, MATL2	
POPTR_0001s45940	MATL4				
POPTR_0004s09280	MATL5				
POPTR_0004s09520	MATL6	local (3)		MATL7, MATL8	POPTR_0001s09540 is a partial BAHD
POPTR_0004s09530	MATL7	local (3)		MATL6, MATL8	
POPTR_0004s09550	MATL8	local (3)	5' and 3'	MATL6, MATL7	
POPTR_0004s10330	MATL9				
POPTR_0004s10920	MATL10				
POPTR_0004s11990	MATL11		3'		
POPTR_0004s18920	MATL12		5' and 3'	MATL10, MATL13 (RE)[c]	
POPTR_0004s19020	MATL13		5' and 3'	MATL10, MATL12 (RE)[c]	
POPTR_0009s02480	MATL14				
POPTR_0009s06800	MATL15				
POPTR_0010s21520	MATL16				
POPTR_0017s13040	MATL17				
POPTR_0019s14060	MATL18				
POPTR_0003s05570	ATL1	local (3)	5' and 3'	ATL2, ATL3	
POPTR_0003s05580	ATL2	local (3)		ATL1, ATL3	

Table 2.5, Continued:

Gene Model v2.0	Name[a]	Duplication Type[b]	Retrotransposon[c]	Gene Partner(s)	Notes
POPTR_0003s05590	ATL3	local (3)		ATL1, ATL2	
POPTR_0006s09870	ATL4	salicoid		ATL11	
POPTR_0008s06520	ATL5	salicoid		ATL6	
POPTR_0010s19980	ATL6	salicoid	3'	ATL5	
POPTR_0012s12890	ATL7	salicoid		ATL10	
POPTR_0014s02560	ATL8	local		ATL9	
POPTR_0014s02570	ATL9	local		ATL8	
POPTR_0015s12810	ATL10	salicoid	3'	ATL7	
POPTR_0016s11990	ATL11	salicoid	5'	ATL4	
POPTR_0001s32660	CERL1				
POPTR_0005s05380	CERL2				
POPTR_0005s19690	CERL3		5' and 3'	CERL1	
POPTR_0013s03730	CERL4				
POPTR_0018s01250	CERL5				
POPTR_0001s31750	AATL1				
POPTR_0002s01180	AATL2	salicoid		AATL4	
POPTR_0004s01720	AATL3				
POPTR_0005s27190	AATL4	salicoid		AATL2	
POPTR_0006s01180	AATL5	local	5'	AATL6	
POPTR_0006s01190	AATL6	local		AATL5	
POPTR_0006s03260	AATL7				
POPTR_0006s03450	AATL8				
POPTR_0007s00580	AATL9				

Table 2.5, Continued:

Gene Model v2.0	Name[a]	Duplication Type[b]	Retrotransposon[c]	Gene Partner(s)	Notes
POPTR_0007s00980	AATL10	salicoid		AATL20	
POPTR_0008s18070	AATL11	salicoid		AATL14	
POPTR_0010s06380	AATL12	local		AATL13	POPTR_0010s06400 is a partial BAHD
POPTR_0010s06390	AATL13	local	3'	AATL12	
POPTR_0010s06410	AATL14	salicoid	5' and 3'	AATL11 (sal), AATL15 (RE)[c]	
POPTR_0010s06640	AATL15		5' and 3'	AATL11, AATL14 (RE)[c]	
POPTR_0011S12490	AATL16	local		AATL17	POPTR_0011s12480, POPTR_0011s12490 and POPTR_0011s12500, POPTR_0011s12510 appear to form duplicate segments
POPTR_0011s12510	AATL17	local		AATL16	
POPTR_0015s14800	AATL18				
POPTR_0015s14850	AATL19				
POPTR_0017s04320	AATL20	salicoid		AATL10	
POPTR_0017s05330	AATL21				
POPTR_0017s05810	AATL22				
POPTR_0019s01520	AATL23	local		AATL24	POPTR_0019s01530 is a partial BAHD
POPTR_0019s01540	AATL24	local		AATL23	
POPTR_0001s01020	CFATL1				
POPTR_0007s15050	CFATL2		5'		
POPTR_0001s33390	ABTL1	salicoid		ABTL13	
POPTR_0002s03350	ABTL2	salicoid		ABTL5	
POPTR_0002s24610	ABTL3		3'		
POPTR_0004s05280	ABTL4		3'		

Table 2.5, Continued:

Gene Model v2.0	Name[a]	Duplication Type[b]	Retrotransposon[c]	Gene Partner(s)	Notes
POPTR_0005s25250	ABTL5	salicoid		ABTL2	
POPTR_0006s16070	ABTL6		3'		
POPTR_0008s07130	ABTL7	salicoid		ABTL9	
POPTR_0010s18720 plus POPTR_0010s18710	ABTL8				
POPTR_0010s19360	ABTL9	salicoid		ABTL7	
POPTR_0014s10890	ABTL10				
POPTR_0014s16460	ABTL11				
POPTR_0015s11290	ABTL12				
POPTR_0017s09970	ABTL13	salicoid	5' and 3'	ABTL1	
POPTR_0019s14700	ABTL14				
POPTR_0003s01420	AMATL1		5'		
POPTR_0013s11650	AMATL2		5' and 3'	AMATL1 (RE)[c]	
POPTR_0001s00980 plus POPTR_0001s00970	AMATL3	local	5'	AMATL4	POPTR_0001s00960 is a partial BAHD
POPTR_0001s00950	AMATL4	local	3'	AMATL3	
POPTR_0001s45150	HMTL1	local x2	3'	HMTL2 (loc1), HMTL3-6 (loc2)	
POPTR_0001s45160	HMTL2	local x2		HMTL1 (loc1), HMTL3-6 (loc2)	
POPTR_0001s45180	HMTL3	local (4) x2	5'	HMTL1-2 (loc2), HMTL4-6 (loc3)	locus inverted relative to other HMTL genes
POPTR_0001s45190	HMTL4	local (4) x2	5' and 3'	HMTL1-2 (loc2), HMTL3 (loc3), HMTL5-6 (loc3)	

Table 2.5, Continued:

Gene Model v2.0	Name[a]	Duplication Type[b]	Retrotransposon[c]	Gene Partner(s)	Notes
POPTR_0001s45200	HMTL5	local (4) x2	5'	HMTL1-2 (loc2), HMTL3-4 (loc3), HMTL6 (loc3)	
POPTR_0001s45210	HMTL6	salicoid, local (4) x2		HMTL1-2 (loc2), HMTL3-5 (loc3), HMTL7 (sal)	
POPTR_0011s15660	HMTL7	salicoid		HMTL6	
POPTR_0013s07220	CHATL1	local (3)		CHATL2, CHATL3	
POPTR_0013s07230	CHATL2	local (3)		CHATL1, CHATL3	
POPTR_0013s07240	CHATL3	salicoid, local (3)	5'	CHATL1-2 (loc), CHATL6 (sal)	
POPTR_0019s01680	CHATL4	local	5'	CHATL5	
POPTR_0019s01700 plus POPTR_0019s01690	CHATL5	local	3'	CHATL4	
POPTR_0019s06040	CHATL6	salicoid	5'	CHATL3	
POPTR_0003s18210	HCT1	salicoid		HCT6	
POPTR_0018s11440 plus POPTR_0018s11450	HCT2		3'		
POPTR_0018s11380	HCT3	local		HCT4	
POPTR_0018s11370	HCT4	local		HCT3	
POPTR_0005s02820	HCT5	local		HCT7	
POPTR_0001s03440	HCT6	salicod		HCT1	
POPTR_0005s02810	HCT7	local		HCT5	
POPTR_0006s17880	SHTL1	salicoid	5' and 3'	SHTL2	
POPTR_0018s11840	SHTL2	salicoid		SHTL1	

[a]Gene names indicated in red are members of large *Populus*-dominated subclades. [b]Numbers in parentheses indicate the number of genes in a local array with more than two members, while "x2" indicates two distinct local duplications. [c]Presence of multiple retrotransposon BLAST hits within 10 kb of the 5' and/or 3' end of the gene model. [d]Likely partner(s) for a putative duplication event by retrotransposition.

Table 2.6: Correspondences of *Populus* BAHD Acyltransferase Genes and Affymetrix Probe Identifiers

Coloration for gene name is assigned according to clade membership in Figure 2.1.

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
MATL1	POPTR_0001s40570	PtpAffx.225268.1.S1_s_at	X	X		
		PtpAffx.32180.1.A1_at	X	X		
		PtpAffx.225268.1.S1_at			X	
MATL2	POPTR_0001s40600	PtpAffx.32180.1.A1_at	X	X		
MATL3	POPTR_0001s40620	PtpAffx.225268.1.S1_s_at	X	X		
		PtpAffx.32180.1.A1_at	X	X		
MATL4	POPTR_0001s45940	PtpAffx.201430.1.S1_s_at	X	X		
		PtpAffx.219934.1.S1_at	X			
		PtpAffx.201430.1.S1_at			X	
MATL5	POPTR_0004s09280	PtpAffx.16402.2.S1_a_at			X	
		PtpAffx.16402.2.S1_at			X	
MATL6	POPTR_0004s09520	Ptp.6874.1.S1_at	X[c]			
		Ptp.6874.1.S1_s_at	X[c]	X		
		PtpAffx.225728.1.S1_at			X	
MATL7	POPTR_0004s09530					X
MATL8	POPTR_0004s09550					X
MATL9	POPTR_0004s10330	PtpAffx.75275.1.A1_at			X	
MATL10	POPTR_0004s10920					X
MATL11	POPTR_0004s11990	PtpAffx.224408.1.S1_at			X	
MATL12	POPTR_0004s18920	PtpAffx.217327.1.S1_s_at			X	
		PtpAffx.218389.1.S1_s_at			X	

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
MATL13	POPTR_0004s19020	PtpAffx.217327.1.S1_s_at			X	
		PtpAffx.218389.1.S1_s_at			X	
MATL14	POPTR_0009s02480	PtpAffx.224453.1.S1_at			X	
MATL15	POPTR_0009s06800	PtpAffx.204833.1.S1_at			X	
MATL16	POPTR_0010s21520	PtpAffx.219344.1.S1_at			X	
MATL17	POPTR_0017s13040	PtpAffx.219511.1.S1_at			X	
MATL18	POPTR_0019s14060	PtpAffx.212510.1.S1_s_at	X	X		
		PtpAffx.222102.1.S1_s_at	X			
		PtpAffx.212509.1.S1_at			X	
		PtpAffx.212510.1.S1_at			X	
ATL1	POPTR_0003s05570	Ptp.2352.1.A1_at	X	X		
		PtpAffx.107444.1.A1_at	X			
		PtpAffx.107444.2.S1_at			X	
ATL2	POPTR_0003s05580	Ptp.2352.1.A1_at	X	X		
		PtpAffx.107444.1.A1_at	X			
		PtpAffx.107444.2.S1_at			X	
ATL3	POPTR_0003s05590	Ptp.2352.1.A1_at	X	X		
		PtpAffx.107444.1.A1_at	X			
		PtpAffx.107444.1.A1_s_at	X	X		
		PtpAffx.107444.2.S1_at			X	
ATL4	POPTR_0006s09870	PtpAffx.608.2.A1_at	X			
		PtpAffx.608.3.S1_at	X	X		
		PtpAffx.608.3.S1_a_at	X			

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
ATL5	POPTR_0008s06520	PtpAffx.207743.1.S1_at			X	
ATL6	POPTR_0010s19980	PtpAffx.209294.1.S1_at			X	
ATL7	POPTR_0012s12890	PtpAffx.7311.3.A1_at	X	X		
ATL8	POPTR_0014s02560	Ptp.3032.2.A1_at			X	
		PtpAffx.12609.1.S1_at			X	
		PtpAffx.64204.1.A1_at			X	
ATL9	POPTR_0014s02570	Ptp.3032.1.S1_s_at	X	X		
		PtpAffx.221194.1.S1_at			X	
		PtpAffx.221194.1.S1_x_at			X	
ATL10	POPTR_0015s12810	PtpAffx.7311.1.S1_at	X	X		
ATL11	POPTR_0016s11990	PtpAffx.608.1.S1_at	X	X		
		PtpAffx.608.3.S1_a_at	X			
		PtpAffx.608.4.S1_a_at	X			
CERL1	POPTR_0001s32660	PtpAffx.201085.1.S1_at			X	
CERL2	POPTR_0005s05380					X
CERL3	POPTR_0005s19690	PtpAffx.205434.1.S1_at	X	X		
CERL4	POPTR_0013s03730	PtpAffx.64491.1.A1_at			X	
CERL5	POPTR_0018s01250	PtpAffx.214182.1.S1_at			X	
		PtpAffx.2581.1.S1_at			X	
		PtpAffx.93272.1.A1_at			X	
AATL1	POPTR_0001s31750	Ptp.3830.1.A1_at			X	
AATL2	POPTR_0002s01180	PtpAffx.93558.1.A1_at	X	X		

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
AATL3	POPTR_0004s01720	PtpAffx.222535.1.S1_s_at	X	X		
		PtpAffx.16622.1.A1_at			X	
AATL4	POPTR_0005s27190					X
AATL5	POPTR_0006s01180	PtpAffx.36103.2.S1_at			X	
AATL6	POPTR_0006s01190	PtpAffx.36103.1.S1_a_at	X			
		PtpAffx.36103.1.S1_at	X	X		
		PtpAffx.149473.1.S1_at			X	
AATL7	POPTR_0006s03260	PtpAffx.206043.1.S1_at	X	X		
AATL8	POPTR_0006s03450	PtpAffx.206052.1.S1_s_at			X	
AATL9	POPTR_0007s00580	PtpAffx.206882.1.S1_at			X	
AATL10	POPTR_0007s00980					X
AATL11	POPTR_0008s18070	PtpAffx.22923.1.A1_at	X	X		
		Ptp.7502.1.S1_at			X	
		PtpAffx.22923.1.A1_s_at			X	
AATL12	POPTR_0010s06380	PtpAffx.136959.1.S1_s_at			X	
		PtpAffx.224634.1.S1_s_at			X	
		PtpAffx.94880.1.A1_at			X	
AATL13	POPTR_0010s06390	PtpAffx.136959.1.S1_s_at			X	
		PtpAffx.224634.1.S1_s_at			X	
		PtpAffx.94880.1.A1_at			X	
		Ptp.7639.1.S1_at			X	
		PtpAffx.136959.1.S1_at			X	
AATL14	POPTR_0010s06410					X

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
AATL15	POPTR_0010s06640	PtpAffx.224633.1.S1_s_at			X	
		PtpAffx.55456.1.A1_at			X	
AATL16	POPTR_0011s12490	PtpAffx.9989.1.S1_at	X	X		
AATL17	POPTR_0011s12510					X
AATL18	POPTR_0015s14800	PtpAffx.225006.1.S1_at			X	
AATL19	POPTR_0015s14850	PtpAffx.225005.1.S1_at			X	
AATL20	POPTR_0017s04320	PtpAffx.222290.1.S1_s_at			X	
AATL21	POPTR_0017s05330	PtpAffx.222290.1.S1_s_at			X	
AATL22	POPTR_0017s05810	PtpAffx.225720.1.S1_at			X	
AATL23	POPTR_0019s01520	Ptp.6192.1.S1_s_at	X			
		PtpAffx.225142.1.S1_s_at	X	X		
		Ptp.6192.1.S1_at			X	
		PtpAffx.22957.1.A1_at			X	
AATL24	POPTR_0019s01540	PtpAffx.215069.1.S1_at	X	X		
CFATL1	POPTR_0001s01020	PtpAffx.220124.1.S1_at			X	
CFATL2	POPTR_0007s15050	PtpAffx.16365.1.A1_at			X	
ABTL1	POPTR_0001s33390	PtpAffx.114321.1.A1_at			X	
ABTL2	POPTR_0002s03350	PtpAffx.201638.1.S1_at			X	
ABTL3	POPTR_0002s24610	PtpAffx.202679.1.S1_at			X	
ABTL4	POPTR_0004s05280	PtpAffx.203851.1.S1_at	X	X		
		PtpAffx.203851.1.S1_s_at	X			
ABTL5	POPTR_0005s25250	PtpAffx.205785.1.S1_at			X	
		PtpAffx.28677.1.S1_at			X	

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
ABTL6	POPTR_0006s16070	PtpAffx.216977.1.S1_at			X	
ABTL7	POPTR_0008s07130	PtpAffx.207772.1.S1_s_at			X	
ABTL8	POPTR_0010s18720 plus POPTR_0010s18710	PtpAffx.135343.1.S1_at	X			
		PtpAffx.209233.1.S1_at	X	X		
		PtpAffx.9895.1.A1_at			X	
ABTL9	POPTR_0010s19360	PtpAffx.209270.1.S1_at			X	
ABTL10	POPTR_0014s10890	PtpAffx.98699.2.S1_a_at	X	X		
		PtpAffx.98699.1.S1_at			X	
ABTL11	POPTR_0014s16460	PtpAffx.211922.1.S1_at	X	X		
ABTL12	POPTR_0015s11290	Ptp.5997.1.S1_at	X	X		
		PtpAffx.133018.1.S1_s_at	X			
		PtpAffx.45793.1.A1_at			X	
ABTL13	POPTR_0017s09970					X
ABTL14	POPTR_0019s14700	PtpAffx.212485.1.S1_s_at	X	X		
		PtpAffx.219012.1.S1_s_at	X			
AMATL1	POPTR_0003s01420	PtpAffx.202806.1.S1_at	X	X		
		PtpAffx.41880.1.S1_s_at	X			
AMATL2	POPTR_0013s11650	PtpAffx.221303.1.S1_at			X	
AMATL3	POPTR_0001s00980 plus POPTR_0001s00970	PtpAffx.114792.1.A1_at			X	
		PtpAffx.114792.1.A1_s_at			X	
		PtpAffx.220128.1.S1_at			X	
AMATL4	POPTR_0001s00950	PtpAffx.220130.1.S1_at			X	
HMTL1	POPTR_0001s45150	PtpAffx.39214.1.A1_at	X	X		

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
HMTL2	POPTR_0001s45160	PtpAffx.39214.1.A1_at	X	X		
HMTL3	POPTR_0001s45180	PtpAffx.39214.1.A1_at	X	X		
HMTL4	POPTR_0001s45190	PtpAffx.39214.1.A1_at	X	X		
HMTL5	POPTR_0001s45200	PtpAffx.39214.1.A1_at	X	X		
HMTL6	POPTR_0001s45210	PtpAffx.144977.1.S1_at	X			
		PtpAffx.59062.1.A1_at	X	X		
HMTL7	POPTR_0011s15660	PtpAffx.121447.1.S1_at	X	X		
		PtpAffx.140872.1.A1_at			X	
CHATL1	POPTR_0013s07220	PtpAffx.6696.4.S1_at	X	X		
CHATL2	POPTR_0013s07230	Ptp.1105.2.A1_s_at	X	X		
		PtpAffx.6696.2.S1_s_at	X			
		PtpAffx.6696.3.S1_at			X	
CHATL3	POPTR_0013s07240	PtpAffx.211140.1.S1_s_at	X	X		
		PtpAffx.6696.2.S1_s_at	X			
		PtpAffx.6696.2.S1_at			X	
CHATL4	POPTR_0019s01680	PtpAffx.215077.1.S1_s_at	X	X		
CHATL5	POPTR_0019s01700 plus POPTR_0019s01690	PtpAffx.215077.1.S1_s_at	X	X		
CHATL6	POPTR_0019s06040	Ptp.4093.1.S1_at	X	X		
HCT1	POPTR_0003s18210	PtpAffx.6492.1.S1_at	X	X		
		Ptp.6327.1.S1_at			X	
		Ptp.6333.2.S1_at			X	

Table 2.6, Continued:

Gene	Model	Probe ID[a]	Present[b]	Used in Figure 2.11	Absent[b]	Not On Array
HCT2	POPTR_0018s11440 plus POPTR_0018s11450	PtpAffx.214461.1.S1_at	X	X		
		PtpAffx.221469.1.S1_s_at	X			
		PtpAffx.139992.1.A1_at			X	
HCT3	POPTR_0018s11380	PtpAffx.162632.1.A1_at			X	
		PtpAffx.162632.2.A1_s_at			X	
		PtpAffx.214458.1.S1_s_at			X	
HCT4	POPTR_0018s11370	PtpAffx.162632.1.A1_at			X	
		PtpAffx.162632.2.A1_s_at			X	
		PtpAffx.214458.1.S1_s_at			X	
HCT5	POPTR_0005s02820	PtpAffx.215992.1.S1_at	X	X		
		PtpAffx.2057.1.S1_s_at	X			
		PtpAffx.218010.1.S1_s_at	X			
		PtpAffx.218010.1.S1_at			X	
HCT6	POPTR_0001s03440	PtpAffx.6492.2.A1_s_at	X	X		
HCT7	POPTR_0005s02810	PtpAffx.218009.1.S1_s_at	X	X		
		PtpAffx.2057.1.S1_s_at	X			
		PtpAffx.218010.1.S1_s_at	X			
SHTL1	POPTR_0006s17880	PtpAffx.215312.1.S1_at			X	
SHTL2	POPTR_0018s11840	PtpAffx.214477.1.S1_at			X	
TOTAL (unique probes)	**N/A**	**139**	**60**	**41**	**79**	**N/A**
TOTAL (unique genes)	**100**	**N/A**	**48**	**48**	**43**	**9**

[a]Probe names highlighted with the same color within a given subclade are identical, indicating potential for cross-hybridizing to multiple BAHD acyltransferases. [b]Probes with raw hybridization intensity $\geq$50 in at least one set of biological replicates are flagged as "Present," otherwise they are considered "Absent."

Table 2.7: Summary of Gene Duplication Events Among *Populus* BAHD Acyltransferases

Clade	Ia	Ib	II	IIIa	IIIb	Va	Vb	Genome
Total genes in clade	18	11	5	24	2	31	9	100
Recent duplication	6 (33%)	11 (100%)	0 (0%)	14 (58%)	0 (0%)	21[a] (68%)	8 (89%)	60 (60%)
Salicoid duplication	0	6	0	6	0	10	4	26
Local duplication	6[b]	5[c]	0	8	0	13[d]	4	36
Retrotransposon association	7 (39%)	4 (36%)	1 (20%)	4 (17%)	1 (50%)	16 (52%)	2 (22%)	35 (35%)
Both 5' and 3' of gene	5	1	1	2	0	3	1	13
Either 5' or 3' of gene	2	3	0	2	1	13	1	22

[a] Two members in Clade Va are associated with both salicoid and local duplications (21=10+13-2); [b] Includes two triplets; [c] Includes one triplet; [d] Includes one triplet and one quadruplet.

Figures

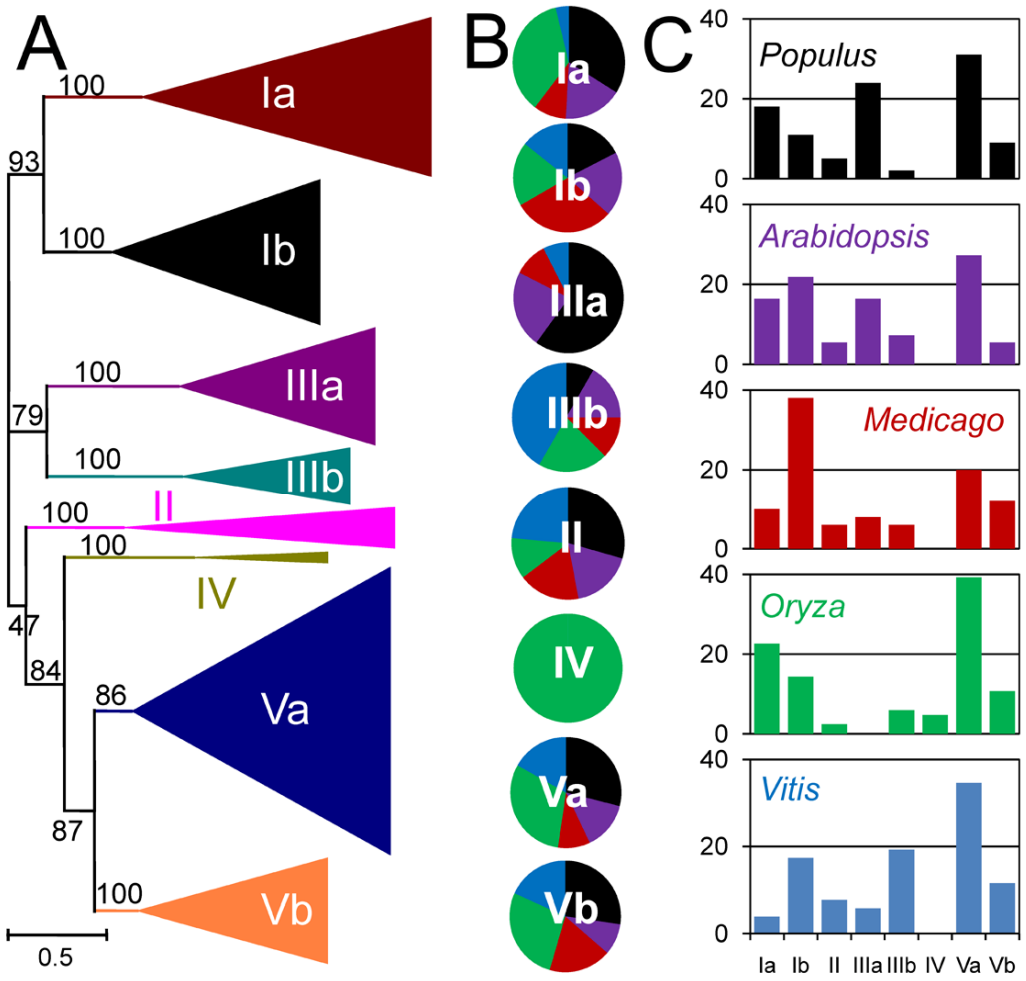

Figure 2.1: Phylogeny and Distribution of BAHD Acyltransferases

A: Protein phylogeny of biochemically characterized BAHD acyltransferases and putative BAHD proteins from *Arabidopsis*, *Medicago*, *Oryza*, *Populus*, and *Vitis* genomes. Phylogeny was constructed using maximum likelihood analysis. B: Percentage representation of putative BAHD acyltransferases across the five taxa within each phylogenetic clade. Colors correspond to the plant taxa as listed in C. C: Percentage representation of clade membership for putative BAHD acyltransferases within each plant genome.

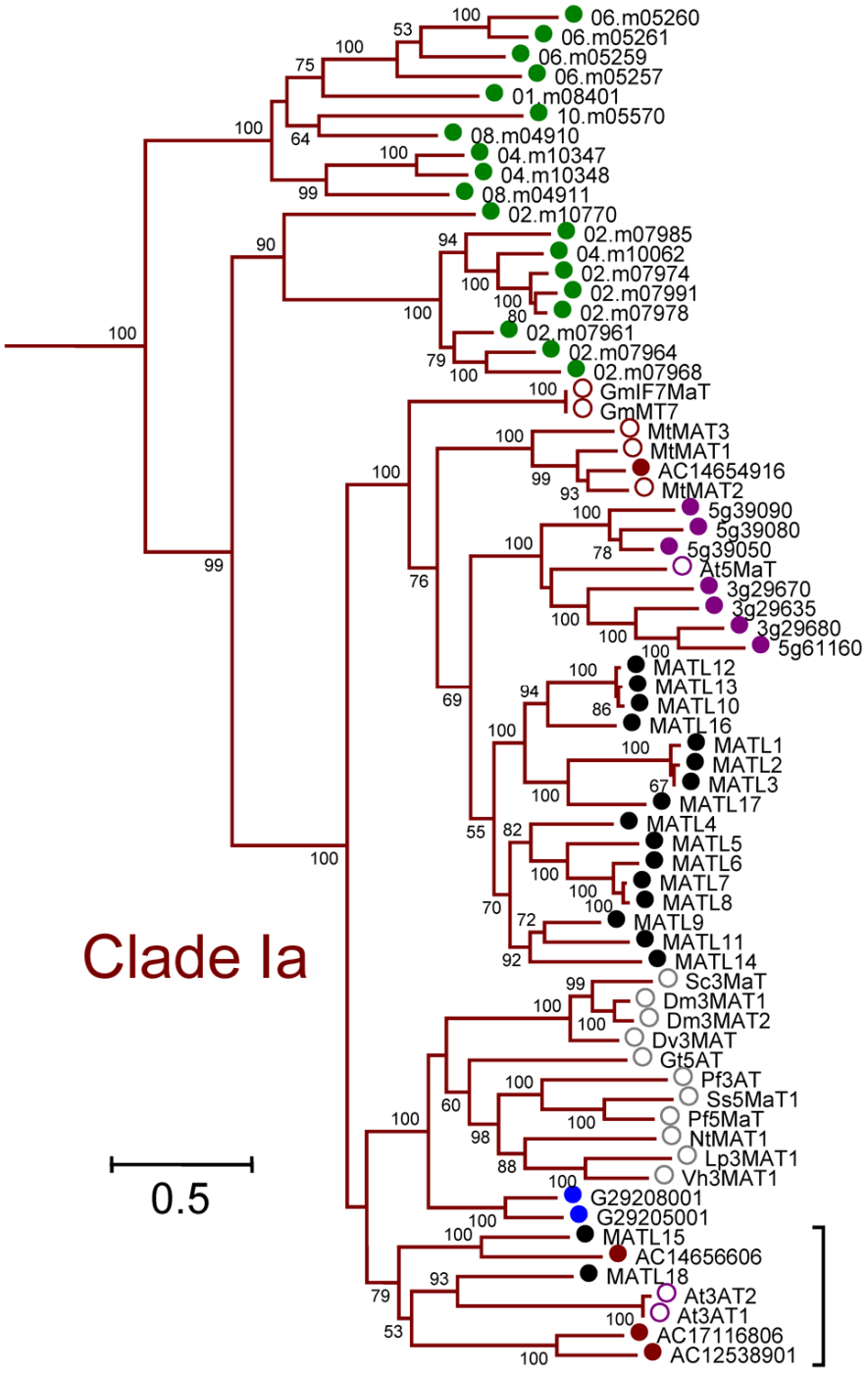

Figure 2.2: Phylogenetic Relationship of Clade Ia Members

Expanded view of all Clade Ia sequences from Figure 2.1A. Bracket indicates region lacking taxon-specific clustering. Filled circles represent putative BAHD acyltransferases, while open circles represent characterized BAHD proteins. Colors

correspond to taxa as listed in Figure 2.1, with gray circles indicating sequences from plants within the Asterids. *Populus* sequence names are provided in Table 2.1. Loci from the other four genomes have been truncated to accommodate text input limitations (e.g., 1g03495 for At1g03495 of *Arabidopsis*, AC1253891 for AC125389_1 of *Medicago*, 01.m08401 for 12001.m08401 of *Oryza*, G29205001 for GSVIVP00029205001 of *Vitis*). GenBank accession numbers and full names for previously characterized proteins are provided in Table 2.3.

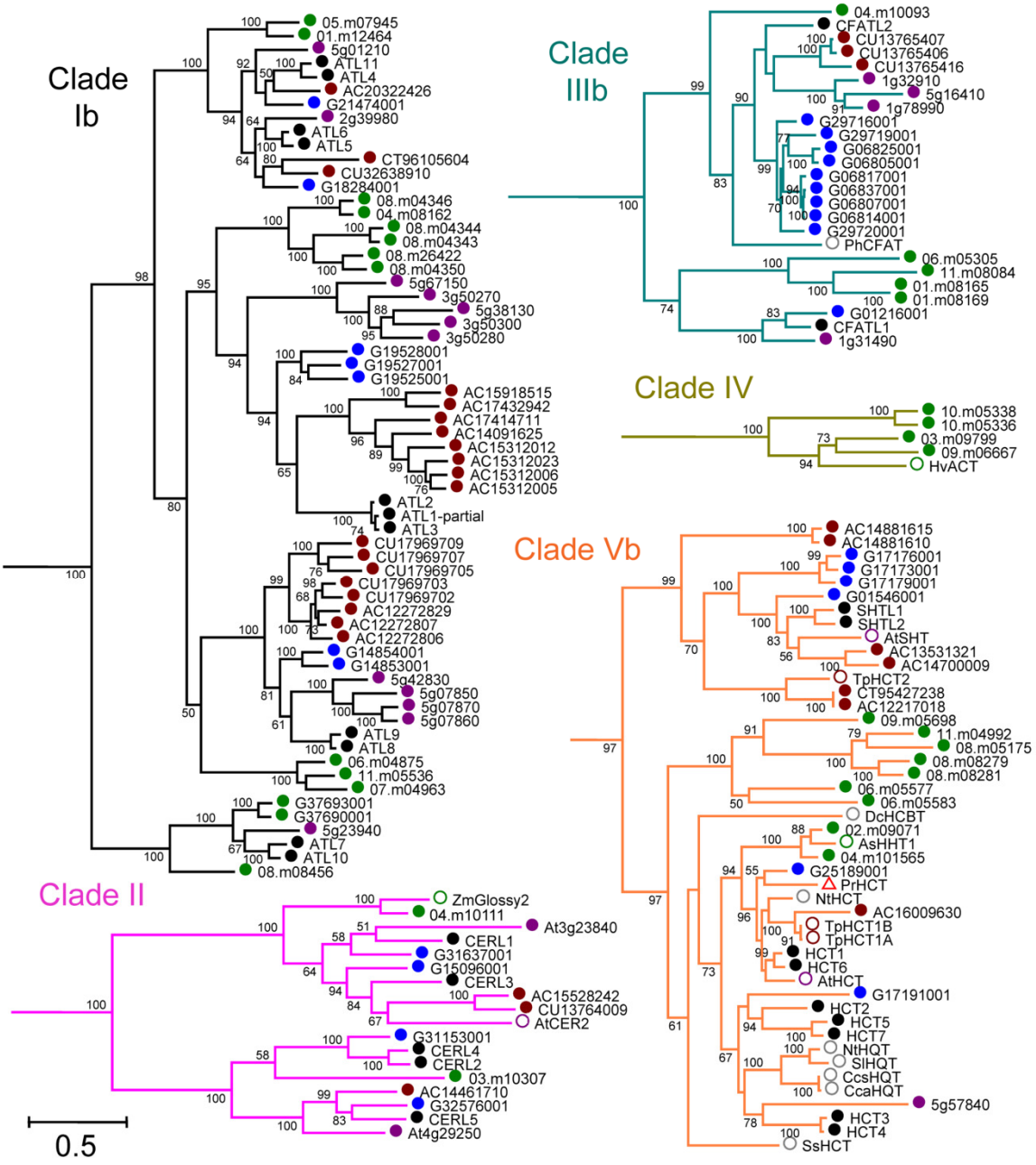

Figure 2.3: Detailed Views of Phylogenetic Relationships Within Clades Ib, II, IIIb, IV, and Vb

Coloration of clades and symbols are as described in Figures 2.1 and 2.2. In addition, red triangle indicates sequence from a gymnosperm.

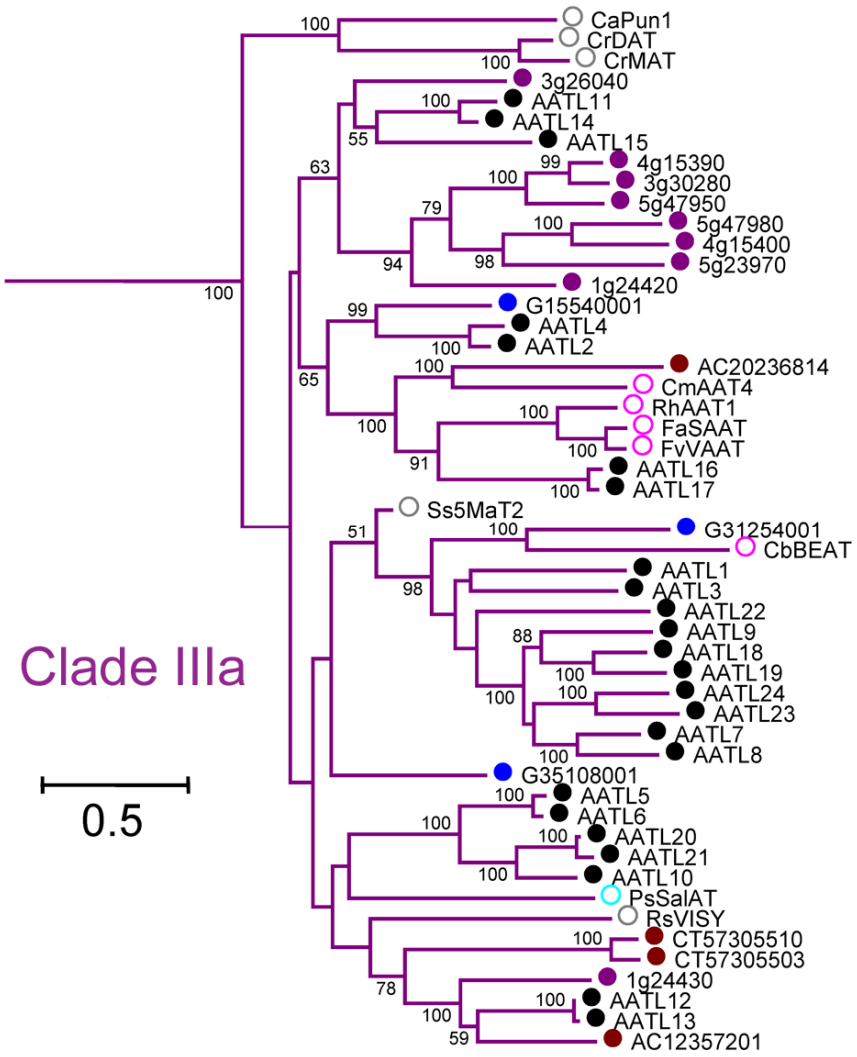

Figure 2.4: Phylogenetic Relationship of Clade IIIa Members

Expanded view of all Clade IIIa sequences from Figure 2.1A. Colors and symbols are the same as in Figures 2.1-2.3. In addition, pink circles indicate sequences from plants within the Rosids, while teal circle indicates sequence from a basal eudicot.

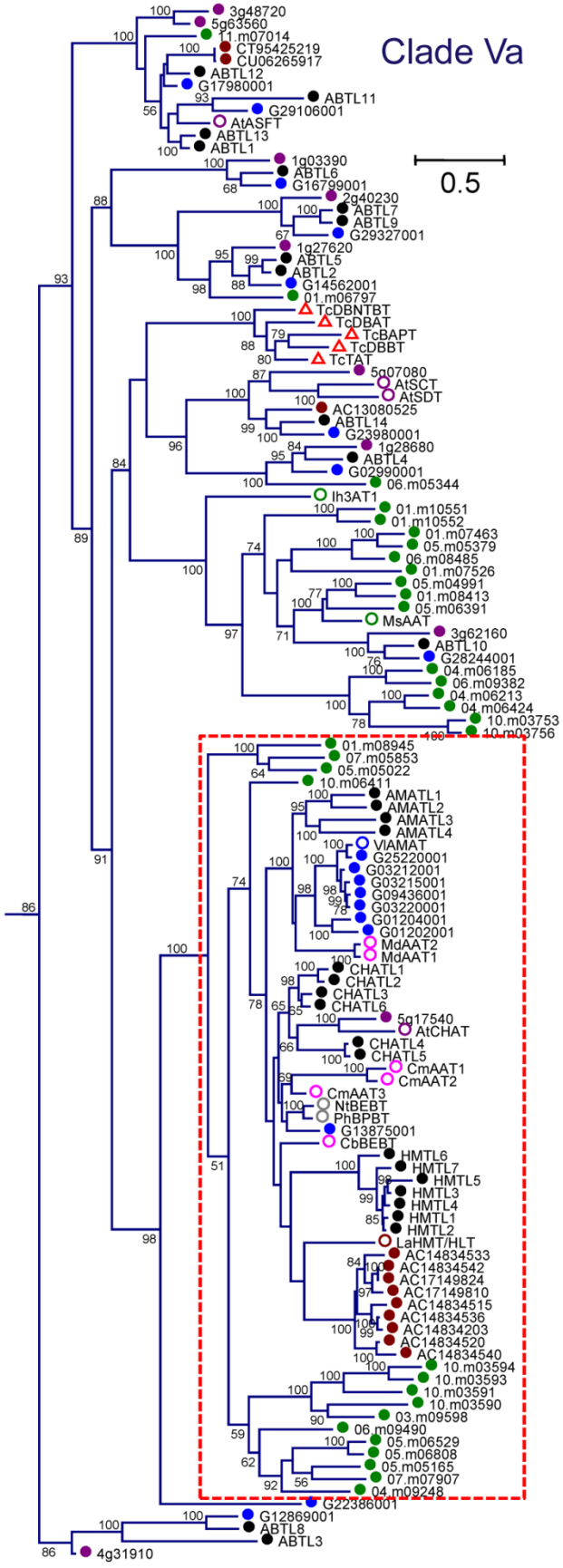
Clade Va
0.5
3g48720
5g63560
11.m07014
CT95425219
CU06265917
ABTL12
G17980001
ABTL11
G29106001
AtASFT
ABTL13
ABTL1
1g03390
ABTL6
G16799001
2g40230
ABTL7
ABTL9
G29327001
1g27620
ABTL5
ABTL2
G14562001
01.m06797
TcDBNTBT
TcDBAT
TcBAPT
TcDBBT
TcTAT
5g07080
AtSCT
AtSDT
AC13080525
ABTL14
G23980001
1g28680
ABTL4
G02990001
06.m05344
Ih3AT1
01.m10551
01.m10552
01.m07463
05.m05379
06.m08485
01.m07526
05.m04991
01.m08413
05.m06391
MsAAT
3g62160
ABTL10
G28244001
04.m06185
06.m09382
04.m06213
04.m06424
10.m03753
10.m03756
01.m08945
07.m05853
05.m05022
10.m06411
AMATL1
AMATL2
AMATL3
AMATL4
VlAMAT
G25220001
G03212001
G03215001
G09436001
G03220001
G01204001
G01202001
MdAAT2
MdAAT1
CHATL1
CHATL2
CHATL3
CHATL6
5g17540
AtCHAT
CHATL4
CHATL5
CmAAT1
CmAAT2
CmAAT3
NtBEBT
PhBPBT
G13875001
CbBEBT
HMTL6
HMTL7
HMTL5
HMTL3
HMTL4
HMTL1
HMTL2
LaHMT/HLT
AC14834533
AC14834542
AC17149824
AC17149810
AC14834515
AC14834536
AC14834203
AC14834520
AC14834540
10.m03594
10.m03593
10.m03591
10.m03590
03.m09598
06.m09490
05.m06529
05.m06808
05.m05165
07.m07907
04.m09248
G22386001
G12869001
ABTL8
ABTL3
4g31910

Figure 2.5: Phylogenetic Relationship of Clade Va Members (Previous Page)

Expanded view of all Clade Va sequences from Figure 2.1A. Colors and symbols are the same as in Figures 2.1-2.4. Boxed region indicates a poorly resolved branch based on bootstrap analysis.

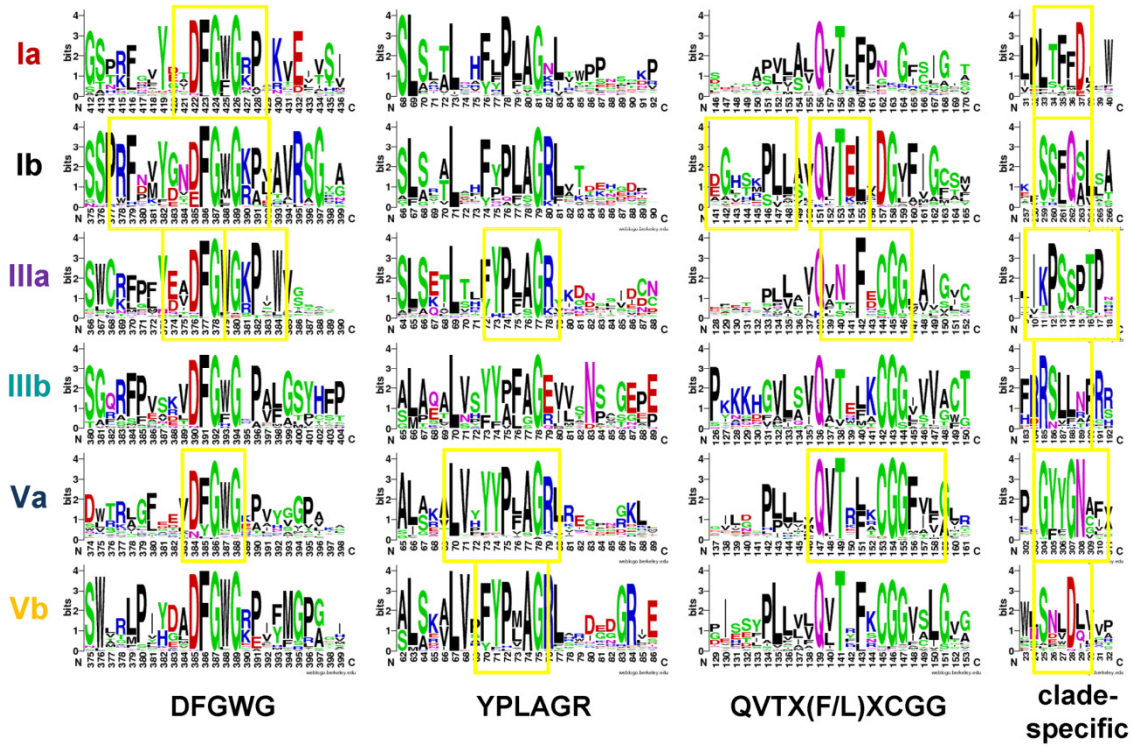

Figure 2.6: Conserved Motifs Within Phylogenetic Clades

WebLogo displays of consensus sequences corresponding to MINER-identified motifs, boxed in yellow. Logos are arranged in rows by phylogenetic clade, named at left, and in columns by motif, labelled at the bottom. The three leftmost columns represent motifs conserved across multiple clades. The rightmost column provides examples of clade-specific motifs; motifs in this column are not aligned relative to one another.

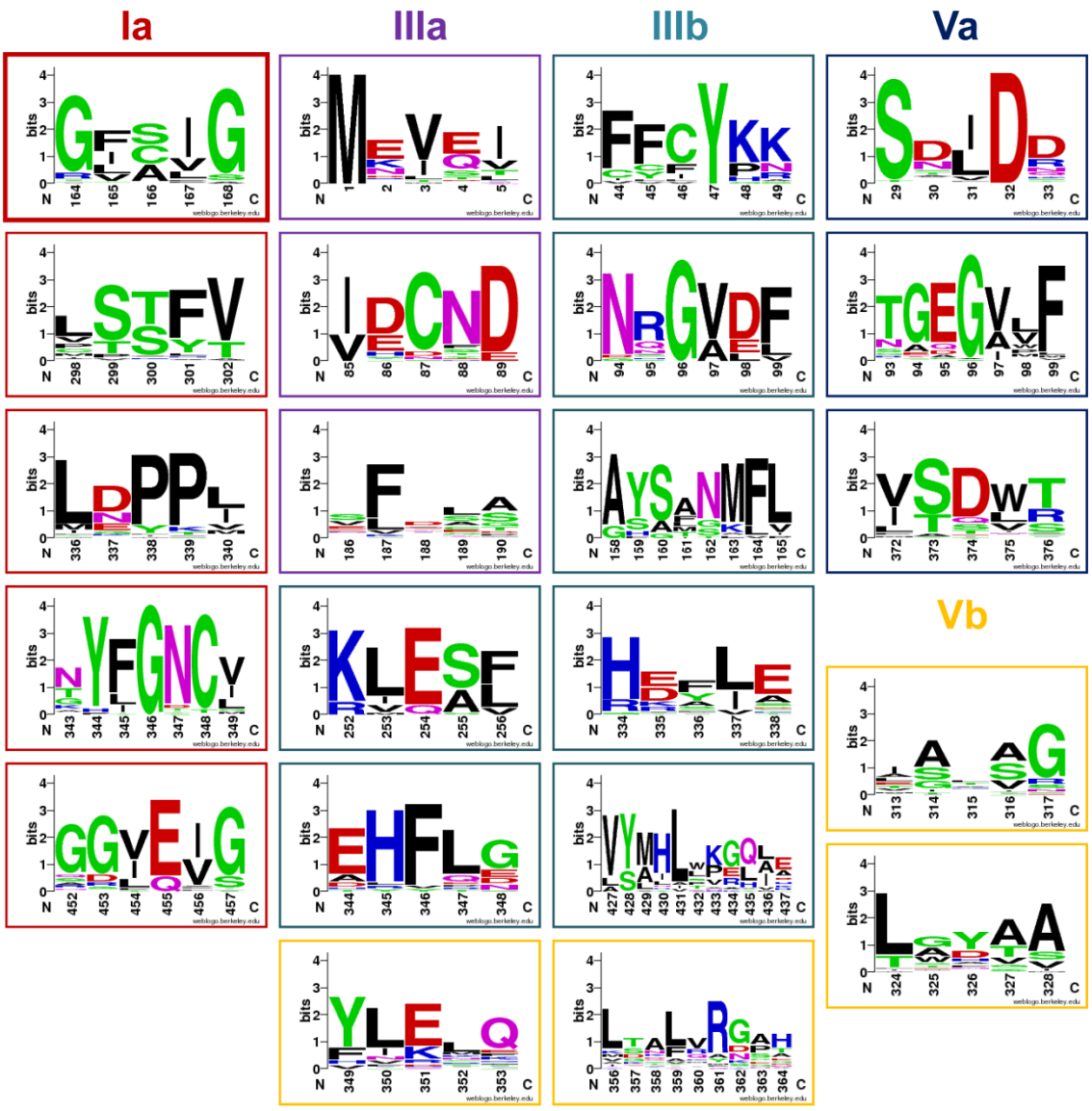

Figure 2.7: Additional Clade-Specific Motifs Identified by MINER

Motifs are arranged by clade, and bordered with the same color scheme as in Figure 2.1. The thickly boxed motif in Clade Ia overlaps with the range for the QVTX(F/L)XCGG motif shown in Figure 2.6. Clade Ib had no additional motifs beyond those shown in Figure 2.6.

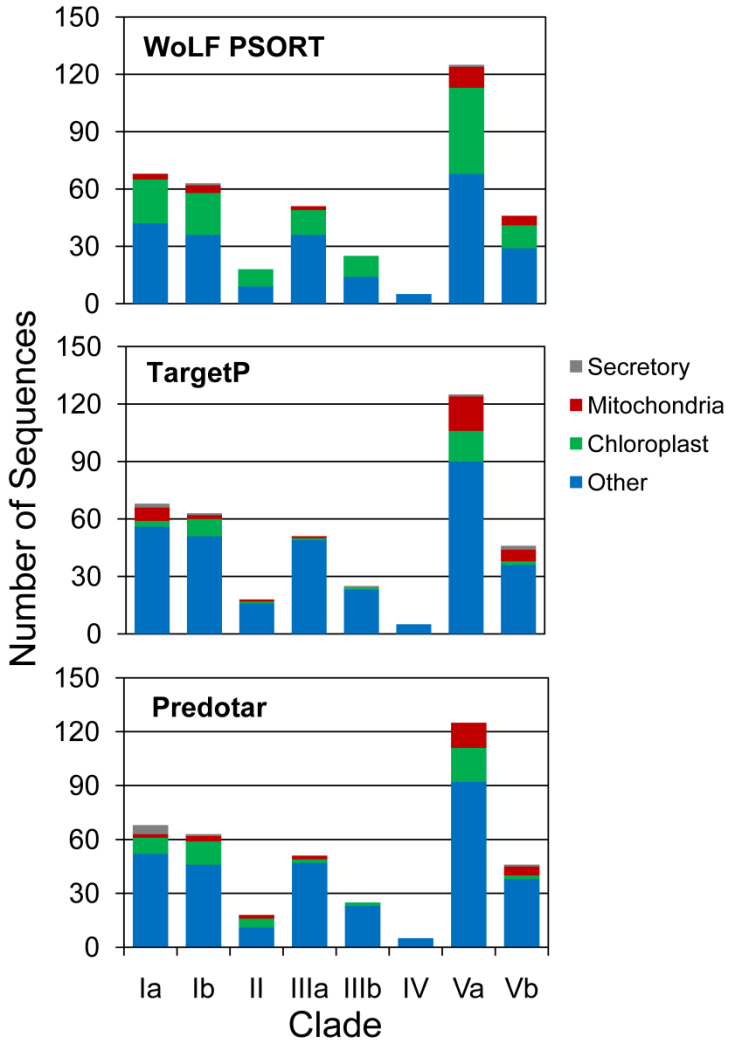

Figure 2.8: Analysis of BAHD Acyltransferase Protein Subcellular Localization

Each chart indicates the results from a different prediction algorithm, with the number of sequences indicated by the y-axis and clade indicated on the x-axis.

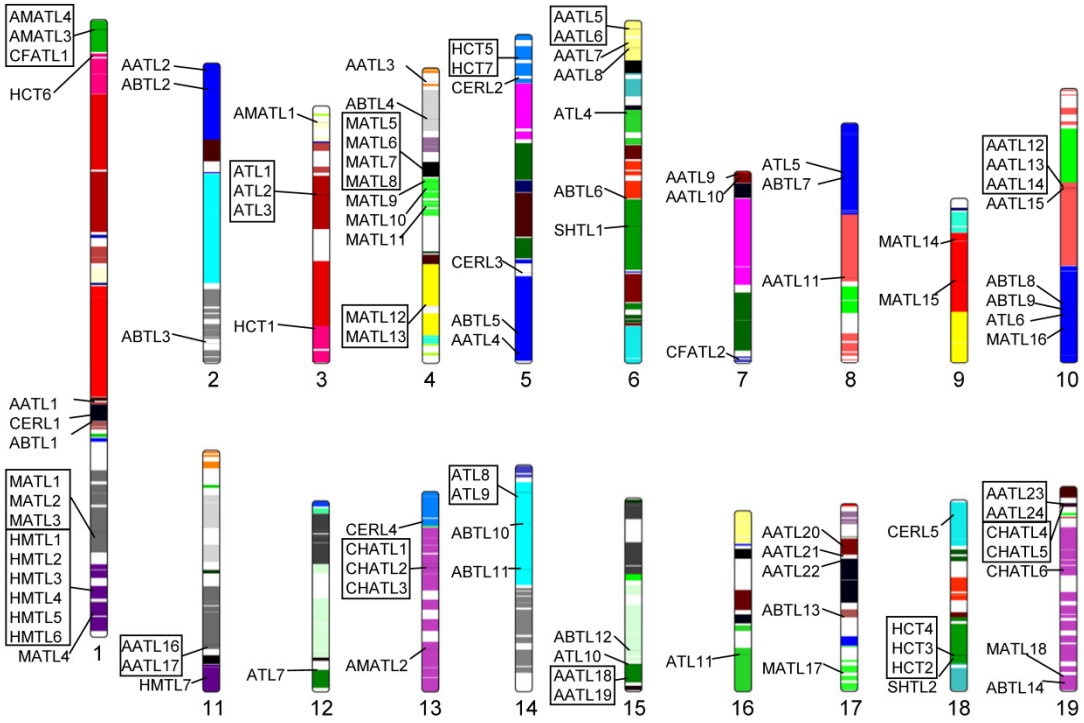

Figure 2.9: Locations of Putative *Populus* BAHD Acyltransferases on Linkage Groups

Homeologous blocks arising from the salicoid genome duplication event are color-coded across the nineteen linkage groups (chromosomes). BAHD acyltransferases in close proximity to one another are boxed for ease of labelling. Note that proximity on a linkage group does not, by itself, indicate a close phylogenetic relationship.

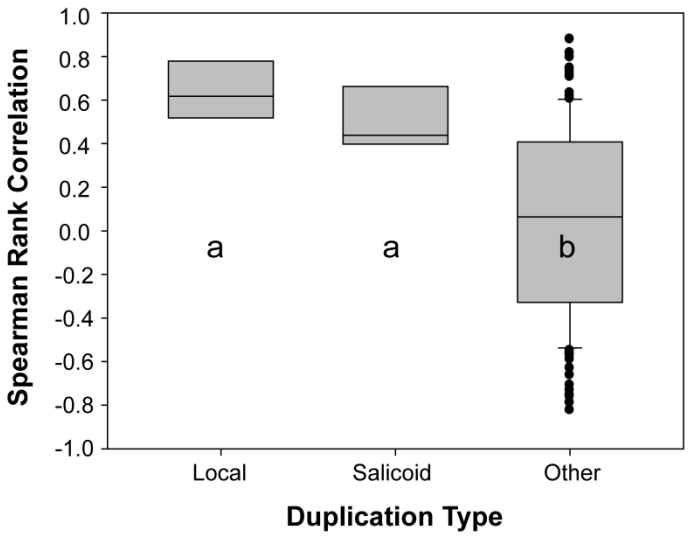

Figure 2.10: Pairwise Gene Expression Correlation Across *Populus* BAHD Acyltransferase Duplication Types

Box plots for Spearman rank correlations of pairwise gene expression by clade across all microarray experiments. Gene pairs are grouped by their association with local duplication, salicoid duplication, or others (all other pairwise combinations). Categories with the same letter had median correlation values that were not significantly different at $\alpha=0.05$ according to Dunn's Multiple Comparison test.

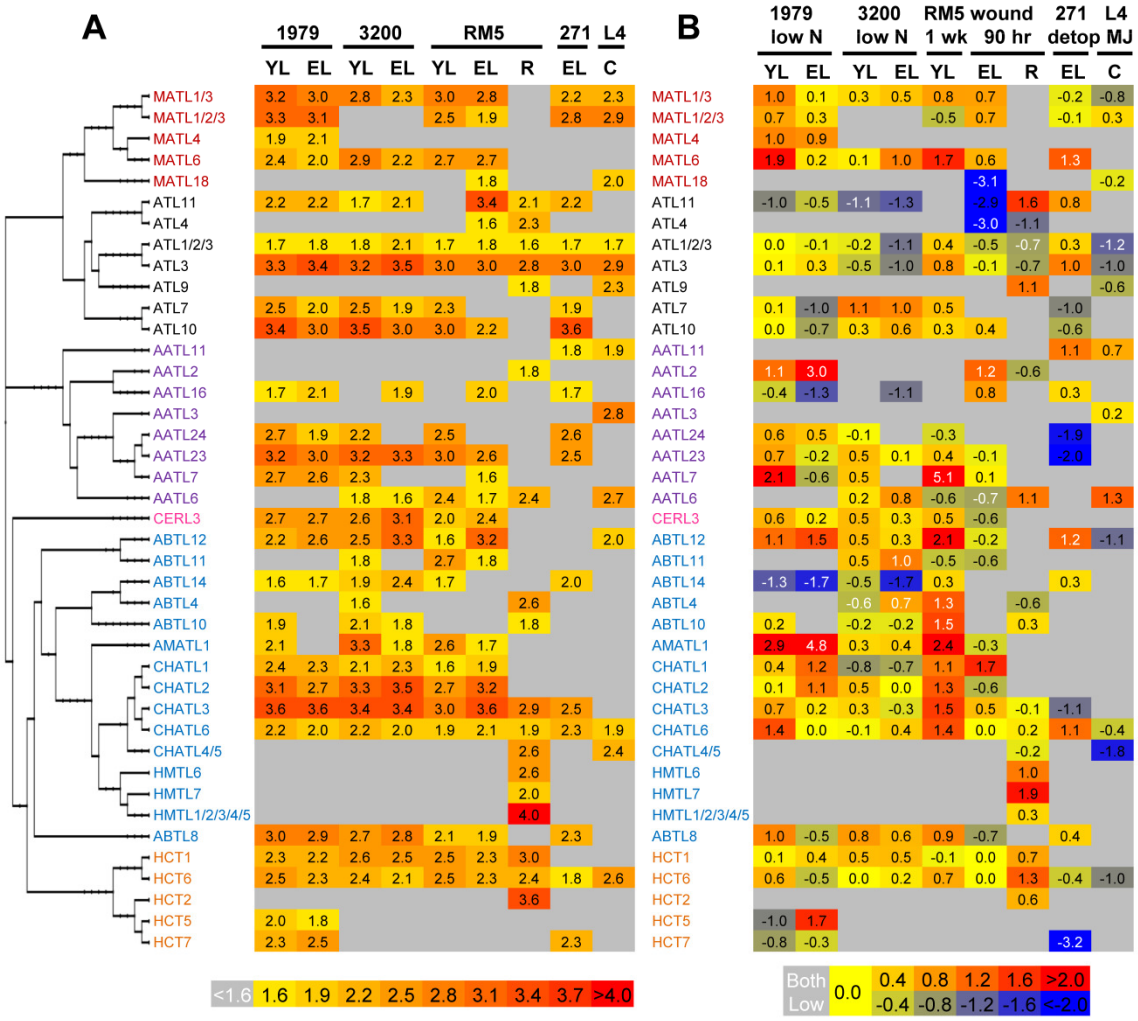

Figure 2.11: Expression of BAHD Acyltransferases in *Populus* Tissues, Organized by Phylogenetic Relationship

A: Expression of BAHD acyltransferase genes across tissues and genotypes. B: Stress responses of BAHD acyltransferase gene expression across tissues and genotypes. Expression data or ratios (stressed vs. control samples) were log-transformed and visualized in heatmaps (see **Methods**). Genes are organized by phylogenetic relationship and labelled by the clade color in Figure 2.1. Genotypes analyzed included: *P. fremontii* x *angustifolia* clones 1979, 3200, and RM5, and *P. tremuloides* clones 271 and L4. Tissues analyzed included: young leaf (YL), expanding leaf (EL), root tips (R), and

suspension cell cultures (C). Stress treatments included: nitrogen limitation (low N), leaf wounding (wound, sampled either 1 week or 90 hours after wounding), removal of shoot up to leaf plastochron index three (detop, 90 hours after removal), and methyl jasmonate elicitation (MJ, Yuan et al. 2009). White text indicates that raw hybridization intensity for either control (for upregulated genes) or stressed treatment (for downregulated genes) samples was below the quantitation limit (see **Methods**).

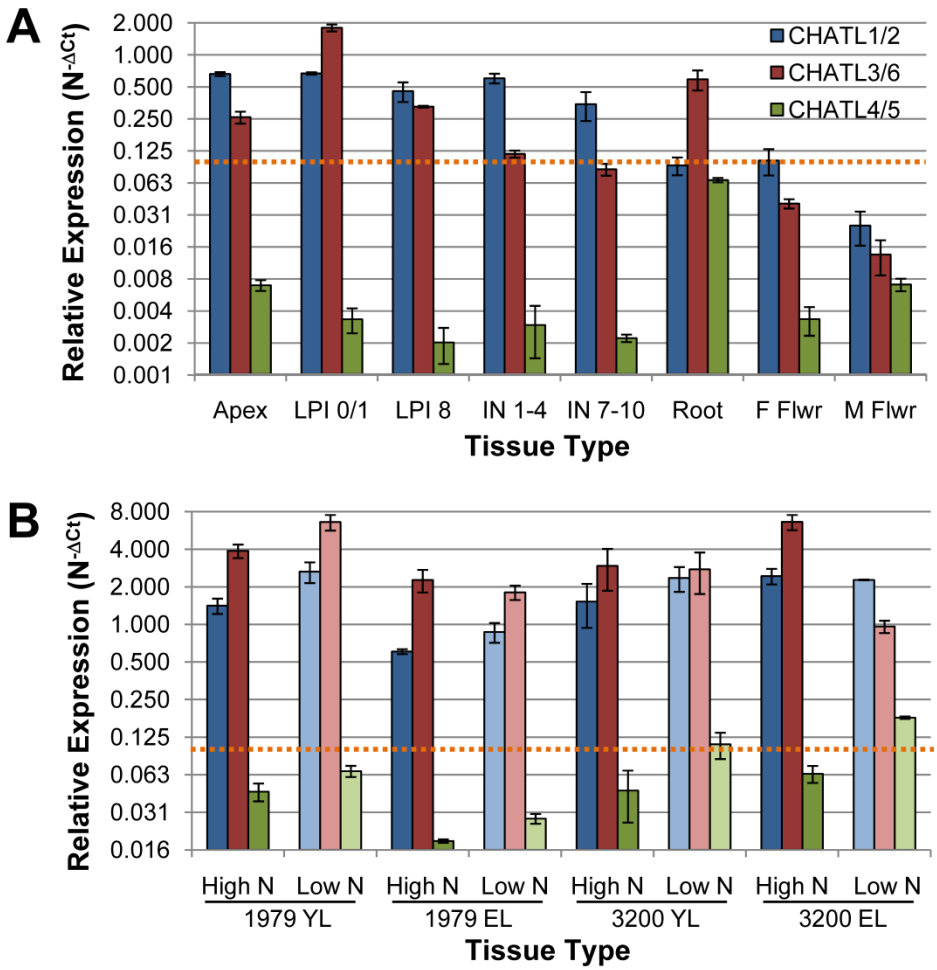

Figure 2.12: QPCR Expression Analysis of *Populus CHATL* Genes

A: Relative expression of the highly similar *CHATL1/2*, *CHATL3/6* and *CHATL4/5* gene pairs in various *P. tremuloides* tissues. Data represent means ± SE of three biological replicates. Tissues examined included apical bud/leaves (Apex), young leaves (LPI 0/1), mature leaves (LPI 8), internodes 1-4 (IN 1-4) and 7-10 (IN 7-10), root tips (Root), female flowers (F Flwr), and male flowers (M Flwr). Dashed orange line indicates an expression level comparable to the presence vs. absence cutoff used in microarray analysis. B: Relative expression of *CHATL* genes in young (YL) and expanding (EL) leaves from the nitrogen stress experiment. Data represent means ± SD

of two biological replicates. Genotypes are listed as in Figure 2.11, with "High N" samples corresponding to non-stressed tissues in Figure 2.11A.

References

BAR Heatmapper Plus Tool. [http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi]

Bayer A, Ma X, Stöckigt J (2004) Acetyltransfer in natural product biosynthesis--functional cloning and molecular analysis of vinorine synthase. *Bioorganic & Medicinal Chemistry* 12: 2787-2795

Berger A, Meinhard J, Petersen M (2006) Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. *Planta* 224: 1503-1510

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding *in vivo* benzenoid metabolism in Petunia petal tissue. *Plant Physiology* 135: 1993-2011

Böhringer S, Godde R, Böhringer D, Schulte T, Epplen JT (2002) A software package for drawing ideograms automatically. *Online Journal of Bioinformatics* 1: 51-61

Bowles D, Lim E-K, Poppenberger B, Vaistij FE (2006) Glycosyltransferases of lipophilic small molecules. *Annual Review of Plant Biology* 57: 567-597

Cao P-J, Bartley LE, Jung K-H, Ronald PC (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. *Molecular Plant* 1: 858-877

Chen F, Liu C-J, Tschaplinski TJ, Zhao N (2009) Genomics of secondary metabolism in *Populus*: Interactions with biotic and abiotic environments. *Critical Reviews in Plant Sciences* 28: 375 - 392

Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E (2009) The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. *BMC Plant Biology* 9: 30

Constabel CP, Lindroth RL (2010) The impact of genomics on advances in herbivore defense and secondary metabolism in *Populus*. In: Jansson S, Bhalerao R, Groover A (eds) *Genetics and Genomics of Populus*. Springer New York, pp 279-305

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Research* 14: 1188-1190

D'Auria JC (2006) Acyltransferases in plants: A good time to be BAHD. *Current Opinion in Plant Biology* 9: 331-340

D'Auria JC, Chen F, Pichersky E (2002) Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. *Plant Physiology* 130: 466-476

D'Auria JC, Pichersky E, Schaub A, Hansel A, Gershenzon J (2007a) Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (*Z*)-3-hexen-1-yl acetate in *Arabidopsis thaliana*. *The Plant Journal* 49: 194-207

D'Auria JC, Reichelt M, Luck K, Svatos A, Gershenzon J (2007b) Identification and characterization of the BAHD acyltransferase malonyl-CoA:anthocyanidin 5-*O*-glucoside-6"-*O*-malonyltransferase (At5MAT) in *Arabidopsis thaliana*. *FEBS Letters* 581: 872-878

Dexter R, Qualley A, Kish CM, Ma CJ, Koeduka T, Nagegowda DA, Dudareva N, Pichersky E, Clark D (2007) Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *The Plant Journal* 49: 265-275

Dhaubhadel S, Farhangkhoee M, Chapman R (2008) Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. *Journal of Experimental Botany* 59: 981-994

Dudareva N, D'Auria JC, Nam KH, Raguso RA, Pichersky E (1998) Acetyl-CoA:benzylalcohol acetyltransferase – an enzyme involved in floral scent production in *Clarkia breweri*. *The Plant Journal* 14: 297-304

El-Sharkawy I, Manríquez D, Flores F, Regad F, Bouzayen M, Latché A, Pech J-C (2005) Functional characterization of a melon alcohol acyl-transferase gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity. *Plant Molecular Biology* 59: 345-362

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* 2: 953-971

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

Greenaway W, English S, May J, Whatley FR (1991) Chemotaxonomy of section *Leuce* poplars by GC-MS of bud exudate. *Biochemical Systematics and Ecology* 19: 507-518

Greenaway W, English S, Whatley FR (1992) Relationships of *Populus* x *acuminata* and *Populus* x *generosa* with their parental species examined by gas chromatography - mass spectrometry of bud exudates. *Canadian Journal of Botany* 70: 212-221

Grienenberger E, Besseau S, Geoffroy P, Debayle D, Heintz D, Lapierre C, Pollet B, Heitz T, Legrand M (2009) A BAHD acyltransferase is expressed in the tapetum of Arabidopsis anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *The Plant Journal* 58: 246-259

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98

Han Y, Gasic K, Korban SS (2007) Multiple-copy cluster-type organization and evolution of genes encoding *O*-methyltransferases in the apple. *Genetics* 176: 2625-2635

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology* 148: 993-1003

Hoffman L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2005) Acyltransferase-catalysed *p*-coumarate ester formation is a committed step of lignin biosynthesis. *Plant Biosystems* 139: 50-53

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: Protein localization predictor. *Nucleic Acids Research* 35: W585-W587

Horton P, Park K-J, Obayashi T, Nakai K (2006) Protein subcellular localization prediction with WoLF PSORT. In: Jiang T, Yang U-C, Chen Y-P, Wong L (eds) *Asia Pacific Bioinformatics Conference APBC06*. Imperial College Press, Taipei, Taiwan, pp 39-48

Huson D, Richter D, Rausch C, Dezulian T, Franz M, Rupp R (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8: 460

Jaillon O, et a (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463-467

JGI *Populus trichocarpa* genome FTP site. [ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v5.0/Ptrichocarpa]

Katoh K, Misawa K, Kuma Ki, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059-3066

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286-298

Kopycki JG, Rauh D, Chumanevich AA, Neumann P, Vogt T, Stubbs MT (2008) Biochemical and structural analysis of substrate promiscuity in plant Mg^{2+}-dependent *O*-methyltransferases. *Journal of Molecular Biology* 378: 154-164

La D, Livesay D (2005a) Predicting functional sites with an automated algorithm suitable for heterogeneous datasets. *BMC Bioinformatics* 6: 116

La D, Livesay DR (2005b) MINER: Software for phylogenetic motif identification. *Nucleic Acids Research* 33: W267-W270

La D, Sutch B, Livesay DR (2005) Predicting protein functional sites with phylogenetic motifs. *Proteins: Structure, Function, and Bioinformatics* 58: 309-320

Lam KC, Ibrahim RK, Behdad B, Dayanandan S (2007) Structure, function, and evolution of plant *O*-methyltransferases. *Genome* 50: 1001-1013

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948

Li D, Xu Y, Xu G, Gu L, Li D, Shu H (2006) Molecular cloning and expression of a gene encoding alcohol acyltransferase (MdAAT2) from apple (cv. Golden Delicious). *Phytochemistry* 67: 658-667

Luo J, Fuell C, Parr A, Hill L, Bailey P, Elliott K, Fairhurst SA, Martin C, Michael AJ (2009) A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in Arabidopsis seed. *The Plant Cell* 21: 318-333

Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y, Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C (2007) Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *The Plant Journal* 50: 678-695

Ma X, Koepke J, Panjikar S, Fritzsch G, Stockigt J (2005) Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *Journal of Biological Chemistry* 280: 13576-13583

Miller MA, Holder MT, Vos R, Midford PE, Liebowitz T, Chan L, Hoover P, Warnow T (2009) CIPRES (Cyberinfrastructure for Phylogenetic Research). [http://www.phylo.org/sub_sections/portal]

Molina I, Li-Beisson Y, Beisson F, Ohlrogge JB, Pollard M (2009) Identification of an Arabidopsis feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiology* 151: 1317-1328

Negruk V, Yang P, Subramanian M, McNevin JP, Lemieux B (1996) Molecular cloning and characterization of the *CER2* gene of *Arabidopsis thaliana*. *The Plant Journal* 9: 137-145

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology* 22: 746-754

Okada T, Hirai MY, Suzuki H, Yamazaki M, Saito K (2005) Molecular characterization of a novel quinolizidine alkaloid *O*-tigloyltransferase: cDNA cloning, catalytic activity of recombinant protein and expression analysis in *Lupinus* plants. *Plant and Cell Physiology* 46: 233-244

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR rice genome annotation resource: Improvements and new features. *Nucleic Acids Research* 35: D883-D887

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* 339: 62-66

Retzel EF, Johnson JE, Crow JA, Lamblin AF, Paule CE (2008) Legume resources: MtDB and Medicago.org. In: Edwards D (ed) *Plant Bioinformatics: Methods and Protocols*. Humana Press, Totowa, NJ, pp 261-274

Richardson PM, Young DA (1982) The phylogenetic content of flavonoid point scores. *Biochemical Systematics and Ecology* 10: 251-255

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4: 1581-1590

Souleyre EJF, Greenwood DR, Friel EN, Karunairetnam S, Newcomb RD (2005) An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. *FEBS Journal* 272: 3132-3144

St. Pierre B, De Luca V (2000) Evolution of acyltransferase genes: Origin and diversification fo the BAHD superfamily of acyltransferases involved in secondary metabolism. In: Romeo JT, Ibrahim R, Varin L, De Luca V (eds) *Recent Advances in Phytochemistry, Vol 34*. Elsevier, pp 285-315

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758-771

Sullivan ML (2009) A novel red clover hydroxycinnamoyl transferase has enzymatic activities consistent with a role in phaselic acid (2-*O*-[caffeoyl]-L-malate) biosynthesis. *Plant Physiology*: 1866-1879

Suzuki H, Nakayama T, Nishino T (2003) Proposed mechanism and functional amino acid residues of malonyl-CoA:anthocyanin 5-*O*-glucoside-6'''-*O*-malonyltransferase from flowers of *Salvia splendens*, a member of the versatile plant acyltransferase family. *Biochemistry* 42: 1764-1771

Suzuki H, Nishino T, Nakayama T (2007) cDNA cloning of a BAHD acyltransferase from soybean (*Glycine max*): Isoflavone 7-*O*-glucoside-6''-*O*-malonyltransferase. *Phytochemistry* 68: 2035-2042

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): Gene structure and function annotation. *Nucleic Acids Research* 36: D1009-D1014

Tacke E, Korfhage C, Michel D, Maddaloni M, Motto M, Lanzini S, Salamini F, Döring H-P (1995) Transposon tagging of the maize *Glossy2* locus with the transposable element *En/Spm*. *The Plant Journal* 8: 907-917

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596-1599

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research* 18: 1944-1954

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The Botany Array Resource: e-northerns, expression angling, and promoter analyses. *The Plant Journal* 43: 153-163

Tsai C-J, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytologist* 172: 47-62

Tsai C, Cseke L, Harding S (2003) Isolation and purification of RNA. In: Cseke L, Kaufman P, Podila G, Tsai C (eds) *Handbook of Molecular and Cellular Methods in Biology and Medicine*. CRC Press, Boca Raton, FL, pp 25-44

Tsai CJ, Ranjan P, DiFazio SP, Tuskan GA, Johnson V (2011) Poplar genome microarrays. In: Joshi CP, DiFazio SP, Kole C (eds) *Genetics, Genomics and Breeding of Crop Plants: Poplar*. Science Publishers, Enfield, NH, pp 112-127

Tuskan GA, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604

Unno H, Ichimaida F, Suzuki H, Takahashi S, Tanaka Y, Saito A, Nishino T, Kusunoki M, Nakayama T (2007) Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *Journal of Biological Chemistry* 282: 15812-15822

Wagner A, Ralph J, Akiyama T, Flint H, Phillips L, Torr K, Nanayakkara B, Te Kiri L (2007) Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase in *Pinus radiata*. *Proceedings of the National Academy of Sciences* 104: 11856-11861

Wang J, De Luca V (2005) The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. *The Plant Journal* 44: 606-619

Xia Y, Nikolau BJ, Schnable PS (1996) Cloning and characterization of *CER2*, an *Arabidopsis* gene that affects cuticular wax accumulation. *Plant Cell* 8: 1291-1304

Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y (2010) Evolution and function of the plant cell wall synthesis-related Glycosyltransferase Family 8. *Plant Physiology* 153: 1729-1746

Yoshihara N, Imayama T, Matsuo Y, Fukuchi-Mizutani M, Tanaka Y, Ino I, Yabua T (2006) Characterization of cDNA clones encoding anthocyanin 3-*p*-coumaroyltransferase from *Iris hollandica*. *Plant Science* 171: 632-639

Yu X-H, Chen M-H, Liu C-J (2008) Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-*O*-glycosidic (iso)flavones in *Medicago truncatula*. *The Plant Journal* 55: 382-396

Yu X-H, Gou J-Y, Liu C-J (2009) BAHD superfamily of acyl-CoA dependent acyltransferases in *Populus* and *Arabidopsis*: Bioinformatics and gene expression. *Plant Molecular Biology* 70: 421-442

Yuan Q, Ouyang S, Wang A, Zhu W, Maiti R, Lin H, Hamilton J, Haas B, Sultana R, Cheung F, Wortman J, Buell CR (2005) The Institute for Genomic Research Osa1 rice genome annotation database. *Plant Physiology* 138: 18-26

Yuan Y, Chung J-D, Fu X, Johnson VE, Ranjan P, Booth SL, Harding SA, Tsai C-J (2009) Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in *Populus* and *Arabidopsis*. *Proceedings of the National Academy of Sciences* 106: 22020-22025

Zheng Z, Qualley A, Fan B, Dudareva N, Chen Z (2009) An important role of a BAHD acyl transferase-like protein in plant innate immunity. *The Plant Journal* 57: 1040-1053

CHAPTER 3.

PERTURBING PHENYLPROPANOID METABOLISM IN *POPULUS* CELL CULTURES USING METABOLIC INHIBITORS AND A DEFENSE ELICITOR

Summary

Phenylpropanoid metabolism is intimately linked with both primary carbon metabolism and nitrogen metabolism due to its biosynthetic origin from the amino acid phenylalanine. In *Populus*, phenylpropanoid metabolism leads to diverse downstream small molecules such as flavonoids, salicinoids, and benzenoids, as well as to major polymeric carbon sinks such as lignin and condensed tannins. We utilized metabolic inhibitors to perturb the first three steps of the core phenylpropanoid pathway, with or without concomitant feeding of the elicitor methyl jasmonate (MeJA), in *Populus* cell suspension cultures to investigate the transcriptional and metabolic consequences of perturbing this pathway. Profiles of core phenylpropanoid metabolites, their glycosylated forms, and putative derivatives, such as arbutin and 3,4-dihydroxybenzoate, in perturbed cells provided evidence for branch points of the pathway due to metabolite diversion. The downstream pathways leading to synthesis of flavonoids, condensed tannins, and lignin showed differential responses to perturbation based on metabolite and gene expression analyses. Feeding of the PAL inhibitor AOPP led to increases in protein, phenlyalanine and several other amino acids, with a particularly strong increase in tyrosine, while MeJA led to reduced protein levels. Expression of nitrate and

ammonium-utilizing genes also responded to phenylpropanoid perturbation, suggesting shifts in inorganic nitrogen utilization contingent on phenylpropanoid biosynthetic demand. The citric acid cycle metabolites succinate and citrate were differentially influenced by MeJA, while AOPP led to increased α-ketoglutarate and malonate. Interestingly, both treatments led to upregulaton of pyruvate dehydrogenase and ATP citrate lyase genes, possibly indicating an indirect influence of phenylpropanoid demand on fatty acid metabolism. The analysis overall supports a model of phenylalanine as a key mediator between phenylpropanoid and amino acid metabolism, both of which can, in turn, influence primary carbon metabolism.

Background

As a secondary metabolic pathway common to vascular plants, phenylpropanoid metabolism is an area of interest to a wide range of plant biologists. The core phenylpropanoid pathway supports multiple downstream branch pathways. For example, an important structural polymer derived from the core pathway is lignin, which imparts cell wall rigidity and hydrophobicity. Lignin is thought to be one of the major evolutionary innovations allowing plants to sustain upright growth while maintaining effective water transport, thus enhancing ability to compete for light after colonizing the land (Kenrick and Crane 1997). Other branch pathways are known to have roles in abiotic and biotic interactions, such as the importance of flavonoids in protection against ultraviolet radiation (reviewed by Bornman et al. 1997) or the roles of phenolic glycosides, condensed and possibly hydrolysable tannins, or coumarins in defense against

herbivores (Bernards and Båstrup-Spohr 2008; Hagerman et al. 1992; Hemming and Lindroth 1995; Roslin and Salminen 2008; Tahvanainen et al. 1985). Volatile benzoates and colorful anthocyanins also play roles in the attraction of pollinators via floral scents and pigmentation, respectively (reviewed by Knudsen and Gershenzon 2006; Miller et al. 2011).

The diversity of vascular plant taxa brings with it a diversity of corresponding evolved secondary metabolic processes that mediate a variety of abiotic and biotic interactions and are dependent upon myriad factors. Within the Salicaceae, phenylpropanoids are by far the dominant secondary metabolites, in terms of both their constitutive quantities (Julkunen-Tiitto 1986; Lindroth and Hwang 1996) and the diversity of their chemical structures (Chen et al. 2009; Greenaway et al. 1991; Greenaway et al. 1992; Tsai et al. 2006b) within and among species in the family. Of particular note is the exclusive occurrence of "salicinoids," a subset of phenolic glycosides, in the *Populus* and *Salix* genera (Boeckler et al. 2011; Julkunen-Tiitto 1986) thought to be derived from the phenylpropanoid pathway (Babst et al. 2010; Tsai et al. 2006b). Phenylpropanoid metabolism exhibits plasticity at multiple levels (reviewed by Bernards and Båstrup-Spohr 2008; Dixon and Paiva 1995). Differences in phenylpropanoid levels can arise in connection with seasonality or developmental cues, with shifts that may be genotype-dependent (Donaldson et al. 2006; Harding et al. 2009; Rehill et al. 2006). Different branch pathways are known to take precedence in a tissue-dependent manner, with, for example, extensive lignification in woody tissues and higher levels of non-structural phenylpropanoids in leaf and root tissues (reviewed by Tsai et al. 2006a). Phenylpropanoid metabolism is inducible under a number of biotic (e.g.,

herbivory) and abiotic (e.g., reduced nitrogen availability, wounding, or increased ultraviolet radiation) stressors (reviewed by Bornman et al. 1997; Chen et al. 2009), with both local and systemic induction reported at the transcriptional level (Babst et al. 2009).

A hierarchical model proposed by Koricheva et al. (1998) for the regulation of carbon-based secondary metabolism provides one possible conceptual framework for analysis of carbon-based secondary metabolism in woody plants. Meta-analysis supports the idea that environmental factors like the availability of resources ("upper" hierarchy levels according to Koricheva et al. 1998), such as light, carbon dioxide, and soil nitrogen, can act as local, unsurmountable limits on the internal metabolic demands ("lower" hierarchy levels) inherent to maintenance, growth, reproduction, and adaptation to multiple, dynamic biotic and abiotic factors. In accordance with a multi-level model, some structurally similar phenylpropanoid derivatives in *Betula pendula* (silver birch), such as different types of quercetin glycosides, myricetin glycosides, or hydroxycinnamoylquinic acids, exhibit differential responses to fertilization (Keinänen et al. 1999). For broad categories of phenylpropanoids, these patterns support previous findings and older models for understanding regulation of carbon-based secondary metabolism such as the carbon nutrient balance hypothesis (Bryant et al. 1983). However, this hypothesis primarily applies at the gross scale, neither generating specific predictions for nor extending to the level of individual metabolites (Keinänen et al. 1999; Koricheva et al. 1998). Such findings emphasize the relevance of multi-level models for understanding metabolic regulation. Particularly important for generating such understanding are models that incorporate finer-scale views of metabolic pathways. Metabolite profiling approaches could be particularly valuable in building an

understanding of partitioning among phenylpropanoid branch pathways, among individual metabolites within those pathways, and between primary and secondary metabolism.

Phenylpropanoid metabolism both contributes to the carbon component of biomass and responds to relative carbon-nitrogen status (Fritz et al. 2006; Matt et al. 2002). The compounds are linked to primary carbon metabolism via the shikimate pathway, which is sourced from erythrose-4-phosphate via the pentose phosphate pathway and glycolysis (Herrmann 1995). Accordingly, one study of induced phenylpropanoid defense responses in *Populus* showed that upregulation of phenylpropanoid biosynthetic gene expression was accompanied by changes in expression of genes linked to organic acid and carbohydrate metabolism, glycolysis, and carbohydrate transport (Babst et al. 2009). Evidence suggests that defense-induced accumulation of phenylpropanoids in developing leaves of *Populus* arises partly from new photosynthate and carbohydrates imported from source leaves (Arnold et al. 2004; Arnold and Schultz 2002; new photosynthate also generates defensive compounds in tobacco according to Hanik et al. 2010). Recent work has also demonstrated that existing carbohydrates are transported from leaves towards the lower stem and roots for storage as part of a foliar defense response in *Populus* (Babst et al. 2005), potentially setting up competition among different carbon sinks. In addition to directly acting as a carbon sink, phenylpropanoids are connected to primary metabolism through glycosylation (Payyavula et al. 2009; reviewed by Vaistij et al. 2009), a process thought to increase stability, improve metabolic channeling, and help limit autotoxicity of defense compounds *in planta* (Rea 2007; Vaistij et al. 2009; Vogt and Jones 2000; Yazaki 2005).

Phenylpropanoid metabolism is also intimately connected with nitrogen metabolism via phenylalanine, an entry point to the phenylpropanoid pathway. Phenylalanine is converted to *trans*-cinnamic acid and ammonia by the action of phenylalanine ammonia lyase (PAL; reviewed by Camm and Towers 1973). PAL may also act on tyrosine as an alternative starting point for phenylpropanoid metabolism to yield directly *p*-coumaric acid in some species, bypassing the cinnamic acid intermediate (reviewed by Tzin and Galili 2010). In either case, the first committed step of phenylpropanoid biosynthesis represents a disarticulation point for carbon and nitrogen metabolism. This forms the basis of the protein competition model, an alternative model for predicting carbon allocation to phenylpropanoid metabolism (Jones and Hartley 1999). The model draws on empirical evidence to suggest that phenylalanine, rather than carbon, is limiting to both protein and phenolic biosynthesis, generating a trade-off between these two competing sinks. However, this model discounts the liberation of inorganic nitrogen via PAL. Given the large amounts of structural phenylpropanoids generated in woody species in general, and of non-structural phenylpropanoids in *Populus* in particular, potentially large pools of inorganic nitrogen can – or arguably, must -- be reassimilated into amino acid pools via glutamine and glutamate (reviewed by Cantón et al. 2005; Razal et al. 1996). Furthermore, increased substrate demand at any point in the phenylpropanoid core or branch pathways should lead to concomitant changes in demand for phenylalanine or tyrosine and release of additional inorganic nitrogen, thereby constituting a metabolic loop between phenylpropanoid and amino acid metabolism. This feed-forward propagation from phenylpropanoid sinks to amino acid pools could potentially act as a nitrogen source rather than competitive sink for amino

acid biosynthesis. Thus, the link between phenylpropanoid and nitrogen metabolism may not be exclusively competitive in nature.

Research on the molecular regulation of phenylpropanoid metabolism has been facilitated by various inductive treatments, taking advantage of the highly responsive nature of phenylpropanoids to biotic and abiotic factors. Jasmonates have been a particularly useful tool for understanding the phenylpropanoid pathway. The ability of these signaling molecules to to activate an array of context-dependent biotic defense responses in diverse plant taxa is well documented (Ballaré 2011; Gundlach et al. 1992; Kessler and Baldwin 2002; Sun et al. 2011). At the metabolic level, jasmonates act as broad-spectrum elicitors of secondary metabolism, stimulating phenylpropanoid as well as glucosinolate, terpenoid, and alkaloid biosynthesis (Gundlach et al. 1992; Reymond and Farmer 1998; Zhao et al. 2005). While defense chemicals are widely thought to have significant ecological and physiological costs, these costs are known to be strongly mediated by genetic background (Bergelson and Purrington 1996; Strauss et al. 2002). Evidence for direct growth tradeoffs resulting from defenses induced by jasmonates has also been more limited than for more indirect, ecological tradeoffs (Strauss et al. 2002). In one set of experiments demonstrating the power of integrative transcript and metabolic profiling, methyl jasmonate (MeJA) treatment of *Arabidopsis* cell suspension cultures has supported the existence of a growth-defense tradeoff while providing new insight into its mechanistic underpinnings (Pauwels et al. 2008). In this work, MeJA led to transcriptional activation of phenylpropanoid metabolism, especially monolignol synthesis, concurrent with but independent of transcriptional repression of cell cycle progression (Pauwels et al. 2008). Such observations suggest the possibility that growth

and defense responses to MeJA may be amenable to some amount of decoupling by methods such as mutation of key promoters.

Enzyme inhibitors have also been used to aid investigation of the phenylpropanoid pathway. Examples include inhibitors for the first three enzymatic steps of the phenylpropanoid metabolism (referred to as the core phenylpropanoid pathway), catalyzed sequentially by PAL, cinnamate 4-hydroxylase (C4H) and 4-coumarate:CoA ligase (4CL). The compound α-aminooxy-β-phenylpropionic acid (AOPP) is an aminooxy analogue of phenylalanine previously shown to inhibit PAL activity *in vitro* as well as *in vivo* in feeding experiments with cell suspension cultures, excised tissues, and intact plantlets (Amrhein and Gödeke 1977; Amrhein et al. 1976; Havir 1981). However, since PAL is negatively regulated by its product cinnamic acid, constituting a negative feedback loop, AOPP can also promote PAL activity *in vivo* via transcriptional upregulation resulting from reduced cinnamic acid levels (Amrhein and Gerhardt 1979; Bolwell et al. 1988; Mavandad et al. 1990; Orr et al. 1993). Piperonylic acid (PIP) is a structural analog of cinnamic acid and functions as a mechanism-based quasi-irreversible inhibitor of C4H (Schalk et al. 1998), which catalyzes the conversion of cinnamic acid to *p*-coumaric acid. Although *in vitro* analysis revealed that PIP inhibition of C4H activity may be reversed by high (100 mM, likely non-physiological) levels of cinnamic acid, both *in vitro* and *in vivo* tests suggest that the reversal, when it occurs at all, is much slower than the formation of the enzyme-inhibitor complex (Schalk et al. 1998; Schoch et al. 2002). PIP has little effect on other cytochrome P450 enzymes (Schalk et al. 1998), further supporting its utility as a C4H-specific inhibitor. Methylenedioxycinnamic acid (MDCA) is structurally similar to hydroxycinnamates and has been shown to act as a

competitive inhibitor of 4CL *in vitro* (Funk and Brodelius 1990). 4CL catalyzes the activation of hydroxycinnamates to their corresponding CoA-thioesters for biosynthesis of monolignols and flavonoids (Hahlbrock and Grisebach 1979), with different isoforms exhibiting variations in substrate specificity . Accordingly, MDCA feeding led to reduced accumulation of lignin in both cell suspension cultures of *Vanilla* and hairy roots of *Daucus*, with a concomitant increase of hydroxybenzoates (Funk and Brodelius 1990; Sircar and Mitra 2009).

While metabolic inhibitors may confer non-specific effects, they complement genetics-based inactivation strategies (e.g., mutant or transgenic gene silencing) and provide several advantages. Since metabolic inhibitors act via direct interaction with enzymes, they are likely to be less discriminating among isoforms than methods based on gene sequences (Schoch et al. 2002). Minor variations in AOPP binding (Havir 1981) and MDCA effects on crude 4CL activity (compare Funk and Brodelius 1990; Sircar and Mitra 2009) have still been observed across species, indicating that differences in isoform discrimination by metabolic inhibitors may remain, although this is still likely to be more subtle than for genetic approaches. A second feature of metabolic inhibitor studies is the greatly reduced time for experimental tissue generation. This consideration is significant for research in long-lived species, such as *Populus*, that require months of for transgenic regeneration and for which mutagenesis studies are logistically difficult. Metabolic inhibitors are particularly powerful when coupled with cell suspension cultures that are highly amenable to feeding manipulation (e.g., Payyavula et al. 2009), and such systems lend themselves well to omics methodologies that permit global assessment of responses to metabolic perturbation.

In this chapter, we aim to characterize the transcriptional and metabolic responses of *Populus tremuloides* cell cultures following chemical perturbation of the core phenylpropanoid pathway using PAL, C4H, or 4CL inhibitors in conjunction with the defense elicitor MeJA. We expected that MeJA feeding would stimulate phenylpropanoid metabolism via transcriptional activation of phenylpropanoid biosynthetic genes, in keeping with previous knowledge of the mechanistic function of the elicitor. Meanwhile, we hypothesized that feeding metabolic inhibitors would result in reduced product accumulation and increased substrate levels for the targeted enzymes. Furthermore, successful inhibition was expected to lead not only to buildup of substrate for the targeted enzyme but also to increased levels of branch pathway metabolites derived from that substrate. This effect was expected to vary depending on the inhibited pathway step, with effects likely to decrease with pathway "distance". Correspondingly, enzyme products and downstream pathway metabolites would be decreased due to inhibition, with effects again attenuating with distance from the inhibited step. These expectations were partly an oversimplified, null hypothesis approach in that metabolic feedback and feed-forward effects are discounted. Data showing major deviations from these expectations can therefore help indicate the possible existence of feedback/feed-forward loops while also helping to identify branch locations for the synthesis of metabolites with poorly understood relationships to the core phenylpropanoid pathway.

Treatments combining individual enzyme inhibitors with MeJA were predicted to show even more dramatic accumulation of the enzyme's substrate than either compound fed by itself, because elicitation should lead to increased substrate levels that cannot be efficiently transformed to product by the inhibited enzyme. Correspondingly, decreased

product levels immediately downstream of the inhibited step were expected unless transcription of later pathway steps is responsive to reduced product levels. Combination treatments should lead to even stronger decreases in downstream metabolite levels in elicitor-activated pathways due to the lack of precursor replenishment, although the relative strength of inhibitor versus elicitor activity at the inhibited step could influence this outcome.

The results suggest that perturbing the early steps of phenylpropanoid metabolism influences not only relative proportions of hydroxycinnamic acids within the core pathway and downstream metabolites such as flavonoids, condensed tannin (CT), and benzoates, but also glycosylation patterns within these pools. Primary nitrogen and carbon metabolism, such as the levels of certain amino acids and citric acid cycle metabolites, were also affected. Several of the patterns identified at the metabolic level are supported by transcriptome data, which also reveal possible changes in inorganic nitrogen metabolism.

This work significantly advances a project initiated by a previous doctoral student in the Tsai lab, Raja Payyavula, who developed the *Populus* cell culture feeding trials. His preliminary trials established useful concentrations of the elicitor and inhibitors, based on quantitative cell viability measures and CT concentrations. He also generated the experimental cells that we used for metabolite and transcript profiling analysis presented in this chapter. We explicitly acknowledge in the text all work that Dr. Payyavula completed.

Methods

Populus Cell Cultures and Elicitor/Inhibitor Feeding Experiments

The cell culture feeding experiments were performed previously by Dr. Raja Payyavula. The information presented here is intended to provide context to the whole experimental design and necessary background for the analyses I carried out using these samples. The initiation and maintenance of a cell culture line established from leaf tissue of *Populus tremuloides* genotype L4 was described in Payyavula et al. (2009). Briefly, 5 mL samples of suspension cell cultures were subcultured every eleven days into 30 mL of fresh WPM liquid medium supplemented with 3% sucrose and 2.2 mg/L of 2,4-D. Growth was monitored where appropriate using replicate cultures maintained in Nephelo flasks. All cultures were maintained at 25^{o}C in the dark on an orbital shaker rotating at 120 rpm and continued under these conditions during experiments.

All feeding experiments were initiated five days post-subculture, previously found to fall within the log phase of culture growth. For Experiment 1, seven replicate flasks were generated for each of four treatment groups: Control, Elicitor, Inhibitor, and Inhibitor+Elicitor treatments. Control samples had 7.0 µl DMSO added to each flask at the initiation of the experimental period, followed by an additional 3.5 µl DMSO added 12 h after initiation, while Inhibitor treatment was achieved by adding 7.0 µl of 0.5 M piperonylic acid (PIP) in DMSO to each treated flask at the initiation of the experimental period, followed by another 3.5 µl of the inhibitor 12 h after initiation. PIP is a known inhibitor of C4H (Schalk et al. 1998). Elicitor treatment was the addition of 190 µl of 4.6 mM MeJA dissolved in DMSO to each treated flask 6 h after the initiation of the experimental period. Samples were harvested 48 h after initiation of the experimental

period. Cells were water-rinsed to remove residual culture medium, with excess liquid removed by vacuum filtration prior to flash-freezing cells in liquid nitrogen for storage at -80^{o}C. For Experiment 2, six replicate flasks were generated for each of six conditions: Control, Elicitor, PAL Inhibitor, 4CL Inhibitor, PAL Inhibitor+Elicitor, 4CL Inhibitor+Elicitor conditions. Control and Elicitor treatments were achieved as described above, with the exception that Control flasks had 8.75 µl of DMSO added at the start of the experimental period. PAL or 4CL Inhibitor treatments were achieved by the addition of 8.75 µl 0.2 M α-aminooxy-β-phenylpropionic acid (AOPP, Havir 1981) or methylenedioxycinnamic acid (MDCA, Funk and Brodelius 1990) in DMSO, respectively, at the start of the experimental period. Harvesting and post-harvesting procedures were the same as for Experiment 1. For both experiments, MeJA-treated cell cultures were maintained in a separate room from cultures not treated with MeJA to avoid cross-sample contamination of jasmonates, which are known volatile signal metabolites.

Quantification of Total Carbon and Nitrogen

Frozen cell cultures were ground under liquid nitrogen and portions of the ground material for all available biological replicates were freeze-dried overnight. Freeze-dried samples were weighed out to the microgram level into tin capsules in amounts between 0.800-1.900 mg. Quantification of total carbon and total nitrogen via micro-Dumas combustion was completed at the Stable Isotope & Soil Biology Laboratory within the Odum School of Ecology at the University of Georgia. The procedure generates total C and total N data on a percentage basis of total mass, so the amount of starting material is relatively unimportant as long as it is accurately quantified. Four biological replicates

were quantified for each treatment group in both experiments. Data for each experiment were analyzed separately using two-way ANOVA with additional *post hoc* testing in JMP when relevant, as described below for metabolic profiling data.

Quantification of Condensed Tannins

Dr. Payyavula determined total condensed tannins on a percent dry mass basis in preliminary cell culture trials as described in previous work (Harding et al. 2005; Payyavula et al. 2009; Porter et al. 1986). Data were analyzed using one-way or two-way ANOVA with *post hoc* testing in JMP as appropriate (described below).

Quantification of Lignin

Portions of freeze-dried cell cultures were washed twice in 95% ethanol, dried, weighed out in 3.0$\pm$1.0 mg samples, and shipped to the National Bioenergy Center Laboratory (U.S. Department of Energy National Renewable Energy Laboratory; Golden, CO) for determination of relative lignin content and ratio of syringyl to guaiacyl monolignols (S/G ratio) using pyrolysis molecular beam mass spectrometry (Sykes et al. 2010). For Experiment 1, two biological samples were pooled to generate a single replicate due to low quantities of available material. The remaining sample for the PIP and MeJA treatments were excluded, while an additional replicate was available for the Control treatment (a total N of 4 for that treatment and 3 for the others). For Experiment 2, each biological sample was used to generate a single replicate (a total N of 6 per treatment). One sample was lost in the MDCA treatment group, leaving an N of 5 for

that treatment. Data were statistically analyzed in JMP using two-way ANOVA and, where appropriate, *post hoc* testing (described below).

Quantification of Soluble Protein

Ground, freeze-dried cell culture samples of 2.0-5.2 mg each were individually weighed on an analytical balance to the 0.1 mg level. To each sample was added 1 mL of buffer containing 50 mM Tris (pH 7.0), 200 mM NaCl and 2% (w/v) PVP. Samples were vortexed briefly, sonicated at room temperature for 13 min, cooled for 2 min on ice, and centrifuged for 10 min to pellet the cells. A portion of supernatant was diluted to 10% strength and used in a Bradford assay (Bradford 1976; reagents from Bio-Rad). Absorbance was measured at 595 nm on a microplate reader. All samples were run in duplicate or triplicate, with 3-4 biological replicates per treatment for Experiment 1 and 6 biological replicates per treatment for Experiment 2. Protein concentration was calculated based on a standard curve of BSA. Data were analyzed individually by experiment in JMP as described below.

Metabolite Extraction from Cell Cultures and Deglycosylation of Advanta-Bound Metabolites

Freeze-dried cell culture samples of 10.0$\pm$0.2 mg each were extracted twice for 5 min under sonication at 70^{o}C in 680 µl methanol:water:chloroform (63:40:33). Each sample was spiked with 0.164 µM *o*-anisic acid, 0.164 µM adonitol, and 1.296 µM resorcinol during the first extraction using a master mix, and one internal standard blank was run along with each separate batch of experimental sample extractions. After each

sonication, samples were cooled to room temperature, centrifuged to separate the phases, and the aqueous methanol phase for each sample retained. The two upper phases from each sample were combined (~490 µl x 2), then approximately 20% (~200 µl) of the total volume was taken for HPLC-MS analysis and stored at -80°C. The remainder (~790 µl) was dried down under vacuum centrifugation, resuspended in 100 µl 40% methanol (aq) under brief sonication, and added to pre-washed Advanta resin (Applied Separations, Inc.). Any excess residue was resuspended in 500 ul of 4% methanol (aq) under vortexing and centrifuged prior to pooling with the first fraction, yielding a 10% methanol (aq) sample that was incubated for 20 min at room temperature on an orbital shaker. Samples were allowed to settle on ice for 20 min, then the supernatant (unbound fraction or "UB", representing relatively nonpolar metabolites not already removed by chloroform extraction) was transferred for storage at -80°C until GC-MS analysis. The resin was washed twice with 10% methanol (aq) to remove the residual unbound metabolites, then incubated 15 min on the orbital shaker with 200 µl acetonitrile to elute the Advanta-bound ("B") metabolite fraction. After settling on ice for 20 min, the supernatant was transferred to a new tube and held on ice while the elution step was repeated on the resin. The supernatant from the second elution was pooled with the first, then UB and B samples were stored at -80°C until GC-MS analysis.

Enzymatic treatment just prior to GC-MS analyses were also used to generate deglycosylated metabolites in B samples ("B-DG"). B sample aliquots of 60 µl were dried down under vacuum centrifugation and resuspended in 10 µl methanol, which was treated with 1 U β-glucosidase from almond (Sigma) and 20 mM sodium acetate at pH 5.0, 40°C for 2 h. Water-saturated ethyl acetate (400 µl) and 200 µl of 0.1 M

hydrochloric acid were added to the digest afterwards and samples vortexed for 60 s followed by centrifugation to separate phases. The upper, ethyl acetate phase was retained and the lower phase extracted again in an equal volume of ethyl acetate; the two upper phases were dried down under vacuum centrifugation, immediately prior to methoxymation and derivitization as described below. UB and B samples were also preprocessed immediately before methoxymation and derivitization by directly drying down 60 µl aliquots under vacuum centrifugation.

Metabolite Profiling via GC-MS and Data Analysis

Dried samples were sequentially subjected to methoxymation and derivitization prior to analysis in order to enhance volatilization and GC-MS detection of compounds containing carbonyl moieties. Samples were resuspended in 10 µl of 40 mg/ml methoxyamine in pyridine, 5 µl n-alkanes in pyridine (yielding a final concentration of 10 ppm) to act as retention index standards, and 45 µl MSTFA. Methoxymation of carbonyl groups was achieved by incubation of resuspended samples at 30^{o}C for 90 min, followed a derivitization step at 60^{o}C for 90 min.

For gas chromatographic separation, methoxymated, derivitized samples were injected in 1 µl volumes (injection speed 50 µl/s) in splitless mode onto a 30 m x 0.25 mm x 0.25 µm HP-5 column of 5% phenyl methyl siloxan in an Agilent 7890A GC System (Agilent Technologies, Inc.) with inlet temperature of 250^{o}C and He carrier gas at 13.8 psi with flow rate of 1.03 mL/min. After a 500 millisecond post-injection delay, the oven ramped from 80^{o}C to 120^{o}C at a rate of 6^{o}C/min, was held at 120^{o}C for 3 min, then ramped to 200^{o}C at a rate of 10^{o}C per minute. After a 2 min hold at 200^{o}C, the

temperature ramped to 260°C at a rate of 10°C/min before another 2 min hold, followed by a final ramping to 310°C at 6°C/min. The final temperature was held for 13 min for a total 50 min run time.

Quadrupole mass spectrometry detection of eluted compounds was achieved using an Agilent 5975C inert XL MSD with triple axis detector (Agilent Technologies, Inc.). Solvent delay of 7.20 min, ion source temperature of 230°C, and quadrupole temperature of 150°C were employed. Data were collected in scanning mode over an m/z range from 50.0 to 500.0 and signal detection threshold of 150 for the first 20 minutes, 50.0 to 600.0 and signal detection threshold of 150 for the next 16 minutes, and 50.0 to 750.0 and signal detection threshold of 300 over the remaining run time.

Raw chromatogram data were exported to AnalyzerPro (SpectralWorks Ltd.), which used the NIST 2008 and Agilent-Fieher RL standards libraries, as well as an in-house library developed from authentic standards, for peak identification. A score based on forward and reverse match accuracy of the mass spectral data was generated for each peak, which was used along with consistency of naming for a peak across all samples and cross-sample mass spectral matching as criteria for inclusion of the metabolite in our analysis. Data were uploaded to a custom metabolic profiling database developed in the Tsai lab known as MetaLab (http://128.192.158.63/x/MetaLab/) for completing this quality control work and facilitating rapid summarizing of quality peaks. Identification of internal standards for quantification, retention index (RI) calculation, and peak alignment are automated processes in MetaLab and were followed by visual examination to ensure accuracy. For each sample type (UB, B, B-DG), files from a single experiment were grouped by treatment and peaks aligned by MetaLab according to RI and mass

fragmentation pattern similarity coefficients. This overall summary data was exported in its entirety for manual curation. Automated calls for fed compounds, hydroxycinnamates, flavonoids, other phenolic-related compounds, amino acids, organic acids, and sugars were quality-checked by referencing the name, individual replicate RT's, and individual replicate normalized peak areas in the summary data against all additional data available in the MetaLab database for that item and all similar peaks, including RT, RI, AnalyzerPro identification by MS, and percent similarity across all replicates within and across treatment groups in the same experiment. This process led to manual correction of any peak grouping errors generated by the automatic MetaLab pipeline. Normalized peak areas for quality-checked peaks were transferred to a curated list, with individual peak data adjusted as needed to correct poor calls or improper alignments, and highlighted as needed to note any borderline or inconsistent calls.

Statistical Analysis of Metabolic Profiling Data

The effect of elicitor and inhibitor treatments within each experiment on broad categories of metabolites as a whole (i.e., phenylpropanoid core metabolites, flavonoids, amino acids, organic acids, and sugars) was analyzed using two-way MANOVAs that incorporated data for all curated metabolite data within the category. For treatment factors (Elicitor overall, Inhibitor overall, or Elicitor by Inhibitor interaction) found to be significant at the α=0.05 level in the MANOVA test, individual ANOVA tests were run on each individual metabolite within the category to determine whether the significant effects were due mainly to specific metabolites or to general trends across multiple metabolites within the group. Individual metabolite ANOVAs were one-way analyses if

only the Inhibitor overall or Elicitor overall factors were statistically significant in the corresponding MANOVA, but two-way analyses were carried out if the interaction factor or multiple factors were significant in MANOVA.

Two-way analyses are a more appropriate tool than one-way analyses for use in 2x2 factorial experiments such as those carried out in this chapter, because any combinatorial treatments (such as an elicitor combined with an inhibitor) are not truly independent from other treatments (such as elicitor alone or inhibitor alone). This semi-dependency among independent variables thus requires a slightly different statistical approach; the advantage of two-way ANOVA in such experiments is its ability to detect consistent overall trends relating to a specific factor while also allowing the ability to detect non-additive differences among different treatments. An "overall" statistical trend in this context is identified by pooling semi-dependent samples; e.g., the effect of PIP overall in Experiment 1 is analyzed by comparing pooled data for PIP and PIP+MeJA against pooled data for Control and MeJA. A "non-additive" effect in this context is identified by examining differences that remain across the four treatment types even after overall effects have been accounted for; e.g., increases in a metabolite in MeJA samples relative to Control samples may no longer be evident when comparing PIP+MeJA samples to PIP samples. Such a result may be of additional biological interest due to the inability to predict it based on data from any experiments lacking a combination treatment of both PIP and MeJA.

When a two-way ANOVA revealed a significant interaction between Elicitor and Inhibitor treatments, additional within-treatment analyses were carried out to separately determine the effect of elicitor feeding in each inhibitor treatment condition. For

example, in Experiment 1 two within-treatment tests would be run to determine (a) whether MeJA-fed samples differed significantly from unfed samples within all samples *not* fed PIP (i.e., Control versus MeJA, written as Elicitor:None), and then again to determine whether MeJA-fed samples different significantly from unfed samples within all samples that *were* fed PIP (i.e., PIP versus PIP+MeJA, written as Elicitor:PIP). This procedure is known as "interaction slicing." To avoid overanalysis of the data, sufficient degrees of freedom must be available for each test; this requires that the degrees of freedom in the overall effect being analyzed (i.e., Elicitor overall) be used in the interaction slicing. For this reason, the overall effect that is again analyzed within multiple interaction slices should be statistically considered only within the slicing context and not as an overall effect. By choosing to slice interactions to determine elicitor effects, Inhibitor overall effects may still be considered on the broader level.

Finally, for significant Inhibitor overall effects in Experiment 2, additional *post hoc* testing was carried out to identify which inhibitor treatments (none, AOPP, or MDCA) were significantly different from each other. This was accomplished using Hsu's "Multiple Comparisons with the Best" (MCB) Test, which determines whether each group is significantly greater than an unknown minimum or significantly less than an unknown maximum (Hsu 1981). All statistical analyses on metabolite data were carried out in JMP v8.0 (SAS Institute, Inc).

RNA Extraction and Microarray Transcriptome Analysis

RNA was extracted from ground cell culture samples using the CTAB method (Tsai et al. 2003), generating a total of three biological replicates from all four treatment

groups in Experiment 1 and four biological replicates in the Control, AOPP, MeJA, and AOPP+MeJA treatment groups in Experiment 2. Samples were quantified via Nanodrop spectrophotometry and quality-checked by running a 1% TAE agarose gel containing 0.003% ethidium bromide stain under electrophoresis at 50V, then 10 µg each of two biological replicates per feeding combination per experiment were subjected to treatment with 1 U TURBO DNase (Ambion) according to manufacturer's instructions to remove any remaining DNA. Treated samples were ethanol-precipitated and resuspended in water to remove any traces of inactivation reagent, then quantified and quality-checked using Nanodrop spectrophotometry, gel imaging, or RNA Nano Chip analysis on an Agilent Bioanalyzer 2100. Aliquots of approximately 10 µg were shipped on dry ice to the Center for the Analysis of Genome Evolution and Function at the University of Toronto for cRNA synthesis and hybridization to Affymetrix Poplar Genome Arrays.

Expression data were uploaded to GeneSpring XI software (Agilent Technologies) using the MAS5 algorithm followed by per-gene normalization. From a starting list of 61,251 probes per array, each experiment was filtered to exclude non-*Populus* probes and any probes that failed to show signal >50 in both replicates for at least one treatment, and the intersection set of the two filtered lists, containing 30,999 probes, was designated as the "QC List." Hierarchical clustering analysis compared QC List data for all Control and all MeJA samples across both experiments to determine the appropriate downstream analytical approach; since samples clustered more strongly by experiment than by treatment, data for each experiment was analyzed separately. Valid *post hoc* comparisons can still be made between the two experiments, however, with the recognition that they may overall be biologically different in unknown ways. A Bayesian

equivalent for two-way ANOVA was carried out for the QC List probe data within each experiment, followed by identification of significant probes using the SLIM algorithm (Wang et al. 2011). The union list of all probes found to be statistically significant according to SLIM in at least one factor within either experiment consisted of 8,298 probes and was named the "SLIM List."

Overrepresented biological process and molecular function GO categories among statistically significant probes were identified for each factor (Inhibitor overall, Elicitor overall, or Elicitor by Inhibitor interaction) within each experiment using GO analysis. Nonredundant lists of gene models corresponding to probes passing the statistical test for a given factor were submitted to the AgriGO website (Du et al. 2010) for reference against the *Populus* v2.2 genome using the Hypergeometric test and Hochberg FDR adjustment for multiple hypothesis test correction, with α=0.05 and a minimum of five mapping genes per category required for significance. Lists of statistically significant categories were further reduced using REViGO (Supek et al. 2011), comparing the lists against the Uniprot database using SimRel semantic similarity and then removing all categories with dispensability scores greater than 0.5. Data on fold overrepresentation were included in the REViGO analysis as an additional criterion for judging dispensability. Top-level GO categories within each domain (such as GO:0008152 - metabolic process within the biological process domain, and GO:0005215 - transporter activity within the molecular function domain) were removed from the final lists after REViGO reduction.

At the level of individual genes, expression ratios for Inhibitor/Control, Elicitor/Control, Inhibitor+Elicitor/Control, and Inhibitor+Elicitor/Elicitor were

calculated for all SLIM List probes and filtering performed to identify probes hybridizing to known phenylpropanoid and flavonoid biosynthetic pathway genes, as well as genes putatively involved with the citric acid cycle, amino acid biosynthesis, nitrate metabolism, or ammonium metabolism.

QPCR Analysis of Expression Patterns in Phenylpropanoid Core Pathway Gene Isoforms and Flavonoid Biosynthetic Genes

Fresh aliquots of RNA from three biological replicates previously extracted were subjected to DNase treatment and concentrated via ethanol-precipitation as described above. Treated samples were quantified using a Nanodrop spectrophotometer and quality checked on a 1% agarose gel in TAE, then 2.25 µg (Experiment 1 samples) or 1.75 µg (Experiment 2 samples) were used for cDNA synthesis. For each sample, total RNA was combined with 0.5 µl 5X First Strand Buffer (Invitrogen), 0.5 µl 0.1 M DTT, 20 U RNase inhibitor (Ambion), 100 µM anchored oligo dT primers, and RNase-free water to yield a volume of 16 µl. RNA secondary structure was denatured by incubation at 65^{o}C for 5 min, followed by rapid chilling on ice and a brief centrifugation. An additional 4.5 µl of 5X First Strand Buffer, 2.0 µl 0.1 M DTT, 20 U RNase inhibitor, 1.0 µl of 10 mM dNTPs, and 200 U SuperScript II RT were added to achieve a final volume of 25 µl. Reverse transcription was carried out by a 10 min incubation at room temperature followed by incubation at 42^{o}C for 2 h. Samples were stored at -80^{o}C prior to use.

QPCR analysis was run on 1.5 ng RNA equivalents of cDNA in 10 ul reactions containing 100 nM each of forward and reverse primer, 0.003% ROX internal standard, and 1X ABsolute^{TM} QPCR SYBR$^{®}$ green master mix (Thermo Scientific, Inc.). Thermal

cycling was performed on the Mx3005P™ Real-Time PCR System (Stratagene) using an initial 15 min at 95°C to activate the Thermo-Start® DNA Polymerase present in the master mix, followed by forty cycles of 15 s at 95°C, 60 s at 57°C, and 60 s at 72°C, with fluorescence quantified at the end of the annealing step in each cycle. At the end of the run a melting curve analysis was also completed from 55°C to 95°C. For each gene, all biological replicates from a single experiment, each with two technical replicates, were analyzed in a single plate; samples from different experiments were run on separate plates.

Data for individual replicates were individually subjected to amplification efficiency analysis using LinRegPCR (Ramakers et al. 2003), and replicates with R^2<0.990 were removed from all downstream analyses. Efficiencies for remaining samples amplifying the same gene fragment were averaged to generate the final amplification efficiency for that gene. Gene expression was quantified as $N^{-\Delta Ct}$ relative to the geometric mean of the housekeeping genes ubiquitin-conjugating enzyme E2 (*UBCc*), cyclophilin (*Cyp*), and actin-related complex protein (*ARP*). Primer sequences are listed in Table 3.1, were described in **CHAPTER 2** (housekeeping genes), or detailed in previous work (for phenylpropanoid and flavonoid related genes; Payyavula et al. 2011; Tsai et al. 2006b).

Results

Detection of Phenylpropanoid Metabolic Inhibitors and Methyl Jasmonate in Extracts of Fed *Populus* Cell Cultures

Populus tremuloides cells treated with one of three phenylpropanoid inhibitors, PIP (Experiment 1), AOPP, or MDCA (Experiment 2), with or without the elicitor MeJA (both experiments), were subjected to metabolite profiling by GC-MS. Unmetabolized PIP and MDCA were detected in the Advanta resin-bound fraction (or B fraction; enriched with relatively polar compounds with phenyl rings) from the aqueous portion of methanol:water:chloroform extracts from cells harvested 48 hours after initial feeding (Figure 3.1, panels A & B). Low levels of a compound with a PIP-like signal were also detected in unfed cells, but peak matching confidence against the NIST library was lower than that for fed cells. MS similarity coefficients of aligned peaks between paired unfed and fed samples were about one fourth lower than those for paired peaks across treated samples, and peak heights in PIP-fed cells were about tenfold higher than the putative corresponding peak in unfed samples ($p<0.0001$). Overall, this suggests that the putative signals from unfed cells are unlikely to be PIP. PIP was also detected in the unbound fraction (UB) of samples treated with the compound, but partitioning was heavily biased towards the B fraction. PIP levels did not differ significantly based on MeJA treatment status. MDCA was detectable only in B fraction of fed samples, although levels were significantly lower in samples that had also been fed MeJA ($p=0.0036$). This suggests a possible effect of MeJA co-feeding on metabolism of MDCA. Unmetabolized MeJA and AOPP were not detected in any samples.

B fraction extracts were also subjected to β-glucosidase digestion to generate the B-DG fraction. This was utilized for GC-MS detection of combined free phenolic acids (as seen in B fraction) and those released from their corresponding β-D-glucosides (only seen after digestion). In addition to the known occurrence of a portion of plant phenolic acids in β-glucoside forms, including salicinoids in *Populus*, glycosylation can also be considered an indicator of cellular uptake and metabolism of exogenous compounds (Vaistij et al. 2009). PIP and MDCA levels in cells treated with each inhibitor were about an order of magnitude greater in the B-DG fraction than in the B fraction (Figure 3.1), suggesting cellular uptake and glycosylation of the two inhibitors. While PIP levels did not differ significantly depending on MeJA status, MDCA levels were reduced by over one-third in MeJA-treated cells ($p=0.0377$; Figure 3.1D). Although AOPP was not identified by NIST mass spectral library search, a putative benzenepropionic acid was identified in the B-DG fraction only for AOPP-treated samples. Levels were much lower than those for MDCA in the same experiment despite similar treatment doses, and MeJA co-feeding further reduced levels by over 70% ($p=0.0033$; Figure 3.1D). Running an authentic sample of AOPP standard on GC-MS under similar conditions would help determine whether this peak represents a possible metabolic or thermal decomposition product of AOPP.

Putative MeJA peaks were also detected in the B-DG fraction extracts from elicited cells (Figure 3.1, panels C & D). Peaks were specific to MeJA-fed cells, although one fed sample did not exhibit the peak. The best NIST library match for this peak in over 1/3 of the samples was methyl jasmonate, with within-group similarity coefficients for mass spectral fragmentation patterns of the chromatogram peak among

paired MeJA-fed samples exceeding 84% for all but one sample in Experiment 1 and 74% for all but one sample within each treatment group in Experiment 2. Despite the high incidence of methyl jasmonate as the best NIST library match, library matching confidence was <60%. Due to this relatively low call confidence, it seems likely that the chromatogram peak originated from cellular uptake and subsequent glycosylation of MeJA *in vivo*. Given the preparation method for the B-DG fraction and the volatility of MeJA, it seems likely that the chromatogram peak arose from a less volatile jasmonate generated from at least one additional metabolic transformation, rather than from MeJA itself. Overall, the levels of this putative jasmonate were significantly lower in Experiment 2 than in Experiment 1 ($p=0.0041$) and responded significantly to inhibitor treatment ($p=0.0362$) according to two-way ANOVA. Dissecting the latter effect within each experiment revealed that this effect was traceable only to AOPP-fed samples, which had about 60% higher putative jasmonate levels than samples not fed any inhibitor according to a Hsu's MCB test ($p=0.0342$; Figure 3.1). Because of the differences in jasmonate levels overall and other discrepancies observed in metabolic and transcriptomic results (described later) between experiments, all subsequent statistical analyses were conducted separately by experiment.

Effects of Inhibitor and MeJA Treatments on Carbon to Nitrogen Ratios

The effects of MeJA and inhibitor treatments on overall nitrogen and carbon metabolism were assessed by comparing C/N ratios in a subset of samples. Within Experiment 1, both PIP ($p=0.0091$) and MeJA ($p=0.0014$) were associated with significant reductions in C/N ratio (Figure 3.2A), but the two factors did not significantly

interact (p=0.3514). Within Experiment 2, MDCA and AOPP were associated with significantly lower C/N ratios overall than for cells not treated with inhibitor (p=0.0007, Figure 3.2B). Slicing the significant interaction between inhibitor and elicitor treatment in Experiment 2 (p=0.0056) revealed that MeJA feeding was associated with significantly higher C/N ratios in samples treated with AOPP (p=0.0083) or no inhibitor (p=0.0007), but not in samples treated with MDCA (p=0.8635). Changes in C/N ratio could be traced primarily to significant changes in %N for Experiment 1, but both %C and %N were changed in Experiment 2 (data not shown). Further interpretation of these results can benefit from separate quantification of major carbon and nitrogen pools within the samples.

Condensed Tannins Show Increased Accumulation Under MeJA Feeding and Reductions Under Metabolic Inhibitor Feeding

One major carbon pool in the cell cultures consisted of CTs, which also represent a major phenylpropanoid end product. In preliminary trials carried out by Dr. Payyavula to optimize MeJA and inhibitor treatments, CTs were elevated by nearly 50% (over 4.5 percentage points) in cells treated with the same level of MeJA used in Experiments 1 and 2 within 24 h after feeding began (p<0.0001; Figure 3.3A), consistent with previously-reported CT accumulation induced by jasmonates (Arnold et al. 2004; Arnold and Schultz 2002). CTs decreased significantly in PIP-fed cells (~5% or 0.6 percentage points; p= 0.0103) and AOPP-fed cells (~30% or 2.8 percentage points; p=0.0009; Figure 3.3B) 24 h after initiation of the trials. A trend towards decreased CT levels under MDCA treatment was not statistically supported at 24 h (~18% or 1.6 percentage points;

p=0.1073; Figure 3.3C) or at 48 h (p=0.0759, data not shown), although it did become significant at 96 h (p=0.0437, data not shown). In a 2x2 factorial trial that included both PIP and MeJA in the same experiment, no significant interaction effect on CTs was found (p=0.5963).

Changes in Lignin Content and Composition Under Elicitor and Inhibitor Feeding are Minor

A second major carbon pool representing phenylpropanoid end products is lignin. Lignin content and syringyl-to-guaiacyl monolignol (S/G) ratio were assessed using pyrolysis molecular beam mass spectrometry (pyMBMS) on samples of freeze-dried, ethanol-washed cells. Measurements using this high-throughput approach are useful for providing relative (but not necessarily absolute) quantities of lignin and S/G ratio to identify treatment effects (Sykes et al. 2010). For Experiment 1, lignin content showed small (~2% or a quarter of a percentage point) but significant (p=0.0089) increases in MeJA-fed samples regardless of PIP status, with no significant effect of PIP (p=0.4754) or MeJAxPIP interaction (p=0.4518; Figure 3.4A). MeJA feeding also led to significantly lower S/G ratios in Experiment 1 overall (p=0.0227; Figure 3.4C), primarily due to increases in G-lignin (data not shown). PIP and PIPxMeJA did not significantly influence S/G ratio. In Experiment 2 a non-significant trend (p=0.1626) toward decreased lignin content was observed in MeJA-fed samples. Meanwhile, inhibitor feeding led to small but significant changes in lignin content, with AOPP-fed cells having about 2% (or two-tenths of a percentage point) lower lignin overall than cells not fed with inhibitor (p=0.0178; Figure 3.4B). The S/G ratio was not influenced by MeJA or

inhibitor feeding in Experiment 2 (Figure 3.4D). Overall, the effects of metabolic perturbation on lignin content and S/G ratio were smaller than those observed for CTs, suggesting the two phenylpropanoid branch pathways had distinct responses to the treatments.

Soluble Protein is Reduced by MeJA and Increased by PAL Inhibitor

Because protein represents a major nitrogen pool in plants, soluble protein levels were measured to identify effects of elicitor and inhibitor feeding. MeJA led to small but statistically significant decreases in protein levels overall in both experiments (Figure 3.5; $p=0.0004$ in Experiment 1; $p<0.0001$ in Experiment 2). In Experiment 1 the effect of PIP feeding was mediated by MeJA status (Figure 3.5A; $p=0.0088$ for MeJAxPIP), leading to decreased protein levels in PIP-fed cells only if MeJA had also been fed. In constrast, AOPP-fed cells exhibited a small but statistically significant increased in protein levels regardless of MeJA status (Figure 3.5B; $p=0.0027$). In summary, both MeJA and metabolic inhibition in the core phenylpropanoid pathway influence soluble protein levels.

Metabolic Inhibitor Effects on Phenylpropanoid Core Pathway Metabolites Suggest Efficacy of PIP and AOPP

We next examined core phenylpropanoid metabolites. Five such compounds (Figure 3.6, top row) were consistently detected in the B fraction of *Populus* cell extracts in both feeding experiments. Hydroxycinnamoyl-CoA conjugates (Figure 3.6, bottom row) were not detected by GC-MS. Based on current understanding of the fed inhibitors

and their enzyme targets, the suite of five metabolites should be sufficient for assessment of inhibitor efficacy in this experimental system. While the amino acid phenylalanine was found more abundantly in the UB fraction, all four hydroxycinnamate free acids were found primarily in the B fraction. However, a portion of hydroxycinnamates can occur *in vivo* not as the free-acid form, but as glucose conjugates (e.g., Lu and Yeap Foo 2000; Nagels et al. 1981). The combined total of glucose conjugates and free-acid form of each individual hydroxycinnamate was determined following β-glucosidase digestion of the B fraction (i.e., the B-DG, or bound deglycosylated, fraction), based on the resulting free acids. As previously noted, this digestion treatment is specific for deglycosylation of β-glucosides, and other glucose conjugates such as hydroxycinnamate glucose esters are unlikely to be released to their free acid forms using this approach.

When each "total" hydroxycinnamate pool (i.e., free acid and glycosylated components of a single hydroxycinnamate species from B-DG) and the phenylalanine pool (from UB) in Experiment 1 were compiled in a single MANOVA, the core phenylpropanoid pathway collectively showed significant responses to MeJA and PIP as well as a significant interaction effect (Table 3.2). Significant responses to MeJA and PIP overall were also observed for compiled hydroxycinnamate free acid data (B), but no interaction was observed in this case (Table 3.2). These results are broadly consistent with an influence of inhibitor and elicitor treatments on the phenylpropanoid core pathway as a whole, as well as an interactive influence of inhibitor and elicitor on the core pathway that cannot be inferred from either treatment alone.

To identify treatment effects on specific metabolites within the pathway, the core pathway compounds were then individually analyzed via two-way ANOVA only for

treatment effects that were significant in MANOVA. Cells fed with PIP had overall more than 100-fold higher levels of total cinnamic acid (free acid plus glycosylated forms) but lower levels of both total *p*-coumaric acid and total caffeic acid (Figure 3.7A; Table 3.3), consistent with inhibition of C4H. PIP was also associated with lower levels of phenylalanine when fed without MeJA. Overall, MeJA feeding led to significantly lower levels of phenylalanine, total caffeic acid, and total ferulic acid, but did not affect total *p*-coumaric acid. MeJA feeding had no significant effect on total cinnamic acid in the absence of PIP (i.e., MeJA vs. Control), but led to increases when PIP was present (i.e., MeJA+PIP vs. PIP, Figure 3.7A; Table 3.3).

When only considering free hydroxycinnamates (i.e., non-glycosylated free acids from the B fraction), similar effects of PIP were observed, with more than 40-fold higher levels of free cinnamic acid and significantly reduced levels of free *p*-coumaric acid and ferulic acid (Figure 3.7B; Table 3.4). A trend towards reduced free caffeic acid was also observed in PIP-fed samples. MeJA feeding led to lower levels of free *p*-coumaric acid, caffeic acid, and ferulic acid, but did not have a significant effect on free cinnamic acid (Figure 3.7B; Table 3.4). While MANOVA showed no statistical support for significant influences of MeJA or PIP on representation of the free acids as a percentage of the total hydroxycinnamates (Table 3.2), a weak trend toward reduced representation in MeJA-treated samples seemed visually evident (Figure 3.7C). Cinnamic acid also seemed to show a trend towards reduced representation of the free-acid form in PIP-fed samples (Figure 3.7C), but statistical support was lacking (data not shown). The lack of statistical support could be attributed in part to the additional variation arising from sample-wise

division across data from two different fractions (B and B-DG) to derive the percentage data for the proportion of free acids.

MANOVA testing for Experiment 2 demonstrated significant collective effects of inhibitors and MeJA for both the total hydroxycinnamates (B-DG) with phenylalanine (UB) and for free hydroxycinnamates (B; Table 3.2). A significant interaction between inhibitor and MeJA was also observed for both groups of data (Table 3.2). This suggests an influence of the separate treatments on the core phenylpropanoid pathway overall, as well as effects on the pathway due to combinations of MeJA and inhibitors that cannot be predicted through the use of any one treatment alone. Since all two-way MANOVA effects were statistically significant, identification of treatment effects on individual metabolite types was carried out using two-way ANOVA.

In Experiment 2, feeding of AOPP resulted in a nearly 30-fold increase in phenylalanine concomitant with decreased total *p*-coumaric acid, total caffeic acid, and total ferulic acid, consistent with PAL inhibition (Figure 3.8A; Table 3.3). Total cinnamic acid levels were relatively low in unfed cells and were not significantly affected by AOPP feeding. Cells treated with the 4CL inhibitor MDCA exhibited an over 50-fold increase in total cinnamic acid, but levels of phenylalanine and each of the other total hydroxycinnamates were not altered (Figure 3.8A; Table 3.3). Responses to MeJA in Experiment 2 were broadly consistent with those for Experiment 1, with overall reductions in phenylalanine and total ferulic acid. Total caffeic acid levels were also significantly reduced under MeJA feeding, except in AOPP-treated samples where the trend was not significant (Figure 3.8A; Table 3.3). MeJA had no effect on total cinnamic acid levels but caused significant increases overall in total *p*-coumaric acid levels (Figure

3.8A; Table 3.3). The shift in total *p*-coumaric acid levels is in contrast to the pattern observed for the same pool in Experiment 1, where the trend was in the opposite direction.

The same trends observed for total hydroxycinnamates were observed for their free-acid forms (aglycones, B): free *p*-coumaric acid, caffeic acid, and ferulic acid were significantly reduced in AOPP-treated cells, while free cinnamic acid was increased in MDCA-treated cells (Figure 3.8B; Table 3.4). MeJA feeding contributed to overall reductions in free caffeic acid but increases in free *p*-coumaric acid, as seen for total caffeic and *p*-coumaric acids (Figure 3.8B; Table 3.4). In addition, free cinnamic acid levels decreased under MeJA treament in MDCA-fed samples, although the free-acid form of cinnamate was not detectable in other samples. Free ferulic acid levels were significantly decreased under MeJA feeding in samples not treated with an inhibitor, and showed a trend towards increasing in samples treated with MDCA (Figure 3.8B; Table 3.4). Across both experiments, then, phenylpropanoid metabolic perturbation affects levels of phenylalanine and both free-acid and β-glycosylated forms of hydroxycinnamates.

Overall, free acids constituted much smaller percentages of total hydroxycinnamates in Experiment 2 (up to ~20%) than in Experiment 1 (up to ~60%) across all treatments. Experiment 2 also showed a significant collective effect of inhibitor on the proportion of free acids in the total pools (Table 3.2). Not surprisingly, cinnamic acid had a significantly higher proportion of free acid in MDCA-fed samples than in the others, where no free cinnamic acid was detected (Figure 3.8C; Table 3.5). Ferulic acid exhibited a significantly lower proportion of free acid in AOPP-fed samples

relative to other samples (Figure 3.8C; Table 3.5), although the magnitude of the change was modest. A similar trend approaching statistical significance was observed for *p*-coumaric acid (Figure 3.8C; Table 3.5). In summary, perturbation of the phenylpropanoid pathway appears to influence the balance between free and β-glycosylated forms of hydroxycinnamates.

Effects of Inhibitor and MeJA Feeding on Benzenoid and Phenolic Compounds

I next examined levels of several other metabolites known or likely to be related to the phenylpropanoid pathway. Catechol glucoside, shikimic acid, quinic acid, arbutin, and phenyl-β-D-glucopyranoside were all found in the UB fraction of cell extracts, while 3,4-dihydroxybenzoate was found in the B fraction. Although quinic acid levels did not show significant changes in response to MeJA or inhibitor feeding in either experiment, the five other compounds examined did respond to at least one factor (Table 3.6).

In Experiment 1, arbutin was reduced by more than 50% in cells treated with PIP relative to those not fed the inhibitor, but none of the other metabolites responded significantly to PIP treatment overall (Table 3.6; Figure 3.9). MeJA treatment also led to arbutin levels ~45% of those in untreated cells, but in PIP-treated cells a weaker trend towards decreased arbutin was not statistically supported due to already low metabolite levels. Catechol glucoside also decreased by ~48% and the most abundant of the six metabolites, 3,4-dihydroxybenzoate, increased by ~36% under MeJA treatment overall (Table 3.6; Figure 3.9).

For Experiment 2, AOPP-treated cells exhibited a nearly 7-fold increase in arbutin in conjunction with statistically significant decreases of at least 33% in phenyl-β-D-

glucopyranoside, shikimic acid, and 3,4-dihydroxybenzoate (Table 3.6; Figure 3.9). None of the metabolites showed statistically significant changes overall in response to MDCA treatment. As in Experiment 1, 3,4-dihydroxybenzoate increased significantly in response to MeJA feeding, this time by ~48% (Table 3.6; Figure 3.9). Catechol glucoside levels were not significantly influenced by MeJA feeding in Experiment 2, but shikimic acid increased by ~53% under MeJA feeding (Table 3.6; Figure 3.9).

Methyl Jasmonate and Metabolic Inhibitors had Differential Effects on Flavonoid Levels

Several major flavonoid or putative flavonoid metabolites were identified by GC-MS in both the B (representing aglycones) and the B-DG (aglycones plus flavonoid β-glucosides, or total flavonoids) fractions. Tentative identifications include catechin, taxifolin, eriodictyol and kaempferol, each with multiple peaks. This likely indicates that closely related flavonoid structures and glucosides were detected. In Experiment 1, MANOVA showed significant collective effects of MeJA and a significant interaction between PIP and MeJA on total flavonoids, but no collective treatment effects on flavonoid aglycones (Table 3.7). Peak-by-peak statistical analysis was only carried out on each total flavonoid pool, due to the lack of any collective treatment effect on the aglycones. Four of five flavonoid peaks in the B-DG fraction significantly decreased under MeJA feeding in Experiment 1, while flavonoid aglycones (B fraction) showed weak trends towards increases in two putative catechin peaks and a putative flavanonol (Table 3.8). The collective interaction effect identified in MANOVA for total flavonoids could not be traced to a significant MeJAxPIP effect for any single total flavonoid pool.

In Experiment 2, overall levels of flavonoids tended to be higher than in Experiment 1. Total flavonoids and flavonoid aglycone pools each showed significant collective effects of MeJA and inhibitor treatments in Experiment 2, with no significant interaction between treatments (Table 3.7). Analysis of individual peaks from the B-DG fraction showed that a putative catechin (RI 2878), one of the most abundant flavonoids, was significantly decreased under MeJA treatment (Table 3.9), unlike the pattern seen in Experiment 1. Meanwhile, six flavonoid aglycones, including four putative catechins, were all significantly increased under elicitation (Table 3.9). Most of these appeared to have much lower abundance than the abundant putative catechin from B-DG in all treatments, but a putative catechin aglycone (RI 2879) was at least 4.5-fold greater in all treatments. Attempted peak matching between B and B-DG fractions suggests that MeJA feeding may lead to increases in the proportion of aglycone flavonoids present (data not shown). In both experiments, the differential shifts in metabolite levels between aglycone and total flavonoid pools and, in most cases, their magnitudinal differences suggest that MeJA feeding had a strong effect on several flavonoid glucosides that were not directly identified from the NIST library search.

AOPP and MDCA feeding in Experiment 2 had varying effects on both flavonoid pools. AOPP feeding caused a significant reduction in all total flavonoids (Table 3.9). Within the aglycone pool, significant reductions in three putative taxifolin peaks in AOPP-fed cells, significant increases in the kampferol peak in MDCA-fed cells, and multiple patterns of response to AOPP and MDCA feeding for putative catechin peaks (Table 3.9). In particular, three of four catechin-like peaks exhibited significant increases in MDCA-fed cells relative to AOPP-fed and/or unfed cells; the fourth showed a

significant increase under AOPP-fed conditions despite a significant decrease in a different catechin-like peak under the same conditions. Overall, these results are consistent with shifts in flavonoid partitioning due to pathway stimulation by MeJA and to phenylpropanoid pathway interference by the various inhibitors.

Feeding of a PAL Inhibitor Influences Amino Acid Pool Composition

Given the strong influence of AOPP on phenylalanine levels and the possibility of feedback and/or feed-forward connections between phenylpropanoid and amino acid metabolism, responses to elicitor and inhibitor treatments were characterized for amino acids and closely related nitrogenous substances that were readily and consistently detected (Tables 3.10 and 3.11). These metabolites were found primarily in the UB fraction of the cell extracts; they were either absent or present at consistently lower levels in the B fraction in all cases. We identified nine amino acids in both experiments, excluding phenylalanine which was presented earlier. Three additional amino acid or amino acid-like metabolites were found only in Experiment 2. MANOVA testing revealed that PIP feeding did not have a significant effect on amino acids overall ($p=0.4352$), but MeJA treatment did ($p=0.0495$). In contrast, inhibitor treatment in Experiment 2 did have a significant effect on amino acids overall ($p=0.0029$), but MeJA did not ($p=0.0997$). Neither experiment exhibited a significant interaction between inhibitor and elicitor feeding on overall amino acid data ($p>0.1$ for both experiments).

Within Experiment 1, aspartic acid, glutamic acid, glycine, serine, and homoserine were lower under MeJA feeding overall, while threonine was higher in this case (Table 3.10). Although the patterns were statistically significant for all five

metabolites, they appeared to be moderated by PIP feeding. Valine also showed a non-significant trend towards reduction under MeJA feeding. Within Experiment 2, β-alanine, 5-hydroxynorvaline, and tyrosine exhibited very large increases in samples treated with AOPP. Isoleucene and lysine also showed smaller but still significant increases, while threonine and valine exhibited modest but significant decreases (Table 3.11). No significant responses were associated with MDCA feeding.

Methyl Jasmonate Feeding Impacts Organic Acids and Sugars More Strongly Than Do Phenylpropanoid Metabolic Inhibitors

To examine responses of primary carbon metabolism, we quantified organic acids, sugars, and sugar phosphates. Six citric acid cycle metabolites were consistently identified in UB fractions of cell extracts, with isocitrate, oxaloacetate, and CoA conjugates not identified. Nine additional organic acids, including some involved with radical scavenging, glycolysis, anaerobic metabolism, were found in UB fractions, with one additional peak identified only in some MeJA-treated cells in Experiment 2. One of the nine metabolites, 3-phenyllactic acid, was also found in substantial quantities in B and B-DG fractions of cell extracts; this metabolite was analyzed in a fraction-specific manner. The major soluble sugars fructose and glucose each appeared in two different separately-analyzed peaks, while sucrose was found as a single peak. Glucose-6-phosphase also appeared in two separate peaks; a ribose phosphate and a ribose-5-phosphate peak were found only in Experiment 1. The metabolites analyzed highlight a suite of central processes with a particular focus on the citric acid cycle and soluble

sugars, but do not constitute a comprehensive analysis of the organic acids, sugars, and sugar phosphates found in the extracts.

MANOVA for all citric acid cycle components in Experiment 1 showed that PIP and MeJA treatments each had a significant collective influence on the citric acid cycle; this was also true for a collective analysis of other organic acids (Table 3.12). Neither set of organic acids exhibited a significant interaction effect between PIP and MeJA. Similar collective analysis of sugars and sugar phosphates revealed a significant response to MeJA only (Table 3.12).

The significant collective response of citric acid cycle components to PIP treatment was traceable mainly to fumaric acid, which was significantly lower under PIP feeding and further reduced in MeJA-treated cells (Table 3.13). Succinic and α-ketoglutaric acids were also 48% and 38% lower, respectively, in MeJA-treated cells regardless of PIP treatment (Table 3.13). Citric acid was the most abundant of the citric acid cycle metabolites and increased by more than 50% in MeJA-treated cells (Table 3.13). Overall, these patterns suggest strong shifts in the citric acid cycle under elicitation, but a minor response to PIP.

PIP did not significantly affect levels of any other individual organic acid in the UB fraction, but the collective effect could be traced to slight decreases in most of the organic acids in cells not treated with MeJA. In MeJA-treated cells, PIP-related trends were reversed toward small increases for the majority of organic acids. Two metabolites, 3-phenyllactic acid and phenylpyruvate, which showed consistent trends toward increase in PIP-treated cells regardless of MeJA status, with the effect reaching statistical significance in the B-DG fraction for 3-phenyllactic acid (Table 3.13). MeJA led to

decreases in ascorbic, dehydroascorbic, malonic, and 3-phenyllactic acids, all by about 30% relative cells not treated with MeJA (Table 3.13). The monosaccharides fructose and glucose both showed a similar, statistically significant decrease under MeJA treatment, although sucrose and sugar phosphates were not affected (Table 3.13).

In Experiment 2, citric acid cycle components collectively showed a significant response to inhibitor treatment, MeJA treatment, and an interaction effect between MeJA and the inhibitors (Table 3.12). For the other organic acids, MANOVA showed a significant collective effect of inhibitor treatment and MeJA treatment, but no interaction. As for Experiment 1, sugars and sugar phosphates collectively responded only to MeJA treatment (Table 3.12).

The significant collective effect of inhibitor feeding on citric acid cycle metabolites in Experiment 2 could be traced mainly to four compounds, citric acid, *cis*-aconitic acid, α-ketoglutarate, and fumaric acid (Table 3.14). Overall, AOPP feeding led to a nearly 30% increase in α-ketoglutarate and decreases of about 65% for fumaric, 33% for *cis*-aconitic, and 22% for citric acid relative to samples fed no inhibitor, although the last of these was not statistically significant (Table 3.14). MDCA treatment also led to significant decreases in fumaric (by ~73%) and cis-aconitic acids (by 1/3) relative to samples fed no inhibitor (Table 3.14). While citric acid was not significantly different between MDCA-fed cells and cells fed no inhibitor, it was significantly elevated by 48% relative to cells fed AOPP. This suggests the trends for both inhibitors likely reflect true metabolic effects associated with the treatments rather than sampling error. As in Experiment 1, citric acid exhibited a strong response to MeJA treatment, increasing about 2.3-fold relative to cells fed no elicitor, while succinic acid exhibited a statistically

significant decrease of about 24%. Fumaric acid exhibited an interaction effect between inhibitor and MeJA treatments, with a 2.2-fold increase associated with MeJA feeding in cells fed no inhibitor ($p=0.0009$) versus decreases of 36 and 40% in MeJA-fed cells co-treated with AOPP ($p=0.0653$) and MDCA ($p=0.0034$), respectively (Table 3.14). Thus, while citric acid cycle components showed a somewhat less extensive MeJA response in Experiment 2 than in Experiment 1, inhibitor-related responses were much more widespread in the former.

The other organic acids also showed broader significant responses to inhibitor treatment than were observed in Experiment 1. Malonic acid and 3-phenyllactic acid (UB) both increased significantly in AOPP-treated cells, with 3.5-fold and ~47% increases, respectively (Table 3.14). In the B and B-DG fractions, 3-phenyllactic acid was also increased by more than 13-fold in AOPP-fed cells compared to cells not fed inhibitors overall (Table 3.14). Phenylpyruvate decreased by over 26% in AOPP-treated cells, but increased by nearly one-third in MDCA-treated cells. In MeJA-treated cells, 3-phenyllactic acid decreased by 25% or more in all fractions, and a putative dehydroascorbic acid peak absent from non-elicited cells was found (Table 3.14). The latter was no longer detected when AOPP was present in elicited cells. The significant collective response of sugars to elicitor feeding in Experiment 2 could be traced only to a significant effect of MeJA feeding on the smaller of the two glucose peaks (Table 3.14). No similar trend was seen in the larger peak, suggesting the overall effect was likely small.

Transcriptome Responses Were More Extensive for MeJA Than Inhibitor Feeding

Microarray analyses were conducted to explore the effects of inhibtor and elicitor feeding at the transcriptome level. Because of the relatively subtle effects of MDCA treatment on the core phenylpropanoid pathway based on metabolic data, only PIP (Experiment 1) and AOPP (Experiment 2) samples for 2x2 experimental designs with MeJA treatments were included. The number of probes from both experiments satisfying the minimum expression threshold is illustrated in a Venn diagram (Figure 3.10A), and the intersection of the two lists consisting of 30,999 probes ("QC List") was used for statistical analysis. A preliminary hierarchical clustering analysis (HCA) of the Control and MeJA-fed samples from both experiments was carried out using the QC List in order to determine the appropriate strategy for statistical analysis of treatment effects on gene expression. Results indicated that samples clustered more strongly by experiment than by treatment (Figure 3.11), and therefore array data from the two experiments should be independently analyzed. This clustering by experiment is consistent with previously mentioned differences in metabolite data across the two experiments. Additional HCA tests for all samples within each experiment showed clustering by treatment, as expected (data not shown). Differentially expressed (DE) probes were identified by a Bayesian analogue of two-way ANOVA in conjunction with SLIM (Wang et al. 2011) for each experiment. Nearly 8300 probes (the "DE List") showed statistically significant expression changes for at least one factor (Elicitor overall, Inhibitor overall, or ExI interaction) in at least one experiment (Figure 3.10B). Informal *post hoc* comparisons for a given probe's expression data between the two experiments are not precluded by the

analytical strategy used here, and were later conducted to assess specific gene expression responses.

Analysis of Experiment 1 showed that 3130 probes responded to MeJA overall, 315 to PIP overall, and 656 exhibited a significant interaction between PIP and MeJA treatments (Figure 3.12A). The majority of significant probes (~76%) responded to MeJA only, the predominant factor. Less than 4% of significant probes exhibited an exclusively PIP-dependent response, and about 9% of significant probes exhibited only a PIPxMeJA interaction but not MeJA or PIP overall effects. Relatively small proportions of probes responded to two of these factors, with 6% for MeJA and PIPxMeJA, about 2% for MeJA and PIP, and <1% for PIP and PIPxMeJA. A small proportion (~1.6%) responded significantly to all three factors.

In Experiment 2, a total of 4970 probes responded significantly to MeJA overall, 2112 to AOPP overall, and 66 showed a significant interaction between AOPP and MeJA feeding (Figure 3.12B). Again, the majority of significant probes (~64%) responded to MeJA only. About 16% of the probes responded to AOPP only, and less than 0.2% of probes exhibited a significant interaction between MeJA and AOPP feeding but no overall effect for either one alone. A relatively large proportion of probes (18%) responded significantly to both MeJA and AOPP, while about 0.2% responded to both MeJA and AOPPxMeJA and about 0.1% to both AOPP and AOPPxMeJA. About 0.6% of probes responded significantly to all three factors in Experiment 2. In both experiments, then, MeJA feeding was associated with a greater transcriptional response, in terms of the number of DE probes, than was inhibitor feeding or the interaction between inhibitor and elictor feeding.

Gene Ontology (GO) enrichment analysis was used to identify over-represented gene functions within each up- and down-regulated DE gene list by treatment and experiment. Because the total number of overrepresented GO categories within each list was as high as 104, REViGO (Supek et al. 2011) was utilized to reduce categorical redundancy, resulting in no more than 37 categories per treatment after data reduction. In Experiment 1, a wide range of categories were overrepresented in genes significantly responding to MeJA, consistent with the overall strong transcriptome response to this treatment. Within the biological process domain, MeJA-upregulated genes exhibited strongest overrepresentation in the aromatic amino acid biosynthetic process category (>10 fold) along with two other categories related to amine and cellular nitrogen metabolism (Table 3.15). Two categories relating to protein biosynthesis, along with one each for sulfur metabolism, oxidative phosphorylation, and glycolysis, were also strongly overrepresented (all >3.9 fold). In the molecular function domain, two threonine-type peptidase activity categories were most strongly overrepresented (>8 fold) in the molecular function domain, followed by categories relating to ribosome structure and to kinase, proton transport, carbon-oxygen lyase, GTPase, N-acetyltransferase, and FMN- and amino acid-binding activities, all overrepresented by >3.2 fold (Table 3.16). Overrepresented biological process categories among genes downregulated by MeJA treatment had smaller fold changes than for genes upregulated by MeJA treatment. Those overrepresented in downregulated genes included oligosaccharide metabolism (>4 fold) and related carbohydrate biosynthesis, two secretion categories, and two localization categories (Table 3.15). Meanwhile, downregulated genes in the molecular function domain were enriched in categories of lipid binding (>4 fold), two peptidase

activities, binding of vitamins, coenzymes, or cofactors, and two ligase activities (Table 3.16).

In accordance with PIP's minor effects on global gene expression (Figure 3.12A), no GO enrichment was observed among DE genes responding to PIP. However, genes responding significantly to MeJAxPIP interactions exhibited a number of overrepresented GO categories. Most of the top ten overrepresented biological process categories among genes showing a positive MeJAxPIP interaction were linked to metabolism, with cellular amino acid biosynthesis most strongly overrepresented (>7 fold; Table 3.15). In the molecular function domain, ATPase-dependent helicase activity was most strongly overrepresented (~5.6 fold), followed by four nucleic acid/nucleotide binding categories, cation transport, acyltransferase activity, and ribosome structure (Table 3.16). GO enrichment of genes showing a negative MeJAxPIP interaction was only observed in the molecular function domain, where helicase (>4 fold), peptidase, and hydrolase activities were overrepresented (Table 3.16). Overall, these results support the idea that MeJA feeding generated a broader transcriptional response across biological processes and molecular functions than did PIP feeding alone or the interaction of the two factors.

Within Experiment 2, genes significantly upregulated by MeJA showed similar patterns of GO enrichment to Experiment 1, with some notable differences. Aromatic amino acid biosynthesis remained the most strongly overrepresented (~8.3 fold) within the biological process domain for upregulated genes, with oxidative phosphorylation, glycolysis, and ion transport also included among the top ten (Table 3.17). Some categories found in the top ten list in Experiment 1, such as amine metabolic processes, were significant but not in the top ten in Experiment 2. Steroid metabolic process,

photosynthesis light reactions, and RNA and DNA related processes were overrepresented in Experiment 2 but were absent or had lower fold-enrichment in Experiment 1. Within the molecular function domain, overrepresented categories again showed some similarity with Experiment 1, including amine and amino acid binding (both ~4.1 fold), hydrogen ion transport, and transferases of nitrogenous groups (Table 3.18). Two oxidoreductase activity categories showed similar fold-enrichment between the two experiments but did not make the top ten in Experiment 1. The most strongly enriched category in the molecular function domain for Experiment 2, hydro-lyase activity (>4.5 fold), was not in the REViGO-reduced list for Experiment 1, although it was found in the non-reduced list. Meanwhile, carboxylic acid binding (~3.6 fold) had a much lower fold-enrichment in Experiment 1. Finally, threonine peptidase categories were not enriched in Experiment 2.

Among genes downregulated by MeJA, overrepresented GO categories showed less overlap between experiments. The overall enrichment patterns were weaker than for upregulated genes, as in Experiment 1. The most overrepresented categories in Experiment 2 were for regulation of Rab GTPase activity (>3.8 fold), glycoside and oligosaccharide metabolism, coenzyme biosynthesis, and protein modification (Table 3.17). The secretion and localization categories from Experiment 1 were not overrepresented in Experiment 2. Molecular functions overrepresented among downregulated genes included two categories each related to signal transduction and ion transport, along with one category each for lipid binding, oxidoreductases, transferases of nitrogenous groups, *O*-methyltransferases, and hydrolyases of acid anhydrides (Table 3.18). Thus, while overrepresented GO categories for genes responding to MeJA did

show overall similarities between the two experiments, the discrepancies indicated differences in response as well.

In Experiment 2, genes significantly upregulated by AOPP showed a broad set of enriched GO categories, many of which were also overrepresented in MeJA-upregulated genes. The top ten overrepresented biological process categories included aromatic amino acid biosynthesis (>16 fold), two coenzyme metabolism categories (~11.1-fold and ~4.7-fold), oxidative phosphorylation, cellular aromatic compound metabolism, and transmembrane transport (Table 3.17). The most strongly overrepresented molecular function categories included hydrogen ion transmembrane transporter activity (~8.4 fold), carbon-oxygen lyase activity, transferases for alkyl/aryl or nitrogenous groups, oxidoreductase activity, and five different binding activity categories (Table 3.18). Among genes downregulated in response to AOPP, biological processes relating to carbohydrate metabolism were strongly overrepresented, especially trehalose biosynthesis (>8.8 fold) and glycoside metabolism, as well as three protein related categories (Table 3.17). Overrepresented molecular functions for AOPP-downregulated genes included ATPase activity (~5.4 fold), oxidoreductases, transferases of nitrogenous groups, and two categories each relating to protein degradation and glycosyl hydrolase activity (Table 3.18). Only oxidoreductase activity was overrepresented among genes exhibiting a positive AOPPxMeJA interaction. Four categories relating to ATP binding (i.e., traceable to the same set of seven probes; data not shown) were overrepresented among genes with significant negative interactions (Table 3.18).

Differential Transcriptional Responses of Core Phenylpropanoid Biosynthetic Genes

Expression of genes involved in the core phenylpropanoid pathway was next surveyed, because these genes have previously been shown to exhibit upregulation under MeJA treatment and because their encoded proteins were targeted for inhibition in this study. Understanding transcriptional responses of the corresponding genes to metabolic perturbation may reveal additional feedback loops modulating phenylpropanoid homeostasis. In Experiment 1, phenylpropanoid core pathway genes *PAL2*, *PAL3*, *C4H1*, *C4H2*, and *4CL2* were upregulated in MeJA-treated cells regardless of PIP status, but they did not exhibit a significant response to PIP or an interaction between MeJA and PIP (Table 3.19). QPCR analysis was conducted on two isoforms each for *PAL*, *C4H*, and *4CL* to cross-check general microarray observations. This also allowed us to probe for differential isoform responses. Since the metabolic inhibitors act at the protein level, for example, they may influence expression of different isoforms of the same gene in different ways. Statistical analysis by two-way ANOVA indicated that MeJA treatment overall in Experiment 1 led to significant changes in expression for all six genes (largest p-value for the set of 6 genes was 0.0162 in *PAL2*). Five of the genes were upregulated under MeJA, with the strongest magnitudinal changes for the more highly expressed genes *PAL1* (2.4-fold), *C4H1* (2.0-fold), and *C4H2* (1.8-fold; Figure 3.13A). For the more weakly-expressed genes, *PAL2* expression increased about 20% and *4CL2* expression about 60% under the same conditions, while *4CL1* was downregulated by about 12% (Figure 3.13B). Across MeJA treatments, PIP feeding caused upregulation of *C4H1* expression by about 19% (p=0.0024), but trends towards upregulation of *PAL1* and downregulation of *4CL2* were not statistically supported (Figure 3.13, panels A&B).

Furthermore, no significant interactions between MeJA and PIP feeding were found. Thus, although C4H isoforms exhibited differential changes in expression under PIP feeding, this inhibitor had only minor impacts on core phenylpropanoid gene expression.

In Experiment 2, microarray analysis showed that MeJA feeding had a significant effect on several core phenylpropanoid genes. *C4H2* and *4CL2* were upregulated, but *PAL4*, *4CL1*, and *4CL3* exhibited significant downregulation (Table 3.20). AOPP treatment led to significant increases in *PAL4*, *C4H2*, and *4CL2* expression regardless of MeJA treatment status, although no significant interactions between MeJA and AOPP were observed (Table 3.20). QPCR analysis of the same six core phenylpropanoid genes tested in Experiment 1 revealed similarities in response to MeJA feeding between the two experiments, with downregulation of *4CL1* by over 36% and upregulation of *PAL1* (1.9-fold), *C4H1* (1.8-fold), *C4H2* (~45%), and *4CL2* (~40%) in MeJA-treated samples overall; the modest increase in *PAL2* expression in Experiment 1 under MeJA feeding was not observed in the second experiment (Figure 3.13, panels C&D). Responses to AOPP feeding consisted of a significant increase in expression for all genes when not accounting for MeJA status (largest p-value within the set was 0.001 for *4CL2*; Figure 3.13, panels C&D). Significant AOPPxMeJA interactions were observed for *PAL1* ($p=0.0016$), *C4H1* ($p=0.001$), *C4H2* ($p=0.0029$), and *4CL1* ($p=0.001$). *PAL1*, *C4H1*, and *C4H2* all showed greater than additive increases in gene expression when treated with both AOPP and MeJA than when fed either one alone (Figure 3.13C). In contrast, the up-regulation of *4CL1* expression under AOPP feeding was abolished in the presence of MeJA (Figure 3.13D). Together, these results suggest that isoforms of PAL and 4CL, but not C4H, were differentially regulated during AOPP and MeJA feeding. Across both

experiments, the data supported the idea that MeJA feeding had a stronger influence on phenylpropanoid core pathway gene expression than did inhibitor feeding, while also providing evidence for significant differences in isoform expression and thus differential regulation of core pathway gene family members.

Expression of Monolignol Biosynthetic Genes is Influenced Mainly by MeJA

Genes involved in monolignol biosynthesis were generally expressed at lower levels than core phenylpropanoid biosynthetic genes in the cell suspension cultures, consistent with the undifferentiated nature of the cells. In Experiment 1, several monolignol biosynthetic genes were influenced by MeJA feeding, but no significant effects of PIP or interactions between PIP and MeJA feeding were found (Table 3.21). Both *C3H3* and *CCoAOMT2* were upregulated by 20% or more in MeJA-treated cells, while *COMT1*, *HCT6*, and *Lac90b* were all downregulated by at least 40% (Table 3.21).

Similar patterns for MeJA feeding were observed in Experiment 2, with upregulation of *C3H3* along with downregulation of *HCT6* and *Lac90b* (Table 3.22). In addition, *CCoAOMT1* exhibited a modest (~14%) but significant decrease in expression under MeJA treatment. AOPP feeding led to decreases in *F5H2* expression of about 32% (Table 3.22). No significant interactions were found between AOPP and MeJA feeding. Overall, microarray data supported the idea that MeJA feeding impacted expression of monolignol biosynthetic genes, with a minor impact of AOPP.

MeJA-Stimulated Upregulation of Flavonoid/CT Biosynthetic Genes is Differentially Modulated by Metabolic Inhibitors

Flavonoid pathway genes responded in patterns similar to those observed for core phenylpropanoid and monolignol biosynthetic genes, with stronger responses to MeJA than to inhibitor feeding. In Experiment 1, MeJA feeding led to numerous statistically significant changes in transcript levels, generally associated with upregulated genes (*CHS2/3*, *CHS6*, *CHI*, *F3'H*, *F3H*, *F3'5'H1*, *DFR1*, *BAN2*, *LAR2*), with the exception of a putative *FOMT1,* which was downregulated (Table 3.23). Genes for which statistical significance varied depending on the probe (*CHIL2* and *ANS2*) showed strong (about 1.9 to 4.6 fold), consistent upward trends in non-significant probes. No flavonoid-related genes showed significant responses to PIP treatment overall, nor did any show a significant interaction between PIP and MeJA feeding (Table 3.23).

Four genes, *CHI*, *F3'H*, *LAR2*, and *BAN1*, were selected for QPCR analysis using additional biological replicates. In the microarray data, all but *BAN1* showed statistically significant responses to MeJA. For QPCR analysis, all four genes were significantly upregulated by MeJA overall (largest p-value 0.0001, for *F3'H*), with increases ranging from ~2.2- to ~3.5-fold (Figure 3.14, panels A&B). The two late-pathway genes *LAR2* and *BAN1* are involved with the synthesis of 2,3-*trans*-flavan-3-ols (e.g., catechin) and 2,3-*cis*-flavan-3-ols (e.g., epicatechin), respectively, both of which are CT precursors. While *BAN1* was expressed at levels about an order of magnitude greater than *LAR2*, both genes exhibited significant responses to PIP feeding without considering MeJA status (p=0.0025 and 0.0378, respectively) and PIPxMeJA interactions with strong or marginal statistical support (p = 0.0107 and 0.0633, respectively). The interaction

effects, however, showed opposite trends, with PIP enhancing the upregulation of *BAN1* by MeJA but attenuating MeJA-related upregulation of *LAR2* (Figure 3.14, panels A&B) Treatment with PIP in cells not fed MeJA did not lead to strong changes in expression in either case.

Microarray results for Experiment 2 showed, as for Experiment 1, that MeJA had a broader overall effect on flavonoid-related gene expression than did inhibitor feeding (Table 3.24). Again, the majority of significant changes in MeJA-fed cells were associated with upregulated gene expression, although which genes passed the significance threshold different slightly (*CHS4*, *CHS6*, *CHI*, *CHIL2*, *F3'H*, *F3'5'H1*, *ANS1/2*, *ANS2*, *BAN1*, *LAR1*, *LAR2*, *LAR3*, and *FOMTL1*; Table 3.24). MeJA-downregulated genes included those that have not been experimentally established as associated with flavonoid biosynthesis, such as *FLR*, *FOMT1*, *FOMT2/7*, and *FOMT7*. AOPP feeding in Experiment 2 had a greater effect on flavonoid-related gene expression than did PIP feeding in Experiment 1. In particular, AOPP led to significant increases in expression of *CHS6*, *CHI*, *F3'H*, *ANS1/2*, *BAN1*, *LAR2*, and *FOMTL1* (Table 3.24). *F3'5'H1* expression was also higher in AOPP+MeJA treated cells than in cells fed only MeJA, although it was below the detection threshold in other treatments. A putative *FLR* was the only gene significantly downregulated by AOPP overall. No statistical support was found for AOPPxMeJA interactions in the microarray data.

Confirmatory analysis by QPCR largely supported the microarray results. Three of the four genes were significantly upregulated by MeJA overall (largest p-value was 0.0021 for *LAR2*), and *BAN1* showed trends toward increased expression ($p=0.0754$; Figure 3.14, panels C&D). A similar pattern was observed for AOPP overall ($p<0.0001$

for all three), but *BAN1* did not show a significant trend (p=0.1204). Finally, a significant interaction effect between AOPP and MeJA treatments on the expression of *CHI*, *F3'H*, and *LAR2* was observed, yielding p-values of 0.0042, <0.0001, and 0.0036, respectively. In all three cases, the interaction was a synergistic, rather than additive, increase in gene expression when cell cultures were fed both compounds (Figure 3.14, panels C&D).

Some Nitrogen and Amino Acid Related Genes Respond to Both MeJA and Metabolic Inhibitors

To determine whether the observed differential treatment effects on soluble protein and amino acid levels could be attributed to transcriptional changes, microarray data associated with ammonium, nitrate, and amino acid metabolism were examined. Some genes implicated in both aromatic amino acid metabolism and phenylpropanoid metabolism via the shikimate pathway were included as well. As for phenylpropanoid-related genes, the most consistent responses were seen for MeJA treatment. In Experiment 1, genes significantly upregulated by MeJA overall included a chorismate mutase (*CM1*), two arogenate/prephenate dehydratases (*ADT/PDT1 & 2*), a tryptophan synthase β-chain (*TSB4*), an ammonium transport protein gene (*AMT1*), and the chloroplastic glutamine synthetase gene *GS2* (Castro-Rodriguez et al. 2011; Table 3.25). Significantly downregulated genes included a high-affinity nitrate transporter (*NRT2.1*; CJ Tsai, personal communication) and two cytosolic glutamine synthetases (Castro-Rodriguez et al. 2011). Only *anthranilate synthase β-chain 2* (*ASB2*) was significantly upregulated in response to PIP feeding across samples, but this and several other genes

exhibited significant interactions between MeJA and PIP feeding. As seen for the effects of PIP on gene expression in other pathways, the magnitudinal changes were relatively small. For three of these genes, *ADT/PDT1*, *ASB2*, and *OMR1*, statistical significance depended on which probe was considered. This may be related to Affymetrix probe design redundancy (Tsai et al. 2011) in that sequence heterogeneity between the experimental (*P. trichocarpa*) species and that used for array design (*P. tremuloides*) may affect hybridization efficiency in a probe-dependent manner. Verification by a second, independent analysis would be necessary to confirm trends for *ADT/PDT1* and *OMR1*. *ASB2* expression was more consistent in that both probes showed similar intensities and a subtle increase in PIP-treated cells not fed MeJA that was further increased in MeJA-fed cells. Three additional genes exhibiting PIPxMeJA interaction, an indole-3-glycerol phosphate synthase (*IGPS1*), *AMT1*, and a putative ammonium transporter gene all showed relatively weak (9-28%) downregulation due to PIP treatment in cells not fed MeJA and moderate (31-45%) upregulation due to PIP treatment in MeJA-fed cells (Table 3.25). While the strongest probe for *Nitrate reductase 2* (*NR2*) was not found in the DE List, a weaker probe was upregulated by 46% due to PIP treatment in cells not fed MeJA and downregulated by a similar amount in MeJA-fed cells (Table 3.25).

In Experiment 2, MeJA responses were similar to those seen in the first experiment, with significant upregulation of *CM1*, *ADT/PDT1 & 2*, *TSB4*, *AMT1*, and *GS2* (Table 3.26). A tryptophan synthase α-chain (*TSA1*), two additional β-chain isoforms (*TSB2/3*), and a second putative ammonia transporter isoform not identified in Experiment 1 were also upregulated under MeJA treatment. More genes were significantly downregulated under MeJA treatment in Experiment 2 than in Experiment

1. A *p*-aminobenzoate synthase/glutamine amidotransferase (*PABAS*), an anthranilate phosphoribosyltransferase (*PrAT1*), and two NADH-dependent glutamate synthase-like isoforms showed decreases in MeJA-treated cells of <25% (Table 3.26), while *NR1*, *NRT2.1*, and *NRT2.7* decreased by 39-53%. A modest response to AOPP feeding was observed similar in extent to that for PIP feeding in Experiment 1, although no significant interactions between AOPP and MeJA were identified. *ADT/PDT1 & 3*, *PrAT1*, and *OMR1* were significantly upregulated in AOPP-fed cells (Table 3.26). The largest change among these genes was for *ADT/PDT1*, which showed 27-45% increases in expression depending on which probe was examined. The only gene significantly downregulated by AOPP feeding overall was *GS1.1a*, reduced by 25-33%.

The *Populus* genome contains two nitrate reductase genes, but microarray data suggested their responses to MeJA feeding varied between the two experiments. *NR1* responded significantly to MeJA only in Experiment 2. Meanwhile, of the two probes for *NR2*, the one with greater signal did not show significant changes in either experiment (data not shown), and the probe with weaker signal was significantly influenced by MeJA overall and by PIPxMeJA interaction only in Experiment 1. QPCR was used to validate these expression patterns. Both methods showed that *NR2* is expressed more strongly than *NR1* in *Populus* cell suspension cultures. QPCR revealed that in Experiment 1, *NR1* responded to MeJA overall and to MeJAxPIP interaction, but not to PIP. In particular, the 43% downregulation of *NR1* by MeJA in cells not fed PIP (p=0.0051) was abolished when PIP was co-fed (p>0.8; Figure 3.15A). In Experiment 2, *NR1* was downregulated by 34% in response to MeJA overall (p=0.0016), but did not exhibit an MeJAxAOPP interaction, instead showing an increasing trend due to AOPP overall (p=0.0503; Figure

3.15B). In constrast to *NR1*, *NR2* was upregulated by 43% in Experiment 1 in response to MeJA overall and by 48% in response to PIP overall (p=0.0154 and p=0.0091, respectively; Figure 3.15A), but did not show a significant MeJAxPIP response. In Experiment 2, *NR2* trended towards upregulation in response to MeJA overall (p=0.0788), but showed no significant response to AOPP or MeJAxAOPP (Figure 3.15B). Overall, the QPCR showed similar trends but different statistical support compared to the microarray data. The discrepancy can be ascribed to both greater specificity (QPCR primer design took into account species-specific sequence variations) and additional biological replication of the QPCR assay relative to the microarray approach. Thus, nitrate reductase isoforms in both experiments exhibited differential responses to both MeJA and inhibitor feeding.

Genes Related to Citric Acid Cycle Metabolites Respond to MeJA and AOPP

Genes related to citric acid cycle metabolites (i.e., genes involved in mitochondrial citric acid cycle as well as genes associated with metabolism of these organic acids outside the mitochondria) showed further evidence for stronger transcriptional responses to MeJA feeding than to inhibitor feeding. In Experiment 1, a citrate synthase, a malate translocator, four ATP-citrate lyase isoforms, a putative isocitrate dehydrogenase, an NAD+ dependent isocitrate dehydrogenase-like gene, two pyruvate dehydrogenase E1β isoforms, and a succinate dehydrogenase iron protein subunit-like gene were all significantly upregulated by MeJA overall. The strongest and most consistent changes were noted for the ATP-citrate lyase isoforms, for which expression nearly doubled under MeJA treatment in most cases (fold changes ranging

from 1.5-2.4, depending on the probe; Table 3.27). Citrate synthase and the NAD+ dependent isocitrate dehydrogenase-like gene both showed 1.6-fold increases as well. In contrast, a succinyl-CoA ligase β-like gene was the only gene downregulated by MeJA, by about 25% overall (Table 3.27). None of these genes showed a significant response to PIP feeding overall, but three, a mitochondrial malate dehydrogenase, an ATP citrate lyase, and an isocitrate dehydrogenase, exhibited a significant PIPxMeJA interaction effect (Table 3.27). Interestingly, the malate dehydrogenase gene was not among the genes responding significantly to MeJA overall. For all three genes, no clear response to PIP feeding was seen in cells not fed MeJA, but trends towards upregulation were seen in response to PIP feeding in MeJA-fed cells (p=0.0936 with 76% increase for malate dehydrogenase, p=0.0845 or 0.1062 with 18-50% increase for isocitrate dehydrogenase, and p=0.0514 and 25% increase for ATP citrate lyase; Table 3.27). Thus, a synergistic response for the two treatments may occur for these genes (Table 3.27).

In Experiment 2, many of the genes related to citric acid cycle metabolites exhibited an overall significant response to both AOPP and MeJA feeding, although the effects of MeJA were generally greater in magnitude (Table 3.28). Genes responding significantly to MeJA were largely the same as those in Experiment 1. Most discrepancies usually involved weakly expressed genes (e.g., malic enzyme genes and a β-alanine/pyruvate aminotransferase-like gene) or probes exhibiting smaller fold-changes (e.g., a microbody NAD+ dependent malate dehydrogenase and a distinct NAD^+ dependent isocitrate dehydrogenase-like gene; Table 3.28). Among genes responding to MeJA, significant upregulation in response to AOPP were also seen for the four ATP citrate lyase isoforms (17-45%), one pyruvate dehydrogenase E1β-like isoform (17-

22%), the microbody NAD+ dependent malate dehydrogenase (~11%), and the β-alanine/pyruvate aminotransferase-like gene (~21%; Table 3.28). Two mitochondrial malate dehydrogenases (15-24%), a citrate synthase (*CIT1*, 9-17%), a succinyl-CoA synthetase (11%), a cytosolic malate dehydrogenase (17%), an additional pyruvate dehydrogenase isoform (12%), and a succinate dehydrogenase subunit (9%) also showed small but significant increases due to AOPP feeding, but did not show clear responses to MeJA (Table 3.28). The only significant AOPPxMeJA interaction was seen in one pyruvate dehydrogenase isoform, which showed relatively weak upregulation under AOPP or MeJA alone, but a synergistic effect when the treatments were combined (~5% vs. ~26%; p=0.0075 for AOPP response in MeJA-fed cells). Overall, AOPP and MeJA feeding seem to have similar, largely additive, impacts on genes linked directly or indirectly to the citric acid cycle.

Discussion

Efficacy of Elicitor and Inhibitor Treatments in Perturbing Phenylpropanoid Metabolism

MeJA elicitation was used here to stimulate the phenylpropanoid pathway in heterotrophic *Populus* cells as a model system to investigate partitioning of phenylpropanoid carbon. Feeding metabolic inhibitors for distinct core phenylpropanoid pathway steps alone or in conjunction with MeJA allowed for further dissection of the effects of metabolic perturbation. These treatments can be considered as tools to increase phenylpropanoid pathway flux as a whole (MeJA) or decrease it at specific enzymatic steps (AOPP, PIP, MDCA), during which "snapshots" were taken using a single harvest

time to assess treatment effects by metabolite and transcript profiling (Figure 3.6). Demonstrating treatment efficacy was necessary before assessing the broader effects of perturbation.

Glycosylated jasmonates have been reported across a range of plant taxa (Miersch et al. 2008; Simko et al. 1996; Yoshihara et al. 1989). The identification of a putative jasmonate specific to deglycoslyated extracts of MeJA-fed cells supports the idea that these cells took up and metabolized MeJA. The broad responses to MeJA feeding, including increased accumulation of CTs and upregulation of several core phenylpropanoid and aromatic amino acid metabolism genes, are consistent with previously reported effects of jasmonates (Arnold et al. 2004; Arnold and Schultz 2002; Babst et al. 2009; Pauwels et al. 2008). The slight but statistically significant differences in glycosylated jasmonate levels across the two experiments could indicate variation in jasmonate metabolism or deactivation (Koo et al. 2011; Miersch et al. 2008; Seto et al. 2009; Suzuki et al. 2007). If so, they would contribute to other discrepancies in MeJA responses between the two experiments. We interpret the overall results as indicating successful MeJA elicitation, with experimental differences partly reflecting extent of elicitation at harvest.

Free and glycosylated PIP and MDCA were both identified in the B and B-DG fractions of treated cells, respectively. The higher levels of these inhibitors in B-DG relative to the B fraction suggested that these compounds were metabolized by the cells. Although we did not identify a clear AOPP peak in the B fraction, the α-oxo-benzenepropanoic acid peak specific to the B-DG fraction in AOPP-fed samples represents a putative deamination and glycosylation product of AOPP. To our

knowledge, neither glycosylation nor metabolic fate of AOPP, PIP, or MDCA has been previously described in plants.

Both AOPP and PIP treatments generated the expected effects on substrate and product levels for PAL and C4H, respectively, supporting the overall efficacy of these treatments. AOPP-treated cells showed drastic increases in phenylalanine, consistent with previous studies (Amrhein et al. 1976; Havir 1981). Cinnamate was not influenced, but the next downstream metabolite, *p*-coumarate, was decreased. The lack of cinnamate response to PAL inhibitors has been previously observed in plant cell culture systems (Orr et al. 1993) and is consistent with a role for cinnamate in modulating PAL activity and gene expression, although intracellular compartmentalization also likely plays a role (Bolwell et al. 1988; Mavandad et al. 1990; Orr et al. 1993). In PIP-treated cells, cinnamate was strongly increased and *p*-coumarate decreased, as expected for C4H inhibition (Schalk et al. 1998; Schoch et al. 2002). MDCA-fed cells did not exhibit significant increases in *p*-coumarate as previously observed (Sircar and Mitra 2009), nor were the downstream hydroxycinnmates caffeate and ferulate affected. However, the upstream metabolite cinnamate increased as seen in PIP-fed cells, a pattern not reported in studies where MDCA inhibitory efficacy was demonstrated (Funk and Brodelius 1990; Sircar and Mitra 2009). Considered with the weak effects of MDCA on CTs and lignin, this suggested the treatment as applied in this study was less effective than AOPP and PIP. The shift in cinnamate could be attributable to thermal breakdown of MDCA or one of its catabolic products during GC analysis rather than true cinnamate *in vivo*. Testing pure MDCA using the analytical procedures employed here would help ascertain whether this is the case.

Influence of Methyl Jasmonate and Phenylpropanoid Inhibitors on Partitioning Between Lignin and Condensed Tannins

Quantifying lignin and CTs can provide a measure of experimental perturbation based on relatively stable phenylpropanoid end products. Lignin is typically polymerized in the secondary cell wall (Vanholme et al. 2010), making it a low-turnover carbon pool. CTs may undergo some slow metabolic turnover, at least in developing tissues (Kleiner et al. 1999), but they comprise a large carbon pool in the form of a complex polymer thought to be relatively stable (Reichardt et al. 1991; Swain 1979), sequestered in the vacuole or associated with proteins and cell wall components (Reed 1986; Stafford 1988). Although other wall-bound phenolics may also constitute a sizeable, low-turnover phenylpropanoid sink, we did not quantify such pools here.

CT levels and most flavonoid pathway genes were upregulated by MeJA elicitation, suggesting increased metabolite flux through part of the flavonoid network towards CT biosynthesis. This is consistent with *de novo* synthesis of CTs in jasmonate-treated *Populus* (Arnold et al. 2004; Arnold and Schultz 2002) and with the known role of MeJA as a phenylpropanoid pathway elicitor (An et al. 2006; Pauwels et al. 2008; Zhao et al. 2005). Interestingly, the more lignin-related *4CL1* and flavonoid/CT-related 4CL2 exhibited contrasting responses to MeJA: *4CL2* transcripts increased but *4CL1* transcripts decreased. This pattern is consistent with the interpretation of preferential allocation of hydroxycinnamoyl-CoA conjugates towards CTs rather than lignin in elicited cells.

All three inhibitor treatments reduced CT levels, consistent with the hypothesis that CT accumulation is sensitive to flux limitations at early phenylpropanoid pathway

steps. However, the effects of PIP and AOPP on flavonoid/CT pathway gene expression differed. PIP had relatively little influence on these genes, while AOPP feeding led to increased transcripts of many flavonoid genes, hinting at possible metabolic reprogramming (see also below). QPCR also detected differential inhibitor responses of two late-pathway genes, *LAR2* and *BAN1*, that are involved in the synthesis of trans- and cis-isomers, respectively, of 2,3-flavan-3-ol, the monomers for CT polymerization (Tanner et al. 2003; Tanner and Kristiansen 1993; Xie et al. 2003). The relative proportions of each type of monomer can vary during plant development and contribute to structural and compositional variation in CTs (Gagné et al. 2009). In *Populus*, monomeric composition and chain length of CTs are known to be primarily under genetic control (Scioneaux et al. 2011), and such variation has been attributed to differential expression of *LAR* and *BAN* genes in some plant taxa (Akagi et al. 2009; Bogs et al. 2005). PIP feeding here led to upregulation of *BAN1* and downregulation of *LAR2* in MeJA-cofed cells relative to cells fed only MeJA. In contrast, AOPP stimulated upregulation of only *LAR2*, a response further potentiated by MeJA cofeeding. It appears that synthesis of 2,3-flavan-3-ol isomers may be sensitive to the position at which phenylpropanoid pathway flux is restricted. The current findings overall suggest that CT accumulation and composition are therefore influenced by phenylpropanoid pathway perturbation.

In contrast to CTs, lignin content and composition changed little in response to MeJA in our experiments, along with most lignin biosynthetic pathway transcripts. The exceptions were *HCT6* and *lac90b*, which were strongly reduced in MeJA-treated cells. Upregulated lignin biosynthetic genes and transient mono- and oligolignol accumulation

have previously been observed in *Arabidopsis* cell cultures treated with methyl jasmonate (Pauwels et al. 2008), although lignin content was not measured. The discrepancy may reflect differences in in phenylpropanoid metabolism in the two species, as CTs comprise a quantitatively significant in phenylpropanoid pool in cultured *Populus* but not *Arabidopsis* cells. Lignin biosynthesis was reduced, with concomitant increases in flavonoid biosynthesis, in *HCT*-silenced *Arabidopsis* (Besseau et al. 2007). Since it was the most strongly expressed *HCT* assessed by microarray, reduced *HCT6* expression seems a likely contributor to reduced *de novo* synthesis of lignin in *Populus* cell cultures here (Tsai et al. 2006b), although the corresponding enzyme has not been biochemically characterized.

Lignin accumulation and composition were little influenced by enzyme inhibition, with only AOPP feeding causing a small decrease in relative lignin content. Relative to this short-term assay, an *Arabidopsis pal1/pal2* double mutant exhibited stronger reductions in lignin and increased S/G ratio (Rohde et al. 2004). Interestingly, AOPP caused both an increase in (QPCR-assessed) *4CL1* expression in the absence of MeJA and a decrease in (microarray-assessed) expression *Lac90b*. Based on demonstrated differences in isoform function linking 4CL1 more closely to lignin and 4CL2 to flavonoid/CT biosynthesis (Hu et al. 1998; Kao et al. 2002), increased *4CL1* expression and *4CL1/4CL2* ratio in cells fed only AOPP predicts a tendency towards the maintenance of lignin pools when flux into the phenylpropanoid pathway at PAL is reduced *in planta*. Overall, the methods of phenylpropanoid perturbation used in this study appear to have differential effects on low-turnover phenylpropanoid pools, with

limited consequences for lignin biosynthesis and stronger influences on CTs in *Populus* cell cultures.

Hydroxycinnamates as Indicators of Flux Through the Core Phenylpropanoid Pathway

Levels of phenylalanine, caffeate, and ferulate were generally reduced under MeJA treatment, with consistent patterns between the free acid and "total" pools of the two hydroxycinnamates. In comparison, free and total *p*-coumarate levels were much higher and responded to MeJA more variably across experiments. Since *p*-coumarate is located near a branch point between competing lignin, flavonoid/CT, and hydroxycinnamate-conjugate biosynthesis pathways, this may be linked to the previously-noted differential sensitivity of these branches to elicitation, possibly arising from different activity ratios among 4CLs, CHSs, and HCTs. Considered in conjunction with the general pattern of increased expression of core pathway and flavonoid network genes and greater CT accumulation, these data likely indicate increased utilization of core phenylpropanoids during a MeJA-directed partitioning favoring CT synthesis. Increased expression of core phenylpropanoid pathway genes under jasmonate elicitation has previously been reported in *Populus* and other plants (Babst et al. 2009; D'Onofrio et al. 2009; Pauwels et al. 2008; Xiao et al. 2009), although this pattern may not be universal (e.g., Suzuki et al. 2005). Given the requirement of HCT for synthesis of caffeic and ferulic acids (Hoffmann et al. 2004; Hoffmann et al. 2003), lower *HCT6* expression was consistent with their lower abundance. However, possible utilization of their CoA-conjugates in flavonoid/CT biosynthesis (Tsai et al. 2006b) could not be ruled out here. Direct surveys of hydroxycinnamates under jasmonate elicitation are scarce but do

suggest that responses can be species- and system-dependent (D'Onofrio et al. 2009; Xiao et al. 2009).

In contrast to phenylalanine and the hydroxylated cinnamic acid derivatives, cinnamic acid itself tended to remain stable under MeJA treatment. In PIP-fed cells, both free and total cinnamic acid levels were increased, with greater-than-additive increases in total cinnamic acid when MeJA was also present. In keeping with a role for cinnamate in feedback-regulation of PAL (Blount et al. 2000), phenylalanine levels were reduced in PIP-fed cells. The previously mentioned reductions in phenylalanine in MeJA-fed cells, likely due to draw-down under increased pathway flux, were not further depressed by PIP in cofed cells despite cinnamic acid accumulation. *PAL* expression was not influenced by PIP regardless of MeJA status, contrasting with previous studies (Blount et al. 2000; Mavandad et al. 1990; Orr et al. 1993), although PAL activity was not measured, limiting a full assessment of cinnamate feedback regulation in the current study. Nevertheless, on the basis of increased CT levels in cofed cells relative to those fed only PIP, it appears that MeJA elicitation is capable of overriding cinnamate feedback effects.

Expression of *C4H* isoforms 1 and 2 showed minor differences in their response to PIP feeding, with *C4H1* showing small but statistically significant upregulation detected by QPCR and the lignin-associated *C4H2* (Lu et al. 2006) showing no significant response. Both isoforms responded similarly to MeJA and to AOPP according to QPCR, however, suggesting their differential sensitivity to PIP may be related more strongly to enzyme-inhibitor interactions than to metabolic perturbation. Possible differences in PIP access to or inhibitory efficacy for distinct C4H isoforms may exist, with likely consequences for enzyme turnover. For example, differential physical

association of C4H isoforms with PAL, previously reported to result in metabolic channeling of phenylalanine to *p*-coumarate (Achnine et al. 2004; Rasmussen and Dixon 1999), would likely reduce PIP access to the more strongly PAL-associated isoform.

AOPP-stimulated increases in phenylalanine were partially counteracted by co-feeding of MeJA. AOPP feeding also led to significant decreases in levels of both free downstream hydroxycinnamates, with the exception of free and total cinnamic acid, which were unchanged relative to controls. Most core phenylpropanoid pathway transcripts were increased by AOPP feeding, with four of these exhibiting synergistic upregulation in MeJA cofed cells. In contrast, AOPP cofeeding did not influence the decrease in *4CL2* transcripts identified by QPCR for MeJA fed cells. Given increased expression of several CT-related flavonoid/CT biosynthetic genes (see below) and reduced lignin and CT accumulation in AOPP-fed cells, this seems to suggest a compensatory transcriptional upregulation due to AOPP restriction of PAL in order to maintain phenylpropanoid synthesis, with CTs favored over lignin. Recent work has identified flavonol-dependent feedback inhibition of *PAL*, *CHS*, and flavonol biosynthetic transcripts using *Arabidopsis* mutants (Yin et al. 2012), indicating the possibility that flavonoid levels can influence transcription in the core phenylpropanoid pathway.

In the present work, all four core phenylpropanoids surveyed were more abundant in the B-DG fraction than in the B fraction, consistent with the presence of glycosylated hydroxycinnamates in the cell cultures. Glycosylation of hydroxycinnamates in the form of glycosides and glucose esters is widespread in plants (Fraissinet-Tachet et al. 1998; Harborne and Corner 1961; Lim and Bowles 2004; Lu and Yeap Foo 2000; Nagels et al.

1981; Paquette et al. 2003; Tian et al. 2005; Wu et al. 2003). Glycosylation also occurs readily in plant cell cultures with exogenous phenolic substrates (Harborne and Corner 1961; Nagels et al. 1981; Payyavula et al. 2009). In addition to modulating metabolic homeostasis, glycosylation is thought to provide a general detoxification mechanism in plants for both autotoxic defensive compounds and xenotoxic compounds such as herbicides (Lim and Bowles 2004; Vaistij et al. 2009; Wink et al. 1997). From the present work, the trends in the proportion of glycosylated hydroxycinnamates under metabolic inhibition are consistent with a role of glycosylation in modulating hydroxycinnamate levels depending on pathway status. For example, AOPP treatment led to a lower proportion of free ferulic acid and a similar trend for free *p*-coumaric acid. PIP feeding led a decreasing trend in the proportion of free cinnamate while the total cinnamate pool increased, indicating increased glycosylation as cinnamic acid accumulated. Previously, increased activity of a cinnamic acid glucosyltransferase in *Phaseolus* cell cultures was observed upon cinnamate feeding (Edwards et al. 1990). Thus, hydroxycinnamate glycosylation in *Populus* appears to be constitutive and to respond to perturbed hydroxycinnamate levels.

Reorganization Between Flavonoid and CT Biosynthesis by Phenylpropanoid Perturbation

Gene expression data suggested broad activation of flavonoid biosynthesis in elicitor-treated cells in both experiments. Some evidence for distinct regulation patterns within the flavonoid biosynthetic "grid" was also found, consistent with well-established observations in *Arabidopsis* (Dixon et al. 2005). For example, the few genes exhibiting

reduced expression in MeJA-fed *Populus* cells were primarily *FOMT*s, putatively involved in methylation of chalcones or flavanones not known to contribute to CT pools (Tsai et al. 2006b). The ability of jasmonates to differentially regulate branch pathways within the flavonoid biosynthetic grid has not previously been reported. However, this may not be the case for other stressors. For example, *Populus* flavonol synthase (*FLS*) genes lack the wound-responsiveness of other flavonoid/CT pathway genes (Mellway et al. 2009; Tsai et al. 2006b). One *FLS* was nonrespsonsive to a fungal pathogen but rapidly activated by light and UV-B irradiation, nearly the opposite of flavonoid genes involved in CT biosynthesis (Mellway et al. 2009). Findings in other plant taxa also support distinct regulation between flavonols and downstream flavonoids such as anthocyanins and CTs (Martens et al. 2010). The current work is consistent with the idea that jasmonates may trigger a tailored transcriptional response within the flavonoid biosynthetic grid to promote CT biosynthesis in *Populus* cell cultures.

A putative 2,3-flavan-3-ol (catechin) in the B-DG fraction was reduced by over 50% in MeJA-fed cells in both experiments. In general, other flavonoids in the B-DG fraction also decreased but flavonoid aglycones (B fraction) were stable under MeJA treatment in Experiment 1, while in Experiment 2 B-DG fraction flavonoids were not influenced by MeJA and flavonoid aglycones tended to increase. This indicates a consistent depletion of glycosylated flavonoids across experiments. In light of the MeJA-activated transcription of flavonoid/CT biosynthetic genes, this is consistent with an interpretation of flavonoid network reprogramming favoring CT accumulation under elicitation. Similarly, transient decreases in anthocyanins (i.e., glycosylated

anthocyanidin flavonoids) are concurrent with CT increases following leaf wounding in *Populus* (Peters and Constabel 2002).

In accordance with the metabolite data, genes associated with glycoside metabolism were strongly overrepresented among those downregulated by MeJA in Experiment 2. Several members of the glycosyltransferase family 1 (GT1), which as a group are known to glycosylate a variety of phenylpropanoids (Bowles et al. 2006; Lim et al. 2002), were among the top forty most strongly MeJA-downregulated genes across both experiments. These included a putative tandem cluster of seven genes on linkage group (LG) XVII homologous to *Arabidopsis* UGT85s, and GT1-317 on LG III, most similar to UGT73s (Li et al. 2007; Yonekura-Sakakibara and Hanada 2011). Previous work has indicated that glycosyltransferase activity can influence CT accrual and CT/flavonoid partitioning. Reciprocal regulation of the glycosyltransferase *UGT78D2* and *BANYULS* in *Arabidopsis* has been shown to partition flavonoid intermediates between anthocyanin and CT biosynthesis (Lee et al. 2005). Transient silencing of an anthocyanidin glycosyltransferase in ripening strawberry fruits led simultaneously to decreases in anthocyanins and increases in CT monomers catechin and epicatechin (Griesser et al. 2008). Flavonoid glycosylation is thought to confer stability and/or facilitate membrane transport for vacuolar storage (Dixon et al. 2005; Vogt and Jones 2000). Glycosylation therefore plays a role in governing overall flavonoid pool composition. Our observations are consistent with an elicitor-induced, glycosylation-mediated partitioning favoring recruitment of flavonoid aglycones to CTs.

AOPP feeding generally reduced the abundance of "total" flavonoids and several flavonoid aglycones, with the former typically showing greater percentage reductions.

Meanwhile, MANOVA indicated that PIP feeding reduced total flavonoid levels in a manner not traceable to any specific peak when fed alone, but this collective effect was not seen when MeJA was co-fed, nor did it extend to aglycone peaks. Cells fed either inhibitor accumulated lower amounts of CTs relative to control cells. Cofeeding of PIP with MeJA also reduced, but did not prevent, the elicitor's stimulation of CT accrual. Preferential partitioning favoring CT over glycosylated flavonoids is supported by the >5-fold GO overrepresentation of glycoside metabolism among downregulated genes and by upregulation of a broad range of CT/flavonoid biosynthetic genes in AOPP-fed cells. The most strongly upregulated genes under either AOPP or MeJA treatment were *F3'H* and *CHI*, each increasing additively when the treatments were cofed. Genes involved in the terminal steps leading to 2,3-flavan-3-ols, *ANS1/2*, *LAR2*, and possibly *BAN1* (Tsai et al. 2006b), were also upregulated by both treatments. Together, these results support a model of partitioning favoring CT biosynthesis under restricted phenylpropanoid flux at an early step in the phenylpropanoid pathway, perhaps increasing the competitive strength of flavonoid biosynthesis relative to other phenylpropanoid pathway branches.

Influence of Phenylpropanoid Perturbation on Amino Acid Pools and Nitrogen Assimilation

Beyond the dramatic increases in phenylalanine already mentioned, tyrosine increased substantially in AOPP-fed cells. We considered a number of explanations, ranging from the possible existence of a tyrosine ammonia lyase (TAL) protein in *Populus* to the possibility that one or more *Populus* PAL isoforms may have tyrosine ammonia lyase activity (Hsieh et al. 2010; Louie et al. 2006; Rosler et al. 1997; Tzin and

Galili 2010; Watts et al. 2006). Given the structural similarity between AOPP, phenylalanine, and tyrosine, AOPP may have inhibited a TAL-like step, if one exists, in the *Populus* cells. AOPP may also have inhibited other tyrosine-utilizing enzymes, as reported for tyrosine decarboxylase and tryptophan aminotransferase (Chapple et al. 1986; Marques and Brodelius 1988; Soeno et al. 2010). Another possibility is that AOPP inhibition of PAL may lead to a diversion of phenylalanine towards tyrosine biosynthesis. Evidence for 4-hydroxylation of phenylalanine to yield tyrosine has been recognized in mammals, some bacteria, and occasionally plants (Endress 1981; Letendre et al. 1975; Nair and Vining 1965; Pribat et al. 2010; Scriver and Clow 1980; Yamamoto et al. 2001). Although AOPP is not as specific for PAL as once thought, our observations support a tight link between AOPP and tyrosine levels.

Gene expression data revealed complex regulation of aromatic amino acid metabolism in response to phenylpropanoid pathway perturbation. Along with several tryptophan synthase genes, multiple shikimate pathway genes (*CM*, several *ADT/PDT*s, *ASB2*) were upregulated in MeJA-treated cells, similar to the increased shikimate pathway transcripts seen in MeJA-fed *Arabidopsis* cell cultures (Pauwels et al. 2008). The upregulation of these genes in MeJA-fed cells are therefore consistent with increased post-chorismate shikimate pathway flux. Interestingly, AOPP feeding also led to upregulation of two *ADT/PDT*s, while PIP seemed to influence expression of *ADT/PDT1* in a MeJA-dependent manner. This is somewhat counterintuitive in that phenylalanine and tyrosine have been reported to negatively regulate the shikimate/aromatic amino acid pathway, although known mechanisms occur at the posttranscriptional level (Eberhard et al. 1996; Tzin and Galili 2010; Yamada et al. 2008). Gene upregulation could potentially

outweigh reported allosteric effects of elevated phenylalanine and tyrosine on ADT/PDT activity, thereby leading to late shikimate pathway contributions to these amino acid pools. However, we did not measure enzymatic activity here, and to our knowledge no previous work has suggested such an interpretation. Despite elevated *ADT/PDT* transcripts, AOPP feeding also reduced shikimate and phenylpyruvate levels, broadly consistent with shikimate pathway repression (Herrmann 1995; Herrmann and Weaver 1999; Maeda et al. 2010). Similarly, transcripts of several shikimate pathway genes were elevated concomitantly with decreased shikimate pools in *Petunia* with RNAi-suppressed *ADT1* (Maeda et al. 2010). Our observations support the previous suggestion of a complex interplay of posttranscriptional and metabolic factors in regulating aromatic amino acid biosynthesis (Maeda et al. 2010).

We observed three- to 17-fold increases in the non-aromatic amino acids β-alanine, 5-hydroxynorvaline, lysine, and isoleucine in AOPP-fed cells, while threonine and valine were significantly decreased by about 30%. Amino acids involved with the GS/GOGAT ammonium assimilation pathway or one-carbon metabolism were unaffected. Significant increases in levels of aromatic and most non-aromatic amino acids were also observed in *Arabidopsis pal1/pal2* double mutants relative to wild type plants (Rohde et al. 2004). The stronger effects relative to this study may be attributed to the greater severity of genetic perturbation on PAL function. Despite the relatively limited effects in AOPP-fed cells, soluble protein level was increased, supporting a cumulative effect of PAL inhibition on amino acid metabolism. Soluble protein was reduced by MeJA in both experiments despite inconsistent amino acid responses, in line with the finding that MeJA treatment inhibits overall protein synthesis in *Cucurbita*

cotyledons (Ananieva and Ananiev 1999). Biosynthesis of threonine, methionine, lysine, and the branched-chain amino acids valine, leucine, and isoleucine is transcriptionally regulated in response to several abiotic stressors (Joshi et al. 2010). Although many amino acid biosynthetic genes are not well annotated in the *Populus* genome, a putative threonine dehydratase/deaminase gene (*OMR1*) was upregulated in AOPP-fed cells. Increased OMR1 activity would be consistent with observed reductions in threonine, OMR's substrate, and increases in isoleucine, its product (Joshi et al. 2010).

Transcriptional changes also suggested altered ammonium reassimilation during phenylpropanoid perturbation, despite the lack of observed pereffects on glutamine and glutamate levels. Reduced expression of the cytosolic glutamine synthetase *GS1.1a* (Castro-Rodriguez et al. 2011) under AOPP feeding likely reflects a constraint of GS/GOGAT cycle flux in response to phenylalanine accumulation (Cantón et al. 2005). Cytosolic GSs are thought to represent the primary route for reassimilation of ammonium released by PAL catalysis (Bernard and Habash 2009; Cantón et al. 2005). However, only the chloroplastic *GS2* (Castro-Rodriguez et al. 2011) was upregulated in MeJA-fed cells despite phenylpropanoid pathway induction. Given their complex regulatory mechanisms (Bernard and Habash 2009), cytosolic GSs may be posttranscriptionally or posttranslationally regulated under MeJA elicitation. Altered expression indicates possible participation of GS2 in reassimilation of PAL-liberated ammonium in *Populus*. Several ammonium transporters were upregulated in MeJA-fed cells, suggesting a possible mechanism for ammonium movement between subcellular compartments. We are unaware of any previous reports of plastidic GS transcriptionally responding to MeJA or altered phenylpropanoid metabolism.

Expression of several nitrate-related genes was also influenced by phenylpropanoid perturbation. Two *NRT*s were downregulated in MeJA-treated cells despite non-limiting carbon and nitrogen supply in the cell culture medium, suggesting altered nitrogen uptake and utilization. Glutamine and ammonium have been reported to negatively regulate expression of *NRT*s (Cai et al. 2007; Thornton 2004; Vidmar et al. 2000), and altered tissue nitrate levels have been inversely correlated with accumulation of phenylpropanoids in tobacco (Fritz et al. 2006). Thus, MeJA elicitation in *Populus* cell cultures here may have reduced nitrate assimilation, either directly via transcriptional regulation or indirectly by phenylpropanoid accumulation or altered ammonium-nitrate ratios. Further supporting this idea, both annotated *NR* genes in *Populus* also exhibited small responses to perturbation. The more weakly-expressed *NR1* showed reduced transcript levels in MeJA-fed cells, while the more abundant *NR2* showed a small increase. Expression of both *NR*s also appeared to be modulated by PIP feeding. NRs mediate assimilation of inorganic nitrate (Beevers and Hageman 1969), and their differential expression patterns here are consistent with functional differentiation for the isoforms *in vivo*. *NR* expression is modulated by amino acid status, and NR activity is subject to posttranslational regulation based on carbon demand (Hey et al. 2010). Nitrate reductases also contribute to synthesis of NO (Lamotte et al. 2005), which is involved with jasmonate signaling and induced defense metabolism (Huang et al. 2004; Wünsche et al. 2011). *NR2* expression patterns were broadly consistent with a response to MeJA and/or amino acid levels, while *NR1* appeared to be coregulated with *NRT*s. Thus, while no previous reports have identified isoform-specific functions for NRs, our data suggest

the isoforms may play distinct roles in nitrate sensing and assimilation (NR1) and stress response and signaling (NR2).

Central Carbon Metabolism Shifts Due to Phenylpropanoid Perturbation

Although the methods employed here do not allow us to distinguish between changes in phenylpropanoid and amino acid metabolism as the proximal cause of changes in central carbon metabolism, we did identify treatment-related effects for this process. Carbohydrate profiles were relatively stable across elicitor and inhibitor treatments, with decreases in fructose and glucose observed in MeJA-fed cells (significant in Experiment 1). However, in both experiments genes associated with carbohydrate and oligosaccharide biosynthesis were overrepresented among downregulated genes, and glycolysis was overrepresented among upregulated ones. Previous work has indicated that newly-acquired carbon accounts for substantial portions of induced defense metabolism in photosynthetic tissues (Arnold and Schultz 2002; Babst et al. 2005; Hanik et al. 2010). The non-limiting sucrose levels in the culture medium probably generated an experimental system in which central carbon metabolism was limited mainly by cellular uptake, rather than by photosynthesis or long-distance transport as expected in whole plants. Metabolic inhibitors caused no clear shifts in sugar or sugar phosphate levels, although trehalose biosynthetic genes were downregulated in AOPP-fed cells. Trehalose-6-phosphate is thought to signal low carbon demand in plants (Hey et al. 2010). Trehalose feeding suppressed stress responsive transcripts, including *NR2*, in heterotrophic *Arabidopsis* cells (Bae et al. 2005a; Bae et al. 2005b), and trehalose accumulation reduced allocation of new carbon to phenylalanine and cinnamic acid pools

in tobacco (Best et al. 2011). Follow-up investigation will be needed to substantiate a link between trehalose levels, the observed transcriptional suppression of trehalose biosynthesis, carbon signaling, and phenylpropanoid pathway perturbation responses.

The dramatic increases in citric acid observed in response to MeJA feeding are consistent with increased glycolytic carbon flux to the citric acid cycle. Several citric acid cycle and oxidative phosphorylation genes were upregulated in MeJA-fed cells, which, along with the previously-mentioned GO enrichment of carbohydrate metabolism, further supports the same interpretation. This is not an entirely novel pattern (Babst et al. 2009; Goossens et al. 2003), but methods distinguishing among subcellular pools of citric acid cycle metabolites could better isolate the effects of MeJA on energetic carbon metabolism. Given the reported simultaneous but independent MeJA-mediated repression of cell cycle genes and activation of phenylpropanoid biosynthetic genes (Pauwels et al. 2008), such work would also help weigh the relative contributions of metabolism versus transcription to the regulation of tradeoffs between growth and induced defenses.

The strong MeJA-upregulation of multiple isoforms of putative *ATP citrate lyase* (*ACL*), known to act primarily in the cytosol (Fatland et al. 2002), suggested that citrate levels in MeJA-fed cells may have shifted in both cytosol and mitochondria. Acetyl-CoA resulting from ACL catalysis could be shunted towards biosynthesis of a variety of primary and secondary metabolites (Fatland et al. 2002; Fatland et al. 2005). For example, malonyl-CoA, which is produced via carboxylation of acetyl-CoA (Fatland et al. 2005), can be utilized in both flavonoid and fatty acid biosynthesis (Luo et al. 2007; Winkel-Shirley 2001). Stimulation of *ACL* is therefore in line with transcriptional

activation of the flavonoid/CT pathway during MeJA elicitation. While we did not quantify CoA conjugates here, recently developed methods (e.g., Qualley et al. 2012) should enable further assessment of this link to phenylpropanoid metabolism.

PIP effected a small, collective reduction of citric acid cycle metabolites, with modestly increased mitochondrial malate dehydrogenase and isocitrate dehydrogenase transcripts only in MeJA cofed cells. Along with small but significant upregulation in several citric acid cycle related genes in AOPP-fed cells, these data are consistent with crosstalk between phenylpropanoid metabolism and the citric acid cycle. Despite the increased transcript levels, α-ketoglutarate levels increased alongside decreased citrate, *cis*-aconitate, and fumarate in AOPP-fed cells. This may indicate distinct partitioning for these metabolites in different subcellular compartments or a change in flux mode for the citric acid cycle favoring alternative pathway routes for these metabolites (reviewed by Sweetlove et al. 2010). Increased levels of *ACL* transcripts as well as malonic acid in AOPP-fed cells also suggested increased acetyl-CoA/malonyl-CoA biosynthesis in the cytosol (Fatland et al. 2005). Coenzyme and lipid metabolism were significantly overrepresented among AOPP-upregulated genes. In conjuction with the observed reductions in flavonoids/CTs in these cells, it seems likely that acetyl-CoA and/or malonyl-CoA contribute primarily to fatty acid biosynthesis under conditions in which defense metabolism is not induced (Fatland et al. 2005).

Novel Insights for Phenolic and Benzoic Pathways

Several phenolic-related metabolites exhibited treatment-specific changes in abundance that may hint at their metabolic origins. The clearest example was a putative

arbutin that exhibited a fourfold increase in AOPP-fed cells along with a slight reduction in PIP-fed cells, similar to the patterns observed for phenylalanine and tyrosine. On this basis, we suggest that the putative arbutin is derived upstream of, and perhaps proximally to, phenylalanine. Loss of C3 organic acids from the C6 ring of chorismate, prephenate, or arogenate could all potentially produce hydroquinone, which is readily glycosylated to arbutin by plant cell cultures (Kittipongpatana et al. 2007; Lutterbach and Stöckigt 1994; Yan et al. 2007), but earlier steps in arbutin biosynthesis have not been well characterized.

Two peaks, putatively assigned as 3,4-dihydoxybenzoate and a phenyl-β-D-glucopyranoside, were decreased in AOPP-fed cells, suggesting a metabolic origin downstream of PAL. Neither compound was affected by PIP or MDCA feeding, leaving open the possibility that neither C4H nor 4CL contribute to their biosynthesis. Isotope feeding has suggested that cinnamate is a precursor for benzenoid and salicinoid biosynthesis in the Salicaceae (Babst et al. 2010; Zenk 1967). The absence of PIP-specific responses here suggests that these metabolites may have been quickly channeled into downstream pathways or may not be sensitive to cinnamate fluctuations. For example, metabolic flux models of *Petunia* metabolism suggest that benzenoids may originate from phenylpyruvate via either phenylalanine or chorismate, bypassing the core phenylpropanoid pathway (Boatright et al. 2004; Colón et al. 2010; Orlova et al. 2006). Along with 3,4-dihydroxybenzoate and the phenyl-β-D-glucopyranoside, phenylpyruvate was also reduced in AOPP-fed cells, consistent with previously-mentioned negative feedback regulation of the late shikimate pathway by phenylalanine (Maeda et al. 2010; Tzin and Galili 2010). While 3,4-dihydroxybenzoate levels increased under MeJA

feeding, the phenolic glucoside did not respond, consistent with reduced glycosylation in MeJA-treated cells. Additionally, 3,4-dihydroxybenzoate may be synthesized upstream of phenylalanine; in some fungi and bacteria this compound is a catabolite of quinate (Herrmann 1995; Herrmann and Weaver 1999). Whether this metabolic route operates in plants is unknown, but 3,4-dihydroxybenzoate and shikimate did respond similarly here to phenylpropanoid perturbation, while quinate pools were surprisingly resistant to perturbation.

Conclusion

The work here demonstrates the power of combining chemical perturbation with metabolite and transcript profiling for investigating phenylpropanoid pathway partitioning and its links to nitrogen and central carbon metabolism. As expected, MeJA elicitation led to increased core phenylpropanoid and flavonoid biosynthetic transcript levels and increased CTs, while metabolic inhibitors were associated with narrower and magnitudinally smaller shifts in gene expression and reductions in phenylpropanoid sinks. Lignin perturbation was weaker than that for CTs, and reductions in some lignin biosynthetic transcripts suggested preferential allocation of phenylpropanoid carbon towards flavonoid/CT biosynthesis in MeJA-fed cells. Flavonoid profiles and differential shifts in expression of flavonoid/CT biosynthetic and glycosyltransferase genes supported the notion that phenylpropanoid perturbation alters flavonoid partitioning between glycosylated and polymerized (i.e., CT) forms. Altered expression patterns for late flavonoid pathway genes suggested possible shifts in the monomeric composition of CTs

as well. Upstream of the phenylpropanoid pathway, differential changes in phenylpyruvate and tyrosine levels accompanied an increase in phenylalanine in cells fed the PAL inhibitor AOPP, supporting a metabolic link between tyrosine and phenylalanine biosynthesis and catabolism in *Populus*. AOPP also influenced levels of several other amino acids and transcripts related to nitrogen transport and assimilation. Within central carbon metabolism, MeJA upregulated glycolysis, oxidative phosphorylation, and citric acid cycle pathway transcripts, indicating a stimulation of energetic metabolism upon elicitation. AOPP, on the other hand, led to increases in several citric acid cycle transcripts but, as also seen for PIP, the corresponding metabolites were generally decreased. Given the heterotrophic nature of the cells, the data suggest the linkage of nitrogen and carbon metabolism via phenylalanine is physiologically relevant even under non-carbon-limited conditions. Finally, large increases in *ATP citrate lyase* transcripts were observed in cells fed MeJA and/or AOPP. This may support a model in which increased production of acetyl-CoA derived from cytosolic citrate leads to greater partitioning to flavonoid or fatty acid biosynthesis, depending on phenylpropanoid pathway status. In summary, perturbing the phenylpropanoid pathway in *Populus* cell cultures has consequences not only for carbon partitioning between downstream branch pathways, but also for nitrogen, central carbon, and likely fatty acid metabolism.

Tables

Table 3.1: Sequence information for additional QPCR primers used in this study.

Gene	Forward Primer	Reverse Primer
4CL2	TATTCCCAAATCGGCTTCTGG	CAGAATGATGGGTTTGTAGTAATT
NR1	CCTCCGCCGATGATTCAATTTGCT	AGGATTAACCAAGTAACAAACCATGC
NR2	AGTGTTGKGTTCRTTACCGAGAGT	TAGATCCGCTCCTCYSTCATCT
ARP	ACTGTGAGGAGATGCAGAAACGCA	GCTGTGTCACGGGCATTCAATGYT

Table 3.2: MANOVA p-values for overall effects on phenylpropanoid core pathway compounds.

Where p-values are not identical for all tests in a given effect, Pillai's Trace is shown. Star indicates statistical significance at the 0.05 level. "Total" indicates data for total hydroxycinnamates (B-DG) and phenylalanine (UB), as seen in Figures 3.7A & 3.8A.

Effect	Total Core Pathway		Hydroxycinnamate Free Acids		% Free Acids	
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
Inhibitor	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.9329	<0.0001*
Elicitor	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.2626	0.2280
I x E	0.0136*	0.0484*	0.2560	0.0001*	0.1484	0.3896

Table 3.3: ANOVA p-values for individual phenylpropanoid core pathway compounds, including phenylalanine and total (B-DG) hydroxycinnamates.

"Type" indicates which treatment was significantly different from the others according to to a *post hoc* Hsu's MCB test. Star indicates statistical significance at the 0.05 level. Gray text indicates source of degrees of freedom for slicing interactions and should be disregarded to avoid overanalysis; see **Methods** for additional background.

Effect	**Phenylalanine**		**Cinnamic Acid**		***p*-Coumaric Acid**		**Caffeic Acid**		**Ferulic Acid**	
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
Inhibitor	0.1513	<0.0001*	<0.0001*	<0.0001*	0.0005*	0.0012*	0.0391*	<0.0001*	0.1395	<0.0001*
Type		AOPP		MDCA		AOPP		AOPP		AOPP
Elicitor	<0.0001*	0.0014	0.0157	0.0925	0.2813	0.0005*	<0.0001*	<0.0001	0.0007*	0.0283*
I x E	0.0314*	0.0028*	0.0161*	0.0531	0.1156	0.5938	0.5805	0.0385*	0.9212	0.1873
Slices	I:None	E:None	E:None					E:None		
	0.0417*	<0.0001*	0.5877					0.0035*		
	I:MeJA	E:AOPP	E:PIP					E:AOPP		
	0.4720	0.0114*	0.0104*					0.3262		
		E:MDCA						E:MDCA		
		<0.0001*						0.0089*		

Table 3.4: ANOVA p-values for free hydroxycinnamates.

"Type" indicates which treatment was significantly different from the others according to a *post hoc* Hsu's MCB test. Star indicates statistical significance at the 0.05 level. Gray text indicates source of degrees of freedom for slicing interactions and should be disregarded to avoid overanalysis; see **Methods** for additional background. N/A indicates item not tested due to failure to reject the null hypothesis in MANOVA.

Effect	**Cinnamic Acid**		***p*-Coumaric Acid**		**Caffeic Acid**		**Ferulic Acid**	
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
Inhibitor	<0.0001*	<0.0001*	0.0309*	<0.0001*	0.0762	0.0006*	0.0376*	<0.0001*
Type		MDCA		AOPP		AOPP		AOPP
Elicitor	0.1949	<0.0001	0.0408*	<0.0001*	0.0017*	<0.0001*	0.0001*	0.0089
I x E	N/A	<0.0001*	N/A	0.2299	N/A	0.1476	N/A	0.0196*
Slices		E:None						E:None
		1.000						0.0020*
		E:AOPP						E:AOPP
		1.000						0.7592
		E:MDCA						E:MDCA
		<0.0001*						0.0702

Table 3.5: ANOVA p-values for percentage detected as the free-acid form for each hydroxycinnamate in Experiment 2.

Only the inhibitor effect in Experiment 2 was found to be statistically significant according to MANOVA, so other factors and Experiment 1 were not tested for individual metabolites. Star indicates statistical significance at the 0.05 level.

Effect	Cinnamic Acid	*p*-Coumaric Acid	Caffeic Acid	Ferulic Acid
Inhibitor	<0.0001*	0.0558	0.2306	0.0001*
Type	MDCA	Trend for AOPP		AOPP

Table 3.6: Results of two-way ANOVA tests for additional phenylpropanoid related compounds within each elicitor/inhibitor feeding experiment.

Starred items indicate statistical significance at the 0.05 level. Gray text indicates source for degrees of freedom for interaction slices; such overall effects should be disregarded.

Compound	Experiment 1		Experiment 2	
Catechol glucoside	Inhibitor	0.2293	Inhibitor	0.8323
	Elicitor	0.0003*	Elicitor	0.4291
	IxE	0.2385	IxE	0.6887
Phenyl-β-D-glucopyranoside	Inhibitor	0.4632	Inhibitor	0.0200*
	Elicitor	0.6009	Elicitor	0.1097
	IxE	0.1547	IxE	0.4295
Arbutin	Inhibitor	0.0003*	Inhibitor	<0.0001*
	Elicitor	0.0037	Elicitor	0.5937
	IxE	0.0287*	IxE	0.4804
	E:None	0.0148*		
	E:PIP	0.1749		
3,4-Dihydroxybenzoate	Inhibitor	0.6401	Inhibitor	0.0001*
	Elicitor	0.0040*	Elicitor	<0.0001*
	IxE	0.4956	IxE	0.1229
Shikimic Acid	Inhibitor	0.3344	Inhibitor	<0.0001*
	Elicitor	0.6316	Elicitor	0.0018*
	IxE	0.3784	IxE	0.1680
Quinic Acid	Inhibitor	0.2496	Inhibitor	0.4853
	Elicitor	0.2810	Elicitor	0.1333
	IxE	0.1001	IxE	0.4672

Table 3.7: MANOVA p-values for overall effects on flavonoids found in B-DG and in B fraction within each experiment.

Where p-values are not identical for a given effect, Pillai's Trace is shown. Star indicates statistical significance at the 0.05 level.

Effect	Flavonoids B-DG		Flavonoids B	
	Exp 1	Exp 2	Exp 1	Exp 2
Inhibitor	0.2846	0.0219*	0.6564	0.0496*
Elicitor	0.0009*	<0.0001*	0.1665	0.0006*
I x E	0.0034*	0.2621	0.9910	0.1174

Table 3.8: Flavonoid levels (Mean+SD, relative to *o*-anisic acid standard) in *Populus* cell cultures from Experiment 1.

N=6 or 7. Numbers beside the named metabolites indicate retention index (RI). ^{E}Statistically significant response to elicitor feeding overall. ^{tr}Trend towards response to elicitor feeding overall.

Flavonoid	Control	PIP	MeJA	PIP+MeJA
Flavonoid B-DG (Total Aglycones & β-Glucosides)				
Catechin 2882^{E}	$2.04\pm0.47 \times 10^{1}$	$1.60\pm0.36 \times 10^{1}$	$0.75\pm0.18 \times 10^{1}$	$0.78\pm0.20 \times 10^{1}$
Taxifolin 2898^{E}	$5.82\pm1.99 \times 10^{-1}$	$5.26\pm1.33 \times 10^{-1}$	$3.67\pm0.42 \times 10^{-1}$	$3.21\pm0.99 \times 10^{-1}$
Taxifolin 2958^{E}	$1.50\pm0.35 \times 10^{1}$	$1.24\pm0.25 \times 10^{1}$	$0.95\pm0.14 \times 10^{1}$	$0.98\pm0.21 \times 10^{1}$
Eriodictyol 2974^{E}	$3.65\pm2.27 \times 10^{-1}$	$2.82\pm1.13 \times 10^{-1}$	$1.09\pm0.43 \times 10^{-1}$	$1.40\pm0.38 \times 10^{-1}$
Kaempferol 3166	$1.69\pm1.92 \times 10^{-1}$	$0.70\pm0.22 \times 10^{-1}$	$0.51\pm0.12 \times 10^{-1}$	$0.67\pm0.31 \times 10^{-1}$
Flavonoid B (Aglycones)				
Taxifolin 1843	$0.92\pm0.24 \times 10^{-1}$	$1.20\pm0.24 \times 10^{-1}$	$1.20\pm0.15 \times 10^{-1}$	$1.12\pm0.21 \times 10^{-1}$
Taxifolin 1886	$1.84\pm0.23 \times 10^{-1}$	$2.11\pm0.41 \times 10^{-1}$	$2.13\pm0.34 \times 10^{-1}$	$2.22\pm0.89 \times 10^{-1}$
Catechin 2852^{tr}	$1.49\pm0.51 \times 10^{-1}$	$1.48\pm0.76 \times 10^{-1}$	$2.29\pm0.37 \times 10^{-1}$	$2.55\pm1.15 \times 10^{-1}$
Kaempferol 2858	$1.36\pm0.56 \times 10^{-1}$	$1.55\pm1.26 \times 10^{-1}$	$1.65\pm0.80 \times 10^{-1}$	$2.18\pm1.75 \times 10^{-1}$
Catechin 2877	$9.68\pm1.97 \times 10^{1}$	$9.19\pm2.61 \times 10^{1}$	$10.8\pm3.32 \times 10^{1}$	$13.3\pm4.24 \times 10^{1}$
Catechin 2887	$3.06\pm1.52 \times 10^{-1}$	$3.77\pm1.23 \times 10^{-1}$	$4.16\pm1.51 \times 10^{-1}$	$3.95\pm2.49 \times 10^{-1}$
Catechin 2911^{tr}	nd	nd	$4.77\pm0.74 \times 10^{-1}$	$4.70\pm0.21 \times 10^{-1}$
Catechin 2929	$2.91\pm0.40 \times 10^{-1}$	$3.44\pm1.02 \times 10^{-1}$	$4.35\pm1.61 \times 10^{-1}$	$4.40\pm2.69 \times 10^{-1}$
Taxifolin 2957^{tr}	$2.33\pm0.59 \times 10^{-1}$	$2.21\pm0.74 \times 10^{-1}$	$4.77\pm1.09 \times 10^{-1}$	$4.13\pm2.12 \times 10^{-1}$

Table 3.9: Flavonoid levels (Mean+SD, relative to *o*-anisic acid standard) in *Populus* cell cultures from Experiment 2.

N=6. Numbers beside metabolite names indicate approximate RI. An entry of nd indicates compound was not found in the sample set. [a]Only one sample in the set contained the compound; SD could not be calculated. [A]Statistically significant response to AOPP feeding overall; [M]significant response to MDCA feeding overall; [L]significant difference between AOPP and MDCA-fed cells overall, but not from either fed set to unfed cells; [E]significant response to elicitor feeding overall; [X]excluded from MANOVA analysis due to low detection.

Flavonoid	Control	AOPP	MDCA	MeJA	AOPP+MeJA	MDCA+MeJA
Flavonoid B-DG (Total Aglycones & β-Glucosides)						
Catechin 2878^{A,E}	$10.8\pm1.95 \times 10^{0}$	$4.89\pm1.76 \times 10^{0}$	$8.96\pm3.51 \times 10^{0}$	$3.34\pm2.26 \times 10^{0}$	$2.33\pm0.74 \times 10^{0}$	$5.24\pm3.38 \times 10^{0}$
Taxifolin 2897^{A}	$2.62\pm0.44 \times 10^{-1}$	$1.29\pm0.31 \times 10^{-1}$	$2.48\pm0.88 \times 10^{-1}$	$1.79\pm0.62 \times 10^{-1}$	$1.64\pm0.31 \times 10^{-1}$	$2.35\pm0.85 \times 10^{-1}$
Taxifolin 2954^{A}	$7.35\pm1.69 \times 10^{0}$	$3.92\pm0.95 \times 10^{0}$	$7.53\pm2.29 \times 10^{0}$	$5.00\pm1.93 \times 10^{0}$	$4.74\pm1.62 \times 10^{0}$	$7.00\pm2.79 \times 10^{0}$
Flavonoid B (Aglycones)						
Taxifolin 1836^{A}	$2.45\pm0.55 \times 10^{-1}$	$2.05\pm0.41 \times 10^{-1}$	$2.84\pm0.32 \times 10^{-1}$	$2.91\pm0.80 \times 10^{-1}$	$2.20\pm0.36 \times 10^{-1}$	$2.66\pm0.18 \times 10^{-1}$
Taxifolin 1881^{A}	$2.85\pm0.13 \times 10^{-1}$	$2.63\pm0.38 \times 10^{-1}$	$3.21\pm0.36 \times 10^{-1}$	$3.31\pm0.95 \times 10^{-1}$	$2.55\pm0.29 \times 10^{-1}$	$3.40\pm0.64 \times 10^{-1}$
Catechin 2853^{A,E}	$4.33\pm1.79 \times 10^{-2}$	$7.06\pm2.01 \times 10^{-2}$	$7.70\pm2.47 \times 10^{-2}$	$6.78\pm1.29 \times 10^{-2}$	$17.6\pm6.94 \times 10^{-2}$	$8.84\pm0.74 \times 10^{-2}$
Kaempferol 2859^{M,E}	nd	nd	7.89×10^{-2a}	$8.99\pm1.72 \times 10^{-2}$	$3.81\pm0.30 \times 10^{-2}$	$14.2\pm6.17 \times 10^{-2}$
Catechin 2879^{L,E}	$4.84\pm1.28 \times 10^{1}$	$3.62\pm1.34 \times 10^{1}$	$5.43\pm0.85 \times 10^{1}$	$7.73\pm1.08 \times 10^{1}$	$6.19\pm1.60 \times 10^{1}$	$8.64\pm1.67 \times 10^{1}$
Catechin 2888^{A,M,E}	$1.44\pm0.56 \times 10^{-1}$	$0.87\pm0.29 \times 10^{-1}$	$2.37\pm1.14 \times 10^{-1}$	$2.70\pm0.46 \times 10^{-1}$	$1.13\pm0.12 \times 10^{-1}$	$3.33\pm1.15 \times 10^{-1}$
Catechin 2853^{L,E}	nd	nd	$0.63\pm0.25 \times 10^{-1}$	$2.33\pm0.46 \times 10^{-1}$	$1.61\pm0.40 \times 10^{-1}$	$2.71\pm0.71 \times 10^{-1}$
Catechin 2930^{X}	nd	nd	nd	$2.51\pm0.50 \times 10^{-2}$	nd	3.78×10^{-2a}
Taxifolin 2959^{X,A,E}	3.41×10^{-2a}	nd	nd	$11.1\pm5.01 \times 10^{-2}$	nd	$8.38\pm3.62 \times 10^{-2}$

Table 3.10: Amino acid levels (Mean+SD, relative to adonitol standard) in *Populus* cell UB extract fraction from Experiment 1.

N=6 or 7. Starred items showed a statistically significant response to MeJA feeding overall.

Amino Acid	Control	PIP	MeJA	PIP+MeJA
β-Alanine	$2.73 \pm 1.30 \times 10^{-3}$	$2.64 \pm 1.43 \times 10^{-3}$	$4.13 \pm 2.11 \times 10^{-3}$	$2.39 \pm 1.18 \times 10^{-3}$
Aspartic Acid*	$3.66 \pm 1.43 \times 10^{-2}$	$2.71 \pm 1.24 \times 10^{-2}$	$1.54 \pm 0.27 \times 10^{-2}$	$2.36 \pm 0.94 \times 10^{-2}$
Glutamic Acid*	$2.41 \pm 1.03 \times 10^{-2}$	$2.32 \pm 0.55 \times 10^{-2}$	$1.56 \pm 0.34 \times 10^{-2}$	$1.61 \pm 0.44 \times 10^{-2}$
Glycine*	$9.47 \pm 3.29 \times 10^{-3}$	$4.59 \pm 2.42 \times 10^{-3}$	$2.43 \pm 1.00 \times 10^{-3}$	$5.64 \pm 2.67 \times 10^{-3}$
Homoserine*	$5.02 \pm 1.31 \times 10^{-2}$	$4.24 \pm 1.17 \times 10^{-2}$	$3.60 \pm 0.65 \times 10^{-2}$	$3.58 \pm 1.03 \times 10^{-2}$
Isoleucine	$1.640.78 \times 10^{-2}$	$0.83 \pm 0.31 \times 10^{-2}$	$0.46 \pm 0.37 \times 10^{-2}$	$1.32 \pm 0.28 \times 10^{-2}$
Serine*	$2.55 \pm 1.34 \times 10^{-2}$	$1.64 \pm 0.56 \times 10^{-2}$	$0.80 \pm 0.56 \times 10^{-2}$	$1.72 \pm 0.55 \times 10^{-2}$
Threonine*	$4.11 \pm 1.46 \times 10^{-2}$	$2.48 \pm 1.00 \times 10^{-2}$	$3.93 \pm 1.15 \times 10^{-2}$	$6.63 \pm 3.21 \times 10^{-2}$
Valine	$5.85 \pm 1.71 \times 10^{-4}$	$4.65 \pm 1.44 \times 10^{-4}$	$4.40 \pm 1.50 \times 10^{-4}$	$3.22 \pm 1.55 \times 10^{-4}$

Table 3.11: Amino acid levels (Mean+SD) in *Populus* cell UB extract fraction from Experiment 2.

N=6 for all treatments. An entry of nd indicates compound was not found in the sample set. [a]Only one sample in the set contained the compound; SD could not be calculated. Starred compounds showed a statistically significant response to inhibitor treatment overall; *post hoc* Hsu's MCB testing suggests a response to AOPP in all cases.

Amino Acid	Control	AOPP	MDCA	MeJA	AOPP+ MeJA	MDCA+MeJA
β-Alanine*	$1.08\pm0.34 \times 10^{-3}$	$18.3\pm6.03 \times 10^{-3}$	$0.91\pm0.21 \times 10^{-3}$	$0.96\pm0.64 \times 10^{-3}$	$28.7\pm10.4 \times 10^{-3}$	$0.54\pm0.04 \times 10^{-3}$
Aspartic Acid	$4.55\pm0.74 \times 10^{-2}$	$3.90\pm1.22 \times 10^{-2}$	$4.47\pm0.48 \times 10^{-2}$	$3.89\pm1.28 \times 10^{-2}$	$3.06\pm1.30 \times 10^{-2}$	$3.67\pm0.53 \times 10^{-2}$
Glutamic Acid	$2.49\pm0.63 \times 10^{-2}$	$2.18\pm0.86 \times 10^{-2}$	$2.56\pm0.33 \times 10^{-2}$	$2.13\pm0.49 \times 10^{-2}$	$1.69\pm0.52 \times 10^{-2}$	$1.88\pm0.42 \times 10^{-2}$
Glycine	$5.52\pm3.15 \times 10^{-3}$	$5.25\pm2.16 \times 10^{-3}$	$5.42\pm3.22 \times 10^{-3}$	$3.84\pm3.45 \times 10^{-3}$	$5.70\pm4.13 \times 10^{-3}$	$3.45\pm2.36 \times 10^{-3}$
Homoserine	$4.01\pm0.80 \times 10^{-2}$	$3.64\pm1.16 \times 10^{-2}$	$4.23\pm0.71 \times 10^{-2}$	$4.28\pm1.59 \times 10^{-2}$	$4.09\pm1.22 \times 10^{-2}$	$3.52\pm0.32 \times 10^{-2}$
5-Hydroxynorvaline*	$0.80\pm0.24 \times 10^{-3}$	$5.77\pm3.30 \times 10^{-3}$	$0.46\pm0.05 \times 10^{-3}$	nd	$6.86\pm2.27 \times 10^{-3}$	$0.37^{a} \times 10^{-3}$
Isoleucine*	$1.01\pm0.49 \times 10^{-2}$	$3.38\pm1.18 \times 10^{-2}$	$1.26\pm0.27 \times 10^{-2}$	$0.95\pm0.31 \times 10^{-2}$	$2.47\pm1.03 \times 10^{-2}$	$0.87\pm0.39 \times 10^{-2}$
Lysine*	$0.55\pm0.21 \times 10^{-2}$	$1.95\pm0.59 \times 10^{-2}$	$0.60\pm0.15 \times 10^{-2}$	$0.38\pm0.12 \times 10^{-2}$	$1.45\pm0.44 \times 10^{-2}$	$0.48\pm0.15 \times 10^{-2}$
Serine	$2.19\pm0.47 \times 10^{-2}$	$1.55\pm0.61 \times 10^{-2}$	$2.73\pm0.76 \times 10^{-2}$	$1.60\pm0.80 \times 10^{-2}$	$1.95\pm0.87 \times 10^{-2}$	$1.27\pm0.32 \times 10^{-2}$
Threonine*	$3.28\pm0.79 \times 10^{-2}$	$2.33\pm1.17 \times 10^{-2}$	$3.58\pm0.75 \times 10^{-2}$	$6.20\pm2.74 \times 10^{-2}$	$3.29\pm1.54 \times 10^{-2}$	$5.50\pm1.11 \times 10^{-2}$
Tyrosine*	$0.50\pm0.13 \times 10^{-2}$	$15.9\pm4.12 \times 10^{-2}$	$0.37\pm0.15 \times 10^{-2}$	$0.40\pm0.11 \times 10^{-2}$	$17.3\pm6.20 \times 10^{-2}$	$0.33\pm0.09 \times 10^{-2}$
Valine*	$0.76\pm0.36 \times 10^{-3}$	$0.52\pm0.22 \times 10^{-3}$	$0.75\pm0.46 \times 10^{-3}$	$1.12\pm0.51 \times 10^{-3}$	$0.49\pm0.17 \times 10^{-3}$	$0.61\pm0.25 \times 10^{-3}$

Table 3.12: MANOVA p-values for overall effects on organic acids and sugars within each experiment.

Where p-values are not identical for a given effect, Pillai's Trace is shown. Star indicates statistical significance at the 0.05 level. "Sugars" includes both sugars and sugar phosphates.

Effect	Citric Acid Cycle		Other Organic Acids		Sugars	
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
Inhibitor	0.0168*	<0.0001*	0.0049*	<0.0001*	0.6018	0.9476
Elicitor	<0.0001*	<0.0001*	0.0006*	0.0067*	0.0389*	0.0003*
I x E	0.2025	<0.0011*	0.1789	0.0824	0.3672	0.2835

Table 3.13: Organic acid and sugar levels (Mean+SD) in *Populus* cell extracts from Experiment 1.

All metabolites are quantified from the UB fraction unless otherwise noted; N=6 or 7. Gray rows were excluded from MANOVA to avoid duplication. [E]Statistically significant response to elicitor feeding overall; [P]significant response to inhibitor feeding overall.

Peak ID	Control	PIP	MeJA	PIP+MeJA
Citric Acid Cycle				
Succinic[E]	$4.21\pm1.47 \times 10^{-2}$	$3.57\pm1.24 \times 10^{-2}$	$2.03\pm0.41 \times 10^{-2}$	$1.98\pm0.66 \times 10^{-2}$
Fumaric[P,E]	$8.63\pm3.30 \times 10^{-3}$	$7.10\pm1.62 \times 10^{-3}$	$8.16\pm1.48 \times 10^{-3}$	$4.36\pm0.77 \times 10^{-3}$
Malic	$3.89\pm1.99 \times 10^{0}$	$4.07\pm0.93 \times 10^{0}$	$4.24\pm0.61 \times 10^{0}$	$4.16\pm1.33 \times 10^{0}$
α-Ketoglutaric[E]	$5.67\pm1.62 \times 10^{-2}$	$5.12\pm1.36 \times 10^{-2}$	$3.17\pm0.39 \times 10^{-2}$	$3.49\pm1.00 \times 10^{-2}$
***cis*-Aconitic**	$3.86\pm0.88 \times 10^{-3}$	$4.01\pm1.51 \times 10^{-3}$	$4.22\pm0.49 \times 10^{-3}$	$4.54\pm1.49 \times 10^{-3}$
Citric[E]	$9.56\pm1.97 \times 10^{-1}$	$9.20\pm3.37 \times 10^{-1}$	$19.9\pm4.53 \times 10^{-1}$	$23.0\pm5.79 \times 10^{-1}$
Other Organic Acids				
Ascorbic Acid[E]	$1.42\pm0.78 \times 10^{-2}$	$1.23\pm0.57 \times 10^{-2}$	$0.93\pm0.37 \times 10^{-2}$	$0.78\pm0.38 \times 10^{-2}$
Dehydroascorbic a[E]	$2.17\pm0.42 \times 10^{-1}$	$1.87\pm0.37 \times 10^{-1}$	$1.54\pm0.16 \times 10^{-1}$	$1.64\pm0.46 \times 10^{-1}$
Glyceric[E]	$2.54\pm0.97 \times 10^{-2}$	$2.30\pm0.59 \times 10^{-2}$	$1.36\pm0.24 \times 10^{-2}$	$1.32\pm0.29 \times 10^{-2}$
Lactic	$9.66\pm4.54 \times 10^{-3}$	$6.64\pm1.30 \times 10^{-3}$	$6.20\pm3.72 \times 10^{-3}$	$7.87\pm3.01 \times 10^{-3}$
Malonic	$7.89\pm3.12 \times 10^{-3}$	$5.19\pm1.97 \times 10^{-3}$	$5.38\pm1.63 \times 10^{-3}$	$5.03\pm2.01 \times 10^{-3}$
Oxalic	$1.37\pm0.60 \times 10^{-1}$	$0.92\pm0.11 \times 10^{-1}$	$0.81\pm0.36 \times 10^{-1}$	$1.10\pm0.31 \times 10^{-1}$
3-Phenyllactic[E]	$8.65\pm2.07 \times 10^{-2}$	$9.06\pm2.52 \times 10^{-2}$	$5.33\pm0.77 \times 10^{-2}$	$5.84\pm1.54 \times 10^{-2}$
3-Phenyllactic (B)[E]	$7.28\pm1.03 \times 10^{-2}$	$7.88\pm1.89 \times 10^{-2}$	$5.11\pm0.55 \times 10^{-2}$	$5.63\pm0.86 \times 10^{-2}$
3-Phenyllactic (B-DG)[P,E]	$1.30\pm0.24 \times 10^{-1}$	$2.03\pm0.36 \times 10^{-1}$	$0.65\pm0.04 \times 10^{-1}$	$0.73\pm0.20 \times 10^{-1}$
Phenylpyruvic	$6.16\pm1.54 \times 10^{-3}$	$6.29\pm2.70 \times 10^{-3}$	$7.07\pm2.12 \times 10^{-3}$	$8.01\pm2.92 \times 10^{-3}$
3-Phosphoglyceric	$7.07\pm2.87 \times 10^{-3}$	$6.86\pm2.72 \times 10^{-3}$	$6.24\pm1.60 \times 10^{-3}$	$6.45\pm3.00 \times 10^{-3}$
Sugars & Sugar Phosphates				
Fructose a[E]	$8.93\pm2.56 \times 10^{0}$	$7.19\pm1.90 \times 10^{0}$	$5.47\pm0.58 \times 10^{0}$	$5.62\pm1.45 \times 10^{0}$
Fructose b[E]	$6.04\pm2.84 \times 10^{0}$	$5.49\pm1.65 \times 10^{0}$	$4.14\pm0.44 \times 10^{0}$	$3.92\pm0.97 \times 10^{0}$
Glucose a[E]	$7.25\pm1.39 \times 10^{0}$	$5.87\pm1.21 \times 10^{0}$	$4.40\pm0.40 \times 10^{0}$	$4.56\pm1.07 \times 10^{0}$
Glucose b[E]	$2.01\pm0.56 \times 10^{0}$	$1.60\pm0.31 \times 10^{0}$	$0.89\pm0.10 \times 10^{0}$	$0.86\pm0.19 \times 10^{0}$
Glucose-6-phosphate a	$1.25\pm0.34 \times 10^{-1}$	$0.92\pm0.38 \times 10^{-1}$	$1.07\pm0.65 \times 10^{-1}$	$1.38\pm0.48 \times 10^{-1}$
Glucose-6-phosphate b	$4.33\pm1.04 \times 10^{-1}$	$3.64\pm0.83 \times 10^{-1}$	$3.79\pm0.67 \times 10^{-1}$	$3.91\pm1.02 \times 10^{-1}$
Ribose phosphate	$1.91\pm0.52 \times 10^{-2}$	$1.51\pm0.31 \times 10^{-2}$	$1.44\pm0.46 \times 10^{-2}$	$1.93\pm0.62 \times 10^{-2}$
Ribose-5-phosphate	$3.41\pm6.46 \times 10^{-2}$	$6.54\pm5.84 \times 10^{-2}$	$3.30\pm4.33 \times 10^{-2}$	$0.20\pm0.07 \times 10^{-2}$
Sucrose	$7.68\pm3.69 \times 10^{0}$	$7.61\pm1.23 \times 10^{0}$	$6.63\pm1.29 \times 10^{0}$	$6.88\pm2.40 \times 10^{0}$

Table 3.14: Organic acid and sugar levels (Mean+SD) in *Populus* cell extracts from Experiment 2.

All metabolites are quantified from the UB fraction unless otherwise noted; N=6. Gray rows were excluded from MANOVA to avoid duplication. [A]Statistically significant response to AOPP feeding overall; [M]significant response to MDCA feeding overall; [L]significant difference between AOPP and MDCA-fed cells overall, but not from either fed set to unfed cells; [E]significant response to elicitor feeding overall; [I]significant interaction between elicitor and inhibitor feeding.

Peak ID	Control	AOPP	MDCA	MeJA	AOPP+ MeJA	MDCA+MeJA
Citric Acid Cycle						
Succinic[E]	4.59±0.76 x10^{-2}	4.29±1.68 x10^{-2}	5.13±0.88 x10^{-2}	3.69±1.66 x10^{-2}	3.65±0.90 x10^{-2}	3.34±0.15 x10^{-2}
Fumaric[A,M,I]	6.49±2.82 x10^{-3}	5.64±2.56 x10^{-3}	3.44±0.67 x10^{-3}	14.1±3.88 x10^{-3}	3.59±1.06 x10^{-3}	2.07±0.80 x10^{-3}
Malic	4.67±1.03 x10^{0}	3.52±2.16 x10^{0}	3.91±0.70 x10^{0}	5.12±1.78 x10^{0}	4.52±0.66 x10^{0}	4.47±0.67 x10^{0}
α-Ketoglutaric[A]	6.38±1.07 x10^{-2}	7.81±3.17 x10^{-2}	7.80±1.14 x10^{-2}	6.01±2.19 x10^{-2}	8.19±2.10 x10^{-2}	5.51±0.54 x10^{-2}
***cis*-Aconitic[A,M]**	4.69±1.77 x10^{-3}	4.39±1.29 x10^{-3}	3.44±1.52 x10^{-3}	7.93±4.06 x10^{-3}	4.12±1.48 x10^{-3}	4.83±2.42 x10^{-3}
Citric[E,L]	0.78±0.32 x10^{0}	0.50±0.20 x10^{0}	0.90±0.49 x10^{0}	1.68±0.44 x10^{0}	1.41±0.60 x10^{0}	1.92±0.60 x10^{0}
Other Organic Acids						
Ascorbic Acid	1.25±0.68 x10^{-2}	1.02±0.65 x10^{-2}	1.42±0.71 x10^{-2}	2.29±0.17 x10^{-2}	1.45±0.78 x10^{-2}	1.62±0.67 x10^{-2}
Dehydroascorbic a	2.39±0.41 x10^{-1}	2.25±0.53 x10^{-1}	2.50±0.50 x10^{-1}	2.27±0.55 x10^{-1}	2.10±0.47 x10^{-1}	2.02±0.30 x10^{-1}
Dehydroascorbic b[A,E]	nd	nd	nd	1.19±0.28 x10^{-2}	nd	1.32±0.14 x10^{-2}
Glyceric	3.04±0.53 x10^{-2}	3.00±1.08 x10^{-2}	3.95±0.82 x10^{-2}	2.98±1.31 x10^{-2}	2.95±0.66 x10^{-2}	2.81±0.41 x10^{-2}
Lactic	3.59±1.17x10^{-2}	4.17±2.81 x10^{-2}	3.31±1.07 x10^{-2}	3.88±1.55 x10^{-2}	3.09±0.93 x10^{-2}	3.14±0.58 x10^{-2}
Malonic[A]	2.89±2.18 x10^{-3}	8.69±5.27 x10^{-3}	2.46±1.52 x10^{-3}	1.93±2.03 x10^{-3}	6.54±1.98 x10^{-3}	1.05±0.13 x10^{-3}
Oxalic	8.35±2.44 x10^{-2}	9.33±6.47 x10^{-2}	5.65±1.11 x10^{-2}	7.99±3.73 x10^{-2}	6.72±1.92 x10^{-2}	8.77±1.52 x10^{-2}
3-Phenyllactic[A,E]	0.75±0.23 x10^{-1}	1.31±0.43 x10^{-1}	1.23±0.32 x10^{-1}	0.70±0.31 x10^{-1}	1.13±0.30 x10^{-1}	0.63±0.60 x10^{-1}
3-Phenyllactic (B)[A,E]	0.40±0.07 x10^{-1}	5.93±1.21 x10^{-1}	0.69±0.23 x10^{-1}	0.24±0.02 x10^{-1}	3.75±0.94 x10^{-1}	0.38+0.09 x10^{-1}
3-Phenyllactic (B-DG) [A,E]	0.68±0.08 x10^{-1}	9.26±1.14 x10^{-1}	1.50±0.32 x10^{-1}	0.39±0.08 x10^{-1}	5.28±0.82 x10^{-1}	0.56±0.08 x10^{-1}

Table 3.14, Continued:

Peak ID	Control	AOPP	MDCA	MeJA	AOPP+ MeJA	MDCA+MeJA
			Other Organic Acids			
Phenylpyruvic[A,M]	7.41±1.59 x10^{-3}	5.68±2.04 x10^{-3}	11.1±2.06 x10^{-3}	7.87±3.13 x10^{-3}	5.58±2.17 x10^{-3}	9.15±3.09 x10^{-3}
3-Phosphoglyceric	0.94±0.35 x10^{-2}	1.35±0.48 x10^{-2}	1.16±0.29 x10^{-2}	1.16±0.62 x10^{-2}	1.35±0.55 x10^{-2}	1.07±0.32 x10^{-2}
			Sugars & Sugar Phosphates			
Fructose a	8.17±3.31 x10^{0}	7.93±3.45 x10^{0}	7.55±0.61 x10^{0}	6.58±2.98 x10^{0}	8.44±1.49 x10^{0}	5.68±0.82 x10^{0}
Fructose b	6.83±2.22 x10^{0}	7.42±2.27 x10^{0}	6.53±1.47 x10^{0}	6.62±3.09 x10^{0}	6.46±0.88 x10^{0}	4.80±0.72 x10^{0}
Glucose a	6.61±2.98 x10^{0}	7.18±2.50 x10^{0}	6.64±1.33 x10^{0}	6.50±2.33 x10^{0}	5.99±0.83 x10^{0}	4.75±0.81 x10^{0}
Glucose b[E]	2.16±0.49 x10^{0}	2.30±0.81 x10^{0}	2.30±0.64 x10^{0}	1.40±0.75 x10^{0}	1.78±0.41 x10^{0}	0.93±0.15 x10^{0}
Glucose-6-phosphate a	1.15±0.46 x10^{-1}	1.07±0.64 x10^{-1}	1.14±0.43 x10^{-1}	1.08±0.88 x10^{-1}	1.28±0.46 x10^{-1}	0.94±0.27 x10^{-1}
Glucose-6-phosphate b	2.67±0.51 x10^{-1}	3.11±1.36 x10^{-1}	3.15±0.91 x10^{-1}	3.31±1.76 x10^{-1}	3.40±1.07 x10^{-1}	2.65±0.44 x10^{-1}
Sucrose	5.00±3.12 x10^{0}	8.08±2.84 x10^{0}	7.04±1.47 x10^{0}	7.79±3.06 x10^{0}	7.61±2.51 x10^{0}	5.59±1.59 x10^{0}

Table 3.15: Top ten overrepresented GO terms in the biological process domain for DE genes identified in Experiment 1, organized by effect type, then by up- (red text) or down-regulation (blue).

Identification of overrepresented terms was conducted on the AgriGO website (Du et al. 2010), and the total list was then reduced using REViGO (Supek et al. 2011). No significant GO category enrichment was found for DE genes responding to the PIP overall effect.

GO Code	Category	Adjusted *p*	Fold Overrepresented
Responding to MeJA			
GO:0009073	aromatic amino acid family biosynthetic process	0.0004	10.009
GO:0006414	translational elongation	0.0047	5.561
GO:0006790	sulfur metabolic process	0.0018	5.005
GO:0051258	protein polymerization	0.0027	4.671
GO:0006119	oxidative phosphorylation	0.0024	4.214
GO:0034220	ion transmembrane transport	0.0036	4.204
GO:0006096	glycolysis	0.0170	3.951
GO:0006818	hydrogen transport	0.0029	3.938
GO:0009308	amine metabolic process	1.0×10^{-9}	3.824
GO:0044271	cellular nitrogen compound biosynthetic process	1.7×10^{-6}	3.630
	Plus 24 additional GO categories	*<0.05*	*3.533 – 1.423*
GO:0009311	oligosaccharide metabolic process	0.0300	4.077
GO:0032940	secretion by cell	0.0480	3.443
GO:0046903	secretion	0.0480	3.443
GO:0016051	carbohydrate biosynthetic process	0.0480	2.530
GO:0065008	regulation of biological quality	0.0340	2.200
GO:0015031	protein transport	0.0140	2.184
GO:0033036	macromolecule localization	0.0099	2.179
GO:0051641	cellular localization	0.0100	2.170
GO:0006519	cellular amino acid and derivative metabolic process	0.0420	1.945

Table 3.15, Continued:

GO Code	Category	Adjusted *p*	Fold Overrepresented
	Responding to MeJA		
GO:0016310	phosphorylation	0.0003	1.621
	Plus 17 additional GO categories	*<0.05*	*1.607 – 1.283*
	Responding to PxM		
GO:0008652	cellular amino acid biosynthetic process	0.0015	7.025
GO:0044271	cellular nitrogen compound biosynthetic process	0.0015	5.095
GO:0009308	amine metabolic process	0.0003	4.776
GO:0006082	organic acid metabolic process	0.0002	4.580
GO:0042180	cellular ketone metabolic process	0.0002	4.567
GO:0034641	cellular nitrogen compound metabolic process	0.0003	4.307
GO:0006412	translation	0.0017	3.062
GO:0044281	small molecule metabolic process	0.0003	3.055
GO:0009058	biosynthetic process	0.0098	1.636
GO:0044249	cellular biosynthetic process	0.0150	1.618
	Plus 1 additional GO category	*<0.05*	*1.362*
N/A	*none*		

Table 3.16: Top ten overrepresented GO terms in the molecular function domain for DE genes identified in Experiment 1, organized by effect type, then by up- (red text) or down-regulation (blue).

Identification of overrepresented terms was conducted on the AgriGO website, and the total list was then reduced using REViGO. No significant GO category enrichment was found for DE genes responding to the PIP overall effect.

GO Code	Category	Adjusted *p*	Fold Overrepresented
Responding to MeJA			
GO:0004298	threonine-type endopeptidase activity	8.6×10^{-5}	8.007
GO:0070003	threonine-type peptidase activity	8.6×10^{-5}	8.007
GO:0019205	nucleobase, nucleoside, nucleotide kinase activity	0.0035	6.825
GO:0003735	structural constituent of ribosome	9.4×10^{-24}	5.057
GO:0010181	FMN binding	0.0140	4.692
GO:0015078	hydrogen ion transmembrane transporter activity	0.0008	4.290
GO:0016835	carbon-oxygen lyase activity	0.0140	4.004
GO:0003924	GTPase activity	0.0015	3.368
GO:0008080	N-acetyltransferase activity	0.0170	3.284
GO:0016597	amino acid binding	0.0250	3.276
	Plus 27 additional GO categories	*<0.05*	*3.276 – 1.325*
GO:0008289	lipid binding	0.0043	4.064
GO:0008234	cysteine-type peptidase activity	0.0150	3.220
GO:0004386	helicase activity	0.0039	2.556
GO:0019842	vitamin binding	0.0410	2.446
GO:0016874	ligase activity	0.0080	2.086
GO:0016879	ligase activity, forming carbon-nitrogen bonds	0.0390	1.961
GO:0050662	coenzyme binding	0.0280	1.794
GO:0048037	cofactor binding	0.0130	1.769
GO:0008233	peptidase activity	0.0280	1.746
GO:0016773	phosphotransferase activity, alcohol group as acceptor	3.3×10^{-6}	1.742
	Plus 9 additional GO categories	*<0.05*	*1.709 – 1.325*

Table 3.16, Continued:

GO Code	Category	Adjusted *p*	Fold Overrepresented
	Responding to PxM		
GO:0008026	ATP-dependent helicase activity	0.0110	5.627
GO:0003723	RNA binding	0.0004	4.471
GO:0022890	inorganic cation transmembrane transporter activity	0.0210	4.424
GO:0008415	acyltransferase activity	0.0330	3.864
GO:0003735	structural constituent of ribosome	0.0039	3.356
GO:0005525	GTP binding	0.0170	3.243
GO:0019001	guanyl nucleotide binding	0.0170	3.165
GO:0016817	hydrolase activity, acting on acid anhydrides	0.0030	2.631
GO:0016787	hydrolase activity	0.0170	1.643
GO:0003676	nucleic acid binding	0.0170	1.599
	Plus 1 additional GO category	*<0.05*	*1.522*
GO:0004386	helicase activity	0.0390	4.065
GO:0070011	peptidase activity, acting on L-amino acid peptides	0.0240	3.445
GO:0008233	peptidase activity	0.0240	3.332
GO:0016817	hydrolase activity, acting on acid anhydrides	0.0320	2.386
GO:0016787	hydrolase activity	0.0240	1.925

Table 3.17: Top ten overrepresented GO terms in the biological process domain for DE genes identified in Experiment 2, organized by effect type, then by up- (red text) or down-regulation (blue).

Identification of overrepresented terms was conducted on the AgriGO website, and the total list was then reduced using REViGO. No significant GO category enrichment was found for DE genes responding to an interaction effect of AOPP and MeJA.

GO Code	Category	Adjusted *p*	Fold Overrepresentation
	Responding to AOPP		
GO:0009073	aromatic amino acid family biosynthetic process	0.0001	16.282
GO:0006752	group transfer coenzyme metabolic process	0.0003	11.101
GO:0006119	oxidative phosphorylation	3.90×10^{-5}	7.712
GO:0034220	ion transmembrane transport	0.0003	6.838
GO:0006725	cellular aromatic compound metabolic process	3.90×10^{-5}	6.601
GO:0006818	hydrogen transport	0.0008	5.605
GO:0006732	coenzyme metabolic process	0.0020	4.684
GO:0009141	nucleoside triphosphate metabolic process	0.0003	4.333
GO:0006091	generation of precursor metabolites and energy	0.0002	4.310
GO:0055085	transmembrane transport	0.0010	3.816
	Plus 19 additional GO categories	*<0.05*	*3.694 – 1.258*
GO:0005992	trehalose biosynthetic process	0.0049	8.836
GO:0016137	glycoside metabolic process	0.0110	5.645
GO:0009311	oligosaccharide metabolic process	0.0110	5.348
GO:0043632	modification-dependent macromolecule catabolic process	0.0110	3.237
GO:0016567	protein ubiquitination	0.0300	3.161
GO:0070647	protein modification by small protein conjugation or removal	0.0310	3.093
GO:0009057	macromolecule catabolic process	0.0088	2.540
GO:0005975	carbohydrate metabolic process	1.20×10^{-5}	2.438
GO:0009056	catabolic process	0.0130	2.166
GO:0006508	proteolysis	0.0110	1.971
	Plus 7 additional GO categories	*<0.05*	*1.590 – 1.295*

Table 3.17, Continued:

GO Code	Category	Adjusted *p*	Fold Overrepresentation
	Responding to MeJA		
GO:0009073	aromatic amino acid family biosynthetic process	0.0004	8.266
GO:0008202	steroid metabolic process	0.0033	5.636
GO:0019684	photosynthesis, light reaction	0.0026	4.988
GO:0009119	ribonucleoside metabolic process	0.0330	3.827
GO:0006119	oxidative phosphorylation	0.0026	3.625
GO:0006260	DNA replication	0.0033	3.502
GO:0006091	generation of precursor metabolites and energy	3.70×10^{-6}	3.495
GO:0006096	glycolysis	0.0340	3.263
GO:0034220	ion transmembrane transport	0.0350	2.893
GO:0044283	small molecule biosynthetic process	1.60×10^{-6}	2.865
	Plus 21 additional GO categories	*<0.05*	*2.839 – 1.189*
GO:0032313	regulation of Rab GTPase activity	0.0220	3.860
GO:0009719	response to endogenous stimulus	0.0094	3.753
GO:0010033	response to organic substance	0.0094	3.753
GO:0016138	glycoside biosynthetic process	0.0250	3.712
GO:0009108	coenzyme biosynthetic process	0.0110	3.652
GO:0016137	glycoside metabolic process	0.0250	3.217
GO:0009311	oligosaccharide metabolic process	0.0310	3.048
GO:0016567	protein ubiquitination	0.0013	3.002
GO:0070647	protein modification by small protein conjugation or removal	0.0016	2.937
GO:0009141	nucleoside triphosphate metabolic process	0.0160	2.179
	Plus 25 additional GO categories	*<0.05*	*2.087 – 1.265*

Table 3.18: Top ten overrepresented GO terms in the molecular function domain for DE genes identified in Experiment 2, organized by effect type, then by up- (red text) or down-regulation (blue).

Identification of overrepresented terms was conducted on the AgriGO website, and the total list was then reduced using REViGO.

GO Code	Category	Adjusted *p*	Fold Overrepresentation
	Responding to AOPP		
GO:0015078	hydrogen ion transmembrane transporter activity	5.3×10^{-7}	8.373
GO:0016835	carbon-oxygen lyase activity	0.0003	7.598
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	0.0004	7.275
GO:0016769	transferase activity, transferring nitrogenous groups	0.0050	4.885
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	0.0011	4.627
GO:0051287	NAD or NADH binding	0.0069	4.440
GO:0016597	amino acid binding	0.0120	4.440
GO:0043176	amine binding	0.0120	4.440
GO:0031406	carboxylic acid binding	0.0190	3.877
GO:0019842	vitamin binding	0.0027	3.856
	Plus 18 additional GO categories	*<0.05*	*3.823 – 1.423*
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	0.0061	5.420
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	0.0170	4.619
GO:0016769	transferase activity, transferring nitrogenous groups	0.0170	4.065
GO:0004842	ubiquitin-protein ligase activity	0.0290	3.027
GO:0030170	pyridoxal phosphate binding	0.0390	2.790
GO:0016874	ligase activity	0.0061	2.345
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	0.0061	2.125
GO:0016798	hydrolase activity, acting on glycosyl bonds	0.0078	2.028
GO:0004175	endopeptidase activity	0.0470	2.024

Table 3.18, Continued:

GO Code	Category	Adjusted *p*	Fold Overrepresentation
	Responding to AOPP		
GO:0008270	zinc ion binding	0.0170	1.559
	Plus 9 additional GO categories	*<0.05*	*1.539 – 1.308*
	Responding to MeJA		
GO:0016836	hydro-lyase activity	0.0021	4.592
GO:0016597	amino acid binding	0.0013	4.133
GO:0043176	amine binding	0.0013	4.133
GO:0031406	carboxylic acid binding	0.0019	3.608
GO:0015078	hydrogen ion transmembrane transporter activity	0.0015	3.542
GO:0051287	NAD or NADH binding	0.0067	3.131
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	0.0019	3.045
GO:0016769	transferase activity, transferring nitrogenous groups	0.0360	2.755
GO:0019842	vitamin binding	0.0067	2.538
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	0.0070	2.516
	Plus 11 additional GO categories	*<0.05*	*2.237 – 1.275*
GO:0005097	Rab GTPase activator activity	0.0210	3.860
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	0.0016	3.860
GO:0005244	voltage-gated ion channel activity	0.0400	3.217
GO:0008289	lipid binding	0.0038	3.164
GO:0004842	ubiquitin-protein ligase activity	0.0004	3.080
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	0.0190	3.071
GO:0016769	transferase activity, transferring nitrogenous groups	0.0110	2.895
GO:0008171	O-methyltransferase activity	0.0260	2.875
GO:0000156	two-component response regulator activity	0.0170	2.673
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	0.0024	2.292
	Plus 17 additional GO categories	*<0.05*	*2.165 – 1.239*

Table 3.18, Continued:

GO Code	Category	Adjusted *p*	Fold Overrepresentation
Responding to AxM			
GO:0016491	oxidoreductase activity	0.0350	3.687
GO:0005524	ATP binding	0.0130	3.828
GO:0001883	purine nucleoside binding	0.0130	3.630
GO:0001882	nucleoside binding	0.0130	3.629
GO:0000166	nucleotide binding	0.0150	3.227

Table 3.19: Microarray gene expression data for phenylpropanoid core pathway genes in Experiment 1 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM (Wang et al. 2011).

Gene Name	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
PAL2	PtpAffx.128701.1.S1_s_at	973±122	1342±445	2021±42	1699±24	0.8909	0.0128	0.1026
PAL3	Ptp.4730.1.A1_s_at	16034±1563	16156±399	26732±3551	31289±6838	0.4480	0.0097	0.4702
C4H1	PtpAffx.150025.1.S1_s_at	25118±2751	31350±1953	37537±2073	38181±543	0.0717	0.0024	0.1192
C4H2	Ptp.336.1.S1_at	29930±5893	36558±1346	43790±6776	46119±60	0.2356	0.0218	0.5398
C4H2	Ptp.5618.1.S1_at	381±56	363±54	626±131	340±118	0.0890	0.1781	0.1198
C4H2	Ptp.6632.1.S1_at	14477±976	18534±329	22844±194	21754±2418	0.1865	0.0034	0.0507
4CL2	PtpAffx.12056.3.S1_a_at	12187±757	13212±840	17878±1109	22026±2486	0.0681	0.0022	0.2084

Table 3.20: Microarray gene expression data for phenylpropanoid core pathway genes in Experiment 2 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Gene Name	Probe ID	Ctrl	AOPP	MeJA	AOPP +MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
PAL4	PtpAffx.2272.1.S1_a_at	8056±890	9351±39	5846±186	7593±382	0.0121	0.0047	0.5531
PAL4	PtpAffx.2272.4.S1_a_at	69.2±7.0	67.9±1.1	56.0±1.4	48.6±8.7	0.3410	0.0153	0.4870
C4H2	Ptp.336.1.S1_at	12107±291	13689±241	12036±838	16754±3344	0.0620	0.2891	0.2700
C4H2	Ptp.5618.1.S1_at	221±29	392±78	379±35	521±4.8	0.0079	0.0109	0.6733
C4H2	Ptp.6632.1.S1_at	6638±877	10659±120	9046±1700	13374±1742	0.0103	0.0489	0.8750
4CL1	Ptp.3043.1.S1_s_at	5664±431	6346±325	4071±114	4749±1069	0.1851	0.0199	0.9957
4CL2	PtpAffx.12056.3.S1_a_at	10653±408	11441±17	12438±525	15384±1197	0.0182	0.0041	0.0897
4CL3	PtpAffx.87600.1.S1_at	101±27	90.0±13.6	23.2±6.5	40.3±9.2	0.7884	0.0052	0.2939

Table 3.21: Microarray relative gene expression data for lignin biosynthetic genes in Experiment 1 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Gene Name	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
C3H3	Ptp.1996.1.S1_s_at	4954±634	4121±1075	5755±886	5134±2172	0.4818	0.3888	0.9155
C3H3	Ptp.4675.1.S1_s_at	1572±424	1813±4.2	2721±211	2286±102	0.6018	0.0091	0.1197
COMT1	PtpAffx.141260.4.S1_at	135±67	191±50	40.6±37.6	33.6±9.0	0.4964	0.0179	0.3925
CCoAOMT2	Ptp.3608.1.S1_s_at	8889±621	8610±589	10643±894	10943±200	0.9830	0.0099	0.5494
HCT6	Ptp.6327.1.S1_at	779±10	833±196	518±112	409±5.6	0.7475	0.0128	0.3672
HCT6	PtpAffx.6492.2.A1_s_at	2097±13	2076±170	992±194	926±141	0.6989	4.12×10^{-4}	0.8383
Lac90b	PtpAffx.2130.2.A1_at	5991+1690	4525+117	1110+148	1157+75	0.3037	0.0024	0.2769

Table 3.22: Microarray relative gene expression data for lignin biosynthetic genes in Experiment 2 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Gene Name	Probe ID	Ctrl	AOPP	MeJA	AOPP +MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
C3H3	Ptp.1996.1.S1_s_at	3254±107	3829±36	4157±86	4281±435	0.0966	0.0138	0.2348
C3H3	Ptp.4675.1.S1_s_at	935±126	1257±134	1214±202	1210±102	0.1970	0.3217	0.1879
F5H2	PtpAffx.1193.1.S1_at	355±35	290±35	343±7.0	186±11	0.0036	0.0325	0.0648
CCoAOMT1	PtpAffx.11239.1.S1_s_at	7924±500	8251±153	6901±125	7036±637	0.4776	0.0192	0.7620
HCT6	Ptp.6327.1.S1_at	510±71	689±115	331±6.1	364±97	0.1459	0.0127	0.2843
HCT6	PtpAffx.6492.2.A1_s_at	1631±80	2007±168	1233±8.5	1289±138	0.0584	0.0024	0.1232
Lac90b	PtpAffx.2130.2.A1_at	7366+57	5234+574	4236+566	1868+293	0.0018	4.35×10^{-4}	0.7176

Table 3.23: Microarray gene expression data for flavonoid biosynthetic genes in Experiment 1 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Gene Name	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
CHS2/3	PtpAffx.7896.3.S1_a_at	35415±791	39229±173	75882±11273	75020±1145	0.7320	$6.87x10^{-4}$	0.5917
CHS6	PtpAffx.7896.2.S1_at	7337±1676	8880±806	18493±198	20376±4057	0.3392	0.0020	0.9194
CHS6	PtpAffx.7896.4.A1_a_at	20449±285	25072±892	53656±5281	65521±9551	0.1003	$6.82x10^{-4}$	0.4027
CHI	PtpAffx.4850.1.A1_s_at	1304±357	1604±392	2860±111	2980±28	0.3370	0.0015	0.6720
CHIL2	Ptp.1512.1.S1_s_at	21626±249	23555±482	39338±5728	40277±289	0.5204	0.0011	0.8200
CHIL2	PtpAffx.932.2.A1_at	148±113	137±62	274±11	265±14	0.8329	0.0506	0.9863
F3'H	Ptp.4863.1.S1_s_at	4022±246	4300±193	10746±1508	14770±3503	0.1870	0.0031	0.2384
F3'H	PtpAffx.120325.1.S1_s_at	1630±15.2	1993±192	4975±653	5827±1382	0.3270	0.0027	0.6770
F3'H	PtpAffx.142603.1.A1_s_at	2305±210	2439±170	5903±5.3	8461±1307	0.0460	$5.21x10^{-4}$	0.0620
F3H	Ptp.323.1.S1_s_at	31189±5003	38713±3006	66315±15220	67475±2402	0.4976	0.0054	0.6139
F3'5'H1	PtpAffx.83404.1.A1_at	69.3±23.5	32.9±36.4	356±102	309±167	0.5860	0.0166	0.9410
DFR1	PtpAffx.37082.1.A1_at	33758±4989	37339±1363	65955±3004	69657±14	0.1603	$1.08x10^{-4}$	0.9785
ANS1/2	PtpAffx.9044.2.S1_at	95.4±27.7	72.0±20.9	423±439	351±94	0.7798	0.1295	0.8861
ANS2	Ptp.6057.1.S1_at	7389±2283	8545±107	13920±676	17819±2480	0.1063	0.0029	0.3227
ANS2	Ptp.6057.1.S1_s_at	88.8±68.6	408±76	420±144	506±356	0.2239	0.2019	0.4537
BAN2	PtpAffx.5092.2.S1_a_at	15489±970	18352±2548	26751±759	29631±3412	0.1409	0.0020	0.9962
LAR2	Ptp.1080.1.S1_at	3059±243	2562±23	5064±1197	7224±971	0.2060	0.0038	0.0737

Table 3.23, Continued:

Gene Name	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
LAR2	Ptp.1080.1.S1_s_at	17685±1999	16038±345	29154±140	33971±1080	0.1234	5.53×10^{-5}	0.0165
LAR2	PtpAffx.6065.3.A1_a_at	4143±156	5325±706	7895±492	9059±2170	0.2290	0.0106	0.9918
FOMT1	PtpAffx.203863.1.S1_at	404±121	552±26	287±29	163±0.7	0.8033	0.0049	0.0391

Table 3.24: Microarray gene expression data for flavonoid biosynthetic genes in Experiment 2 (Mean±SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Gene Name	Probe ID	Ctrl	AOPP	MeJA	AOPP +MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
CHS4	PtpAffx.92231.1.S1_at	253±18	414±90	563±12	691±171	0.1015	0.0128	0.8188
CHS6	PtpAffx.7896.2.S1_at	8797±208	10057±406	10281±521	13276±1268	0.0141	0.0100	0.1644
CHS6	PtpAffx.7896.4.A1_a_at	18391±1235	20561±1584	19868±1677	25642±6037	0.1629	0.2314	0.4818
CHI	PtpAffx.4850.1.A1_s_at	1107±75	1781±105	2020±308	2928±377	0.0113	0.0045	0.5470
CHIL2	Ptp.1512.1.S1_s_at	12670±1191	13923±1245	14580±626	19054±1898	0.0373	0.0195	0.1594
CHIL2	PtpAffx.932.2.A1_at	65.5±17.8	90.7±1.3	119±9.5	179±41	0.0577	0.0119	0.3407
F3'H	Ptp.4863.1.S1_s_at	8111±406	10919±226	11536±309	15758±1518	0.0036	0.0019	0.2838
F3'H	PtpAffx.120325.1.S1_s_at	1542±199	2680±95	2871±520	5544±609	0.0029	0.0020	0.0590
F3'H	PtpAffx.142603.1.A1_s_at	6269±851	9004±461	9510±356	14810±592	6.69×10^{-4}	4.23×10^{-4}	0.0380
F3'5'H1	PtpAffx.83404.1.A1_at	45.1±8.4	29.3±31.6	206±17	355±18	0.0103	7.36×10^{-5}	0.0050
FLR	PtpAffx.202157.1.S1_at	6983±377	5804±71	4446±31	4085±207	0.0076	1.61×10^{-4}	0.0572
ANS1/2	PtpAffx.9044.2.S1_at	91.4±40.1	230±99	341±36	975±224	0.0119	0.0049	0.0486
ANS2	Ptp.6057.1.S1_at	6257±142	7537±894	8683±802	12509±2313	0.0505	0.0160	0.2399
ANS2	Ptp.6057.1.S1_s_at	111±111	151±21	238±3.4	358±50	0.1404	0.0189	0.4169
BAN1	PtpAffx.5092.1.A1_at	2507±262	3129±361	3369±440	5285±431	0.0092	0.0049	0.0737
LAR1	PtpAffx.6065.2.S1_at	376±51	309±43	506±100	759±88	0.1543	0.0053	0.0387
LAR2	Ptp.1080.1.S1_at	11131±871	13487±146	13166±457	16338±1210	0.0075	0.0116	0.5024

Table 3.24, Continued:

Gene Name	Probe ID	Ctrl	AOPP	MeJA	AOPP +MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
LAR2	Ptp.1080.1.S1_s_at	14680±295	16652±695	15745±529	20085±8855	0.0623	0.1415	0.3903
LAR2	PtpAffx.6065.3.A1_a_at	5446±1577	6720±1327	6086±92	7833±226	0.1086	0.2984	0.7632
LAR3	PtpAffx.18705.2.A1_a_at	48.4±2.3	57.6±6.2	89.6±3.8	150±42	0.0813	0.0112	0.1646
FOMT1	PtpAffx.203863.1.S1_at	543±106	385±96	257±26	148±18	0.0621	0.0073	0.6603
FOMT2/7	PtpAffx.218018.1.S1_s_at	1840±331	2164±221	1070±173	905±144	0.6478	0.0033	0.2047
FOMT7	PtpAffx.218018.1.S1_at	461±66	492±12	250±35	219±12	0.9950	8.81×10^{-4}	0.3188
FOMTL1	PtpAffx.113180.1.S1_at	1045±92	1755±152	1527±8.9	1982±85	0.0011	0.0071	0.1420
FOMTL1	PtpAffx.143031.1.S1_at	13.7±6.8	29.7±14.8	70.9±28.2	70.8±24.4	0.6100	0.0268	0.6054
FOMTL1	PtpAffx.204504.1.S1_at	219±19	328±0.2	520±102	516±98	0.3551	0.0084	0.3251
FOMTL1	PtpAffx.204504.1.S1_s_at	1416±158	1984±180	1774±52	2127±216	0.0163	0.0961	0.4063

Table 3.25: Microarray gene expression data for nitrogen-related genes in Experiment 1 (Mean±SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Probe Annotation	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
CM1	PtpAffx.46815.1.S1_at	2373±27	2041±47	1724±856	1233±29	0.2462	0.0743	0.8073
CM1	PtpAffx.46815.2.S1_a_at	5230±569	5771±407	8330±140	7733±153	0.9181	6.05×10^{-4}	0.0922
ADT/PDT1	Ptp.3045.1.S1_at	4419±919	5657±17	10232±406	8977±913	0.9860	6.83×10^{-4}	0.0600
ADT/PDT1	Ptp.3045.1.S1_s_at	4883±260	5158±68	10376±94	12300±2082	0.2131	0.0010	0.3291
ADT/PDT1	PtpAffx.6463.1.S1_at	5031±952	3713±70	8123±384	11081±44	0.0873	1.36×10^{-4}	0.0042
ADT/PDT2	Ptp.6414.1.S1_at	3874±297	3525±172	5656±358	6082±954	0.9249	0.0047	0.3657
ADT/PDT2	PtpAffx.52548.1.A1_at	2401±290	3060±17	4049±352	4745±905	0.1314	0.0096	0.9615
ASB2	PtpAffx.208095.1.S1_at	20063±630	21195±320	20356±412	24259±304	0.0012	0.0055	0.0109
ASB2	PtpAffx.208095.1.S1_s_at	18240±1302	19401±93	19255±498	21805±127	0.0201	0.0261	0.2338
IGPS1	PtpAffx.41412.1.A1_at	4386±149	3981±142	3391±278	4433±95	0.0659	0.0992	0.0047
TSB4	PtpAffx.61105.1.S1_at	1113±252	1194±61	1752±623	1949±9.1	0.3492	0.0059	0.6816
NRT2.1	Ptp.5713.2.S1_a_at	1263+288	1192+201	562+66	685+110	0.8540	0.0102	0.5040
NR2	PtpAffx.102478.1.A1_at	2107±48	3072±136	2467±327	1297±433	0.6330	0.0235	0.0058
GS1.1b	PtpAffx.4733.7.S1_s_at	79±9.3	226±52	96±22	19±4.7	0.1637	0.0096	0.0054
GS1.2	Ptp.2660.1.A1_at	11820±117	12236±879	7355±283	9857±1797	0.1158	0.0093	0.2256
GS2	Ptp.2169.2.S1_s_at	1612±171	1672±309	2575±292	2908±108	0.3034	0.0027	0.4588
OMR1	Ptp.7834.1.S1_s_at	5716±208	5371±93	5645±184	6833±386	0.0695	0.0153	0.0110

Table 3.25, Continued:

Probe Annotation	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
OMR1	PtpAffx.60077.1.S1_at	2119±79	2513±8.8	2079±100	2050±78	0.0261	0.0090	0.0163
Putative ammonium transporter a	PtpAffx.225140.1.S1_s_at	2312±126	1895±100	1783±100	2586±349	0.2403	0.5961	0.0121
AMT1	PtpAffx.3665.1.S1_a_at	2160±429	1549±158	3215±395	4532±103	0.1797	7.41×10^{-4}	0.0113

Table 3.26: Microarray gene expression data for nitrogen-related genes in Experiment 2 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM.

Probe Annotation	Probe ID	Ctrl	AOPP	MeJA	AOPP+ MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
CM1	PtpAffx.46815.1.S1_at	3821±368	3821±53	4632±362	5025±532	0.4960	0.0185	0.4972
CM1	PtpAffx.46815.2.S1_a_at	3175±173	3263±153	3820±303	4695±641	0.1420	0.0170	0.2099
ADT/PDT1	Ptp.3045.1.S1_at	2536±68	3354±346	3798±473	5838±929	0.0214	0.0086	0.1910
ADT/PDT1	Ptp.3045.1.S1_s_at	5499±409	7025±78	7607±654	11099±230	9.29×10^{-4}	4.15×10^{-4}	0.0263
ADT/PDT1	PtpAffx.6463.1.S1_at	8229±226	10249±487	11254±462	14504±289	6.21×10^{-4}	1.76×10^{-4}	0.0853
ADT/PDT2	Ptp.6414.1.S1_at	5190±180	5714±117	7115±469	8003±407	0.0384	8.17×10^{-4}	0.4782
ADT/PDT2	PtpAffx.52548.1.A1_at	1829±144	2310±207	2710±483	3002±251	0.1427	0.0207	0.6784
ADT/PDT3	PtpAffx.140097.1.A1_a_at	5115±276	5660±250	5074±29	6025±214	0.0079	0.3464	0.2529
PABAS	PtpAffx.209398.1.S1_at	826±19	855±50	720±19	742±57	0.4194	0.0191	0.8993
PrAT1	PtpAffx.208680.1.S1_at	201±2.6	225±8.0	187±2.6	198±7.4	0.0136	0.0069	0.1999
TSA1	Ptp.3195.1.S1_s_at	1319±64	1269±66	1524±54	1505±6.5	0.4063	0.0042	0.7059
TSB2/3	PtpAffx.63404.1.A1_at	979±178	966±121	1354±19	1328±49	0.8156	0.0093	0.9407
TSB4	PtpAffx.61105.1.S1_at	790±94	804±19	1494±20	1609±19	0.1424	2.86×10^{-5}	0.2267
NRT2.1	Ptp.5713.2.S1_a_at	1436±47	1309±24	691±67	592±15	0.0214	1.85×10^{-5}	0.6830
NRT2.7	PtpAffx.141124.2.S1_s_at	179±5.5	215±4.9	123±8.4	113±40	0.4244	0.0058	0.1888
NR1	Ptp.10.1.A1_at	897±41	779±76	592±82	426±28	0.0306	0.0016	0.6063
GS1.1a	Ptp.848.1.S1_at	3373±169	2290±293	3221±105	2112±255	0.0021	0.3462	0.9352

Table 3.26, Continued:

Probe Annotation	Probe ID	Ctrl	AOPP	MeJA	AOPP+ MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
GS1.1a	Ptp.848.1.S1_x_at	3089±352	2551±188	3033±236	2004±362	0.0197	0.2209	0.3032
GS2	Ptp.2169.2.S1_s_at	1615±103	1671±74	1744±53	2038±46	0.0268	0.0084	0.0819
OMR1	Ptp.7834.1.S1_s_at	5024±204	5457±166	4878±76	5444±133	0.0098	0.5009	0.5698
OMR1	PtpAffx.60077.1.S1_at	2216±331	2626±565	2314±194	2962±104	0.0963	0.4240	0.6525
Putative ammonium transporter b	PtpAffx.20015.1.A1_s_at	219±23	220±12	328±4.7	340±36	0.7096	0.0020	0.7374
AMT1	PtpAffx.3665.1.S1_a_at	4603±106	4034±604	6009±347	5477±14	0.0915	0.0046	0.9427
NADH-dependent Glu synthase a	PtpAffx.213253.1.S1_x_at	224±18	210±11	153±8.5	187±30	0.5064	0.0236	0.1529
NADH-dependent Glu synthase b	PtpAffx.85882.1.S1_s_at	814±36	958±55	853±39	694±46	0.8296	0.0229	0.0085

Table 3.27: Microarray gene expression data for genes related to citric acid cycle metabolites in Experiment 1 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM. *Not directly related to the mitochondrial citric acid cycle.

Probe Annotation	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
citrate synthase	PtpAffx.5281.1.S1_a_at	5734±224	5867±662	8925±302	9523±191	0.2582	$2.48x10^{-4}$	0.4491
malate dH 1, mitochondrial	Ptp.2139.1.S1_a_at	2229±551	1581±231	1778±108	3132±262	0.2062	0.0782	0.0128
malate translocator	PtpAffx.640.1.S1_at	4455±6.3	4518±115	5463±110	6504±453	0.0313	$9.10x10^{-4}$	0.0448
ATP citrate lyase a	PtpAffx.138813.1.A1_s_at	3890±56	3642±133	6599±70	7920±2059	0.5035	0.0087	0.3432
ATP citrate lyase a	PtpAffx.138813.2.A1_at	6421±631	5615±243	8809±611	10987±393	0.1226	$3.80x10^{-4}$	0.0131
ATP citrate lyase a	PtpAffx.158517.1.S1_at	4318±13	4687±168	6063±216	7193±526	0.0232	$5.31x10^{-4}$	0.1437
ATP citrate lyase b	Ptp.2332.2.A1_a_at	6256±302	6473±266	10378±1693	12156±1738	0.3153	0.0049	0.4201
ATP citrate lyase c	Ptp.2370.1.A1_at	4901±524	5018±296	8626±937	9189±1917	0.6860	0.0073	0.7896
ATP citrate lyase c	Ptp.4670.1.S1_s_at	4156±194	4410±81	7035±411	7889±4668	0.7553	0.1277	0.8654
ATP citrate lyase c	PtpAffx.83148.2.S1_a_at	187±73	377±108	562±71	332±50	0.7283	0.0402	0.0190
ATP citrate lyase c/d	PtpAffx.83148.1.S1_s_at	1967±283	2043±238	4994±660	4796±1325	0.9153	0.0058	0.8119
ATP citrate lyase c/d	PtpAffx.83148.2.S1_s_at	3037±150	3786±42	6389±211	6175±1662	0.6770	0.0080	0.4640
succinyl-CoA ligase β	PtpAffx.150471.1.S1_s_at	1338±81	1418±38	948±84	971±65	0.3540	0.0010	0.5907
isocitrate dH	Ptp.2088.1.S1_at	11452±290	10540±30	11813±633	13935±685	0.1550	0.0056	0.0118
isocitrate dH	Ptp.5871.1.S1_at	5224±72	4094±710	4200±155	6284±472	0.1955	0.1305	0.0064

Table 3.27, Continued:

Probe Annotation	Probe ID	Ctrl	PIP	MeJA	PIP+ MeJA	PIP *p*-value	MeJA *p*-value	PxM *p*-value
*NAD^+ dependent isocitrate dH subunit 1 a**	Ptp.8073.1.A1_s_at	743±125	697±126	1016±12	1281±213	0.3264	0.0121	0.1888
pyruvate dH E1β a/b	PtpAffx.450.1.S1_a_at	10667±493	11040±296	12336±1986	15005±394	0.1104	0.0193	0.1978
pyruvate dH E1β a	PtpAffx.450.1.S1_at	7121±24	5025±1021	5697±318	7817±4720	0.9950	0.7098	0.2854
pyruvate dH E1βa	PtpAffx.450.6.S1_a_at	2935±108	3487±199	3884±211	4552±475	0.0382	0.0074	0.7862
succinate dH iron-protein subunit-like	Ptp.6262.1.S1_at	5742±42	6349±396	6984±13	7448±251	0.0323	0.0022	0.6906

Table 3.28: Microarray gene expression data for genes related to citric acid cycle metabolites in Experiment 2 (Mean+SD, N=2).

Only probes found in the DE List are shown. Red text indicates significant *p*-value according to statistical analysis using a Bayesian equivalent of two-way ANOVA in conjunction with SLIM. *Not directly related to the mitochondrial citric acid cycle.

Probe Annotation	Probe ID	Ctrl	AOPP	MeJA	AOPP+ MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
malate dH 1, mitochondrial	Ptp.2139.1.S1_a_at	2196±131	2274±117	1996±13	2537±77	0.0104	0.6725	0.0270
malate dH 2, mitochondrial	PtpAffx.7625.4.S1_a_at	1502±64	1832±70	1605±63	2031±183	0.0077	0.1186	0.5690
malate translocator	PtpAffx.640.1.S1_at	6690±188	6926±101	6998±16	7916±443	0.0294	0.0203	0.1213
malic enzyme 1	Ptp.3086.1.S1_s_at	844±13	942±30	1079±45	1084±67	0.1677	0.0035	0.2045
malic enzyme 2	PtpAffx.100.1.S1_at	476±20	471±15	558±8.7	584±1.5	0.3236	4.86×10^{-4}	0.1679
CIT1	PtpAffx.60905.1.S1_at	5432±70	5679±49	5147±150	5856±244	0.0107	0.6352	0.0942
CIT1	Ptp.4502.1.S1_at	2081±69	2427±203	2154±85	2542±48	0.0117	0.3235	0.8163
ATP citrate lyase a	PtpAffx.138813.1.A1_s_at	3406±224	4690±112	5475±482	7502±393	0.0022	5.02×10^{-4}	0.1919
ATP citrate lyase a	PtpAffx.138813.2.A1_at	8105±246	9208±553	9428±206	11331±525	0.0068	0.0042	0.2426
ATP citrate lyase a	PtpAffx.158517.1.S1_at	3927±331	5102±26	5308±286	7331±367	0.0014	8.71×10^{-4}	0.1041
ATP citrate lyase b	Ptp.2332.2.A1_a_at	6532±48	8004±473	7886±302	10171±439	0.0017	0.0022	0.1829
ATP citrate lyase c	Ptp.2370.1.A1_at	5148±84	5682±188	6878±86	8961±1261	0.0445	0.0052	0.1624
ATP citrate lyase c	Ptp.4670.1.S1_s_at	5896±241	6782±200	7251±378	9672±328	0.0014	5.26×10^{-4}	0.0213
ATP citrate lyase c	PtpAffx.83148.2.S1_a_at	253±22	317±61	419±41	564±67	0.0448	0.0046	0.3254
ATP citrate lyase c/d	PtpAffx.83148.1.S1_s_at	2578±656	3371±323	3799±457	5771±577	0.0196	0.0079	0.1835

Table 3.28, Continued:

Probe Annotation	Probe ID	Ctrl	AOPP	MeJA	AOPP+ MeJA	AOPP *p*-value	MeJA *p*-value	AxM *p*-value
ATP citrate lyase c/d	PtpAffx.83148.2.S1_s_at	2014±264	3030±419	3184±508	4601±33	0.0080	0.0050	0.4680
succinyl-CoA synthetase	PtpAffx.11524.3.S1_a_at	5775±128	6014±97	5650±122	6691±227	0.0040	0.0622	0.0202
*microbody NAD^+-dependent malate dH**	Ptp.1238.1.S1_at	1517±66	1630±7.5	1716±38	1943±70	0.0098	0.0022	0.1973
cytosolic malate dH	PtpAffx.390.6.S1_s_at	8263±378	9489±107	8023±290	9644±237	0.0018	0.8368	0.3619
*NAD^+ dependent isocitrate dH subunit 1 b**	PtpAffx.164055.1.S1_at	4121±132	3941±276	3381±140	3173±93	0.1912	0.0036	0.9142
putative β-Ala/ pyruvate aminotransferase	Ptp.2105.1.S1_at	1104±133	1370±52	891±26	1038±30	0.0171	0.0065	0.3212
pyruvate dH E1β a/b	PtpAffx.450.1.S1_a_at	9200±414	9908±359	9565±188	11098±466	0.0130	0.0417	0.1920
pyruvate dH E1β a	PtpAffx.450.1.S1_at	7345±48	7753±158	7671±166	9797±202	3.18×10^{-4}	4.11×10^{-4}	0.0014
pyruvate dH E1β a	PtpAffx.450.6.S1_a_at	2856±105	3602±224	3651±195	4344±71	0.0032	0.0025	0.8317
succinate dH iron-protein subunit-like	Ptp.6262.1.S1_at	7241±26	7864±116	7336±94	7978±184	0.0017	0.2822	0.9159

Figures

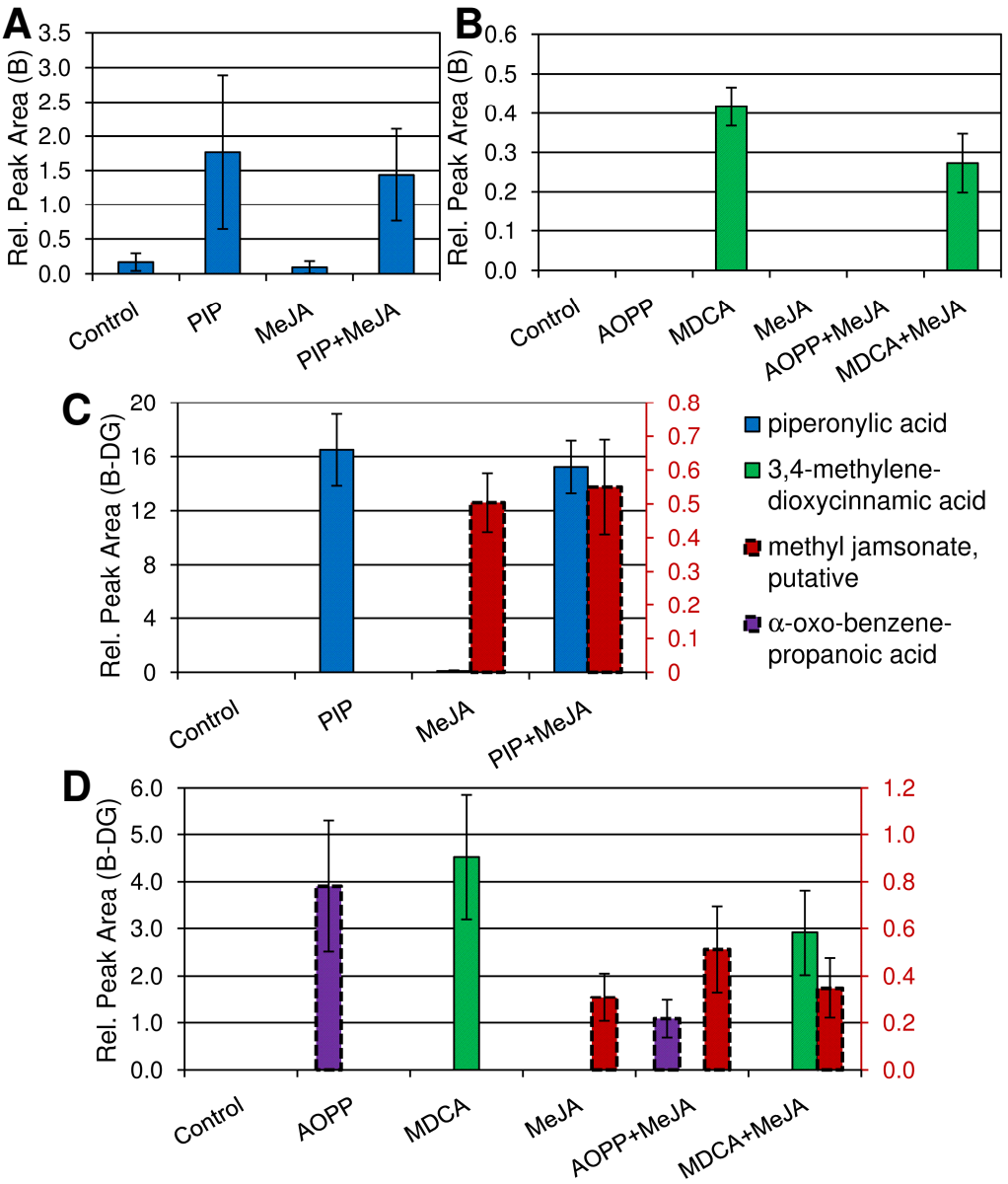

Figure 3.1: Fed elicitors and inhibitors detected in methanolic extracts of *Populus* cells.

Total ion chromatogram (TIC) peak areas (means±SD) from GC-MS analysis of Advanta-bound fraction of methanolic extracts were normalized relative to the internal standard *o*-anisic acid. Bars with dashed border are scaled to the y-axis on the right side of the graph. (A) Experiment 1, B fraction; (B) Experiment 2, B fraction; (C) Experiment 1, B-DG fraction; (D) Experiment 2, B-DG fraction. The putative jasmonate (called as

methyl jasmonate by AnalyzerPro software) was found only in samples subjected to β-glucosidase treatment. AOPP was not found in any samples, but a putative α-oxo-benzenepropanoic acid is a reasonable candidate for testing against a known AOPP standard.

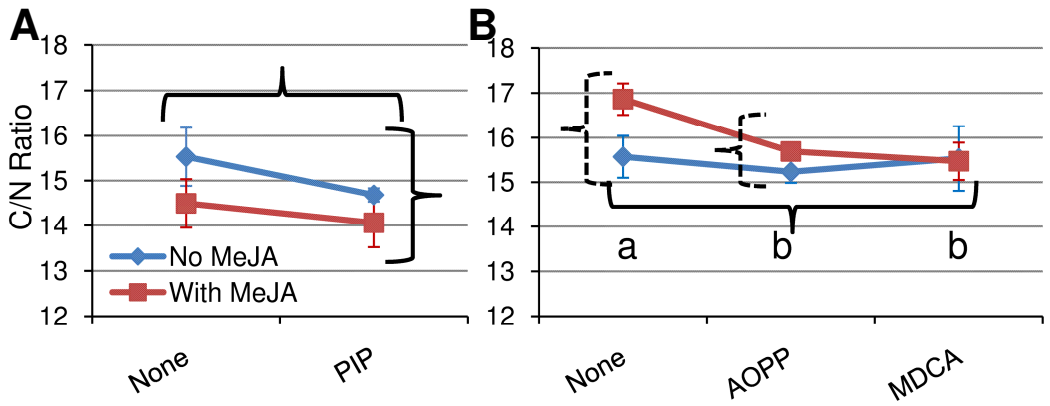

Figure 3.2 Carbon to nitrogen ratio in treated *Populus* cell cultures.

Data show means±SD; N=4 for each treatment. Solid brackets indicate significant overall effects according to two-way ANOVA; samples with the same letter are not significantly different according to Hsu's MCB. Dashed brackets indicate significant MeJA effects when slicing the MeJAxInhibitor interaction effect. (A) Experiment 1; (B) Experiment 2.

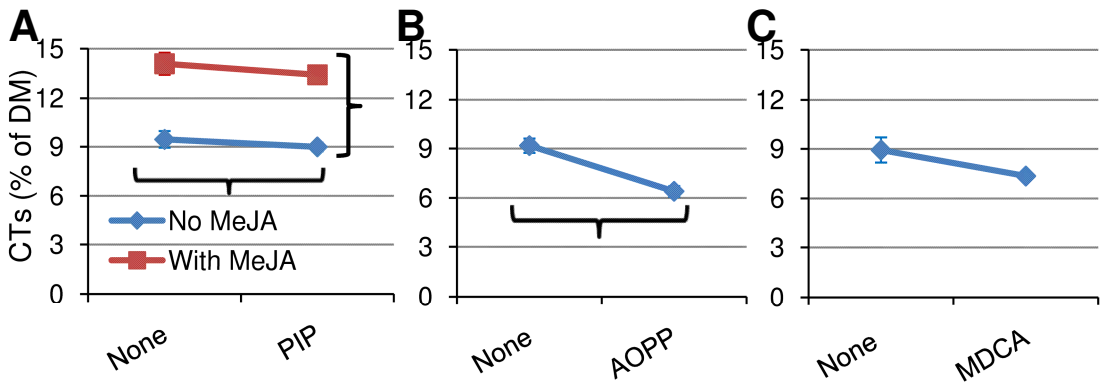

Figure 3.3 Treatment effects on condensed tannin levels in *Populus* cell suspension cultures after 24 h, as quantified by Dr. Payyavula in preliminary trials.

Data show means±SD; brackets indicate significant overall effects according to ANOVA. (A) Trial for PIP and MeJA feeding, N=4; (B) Trial for AOPP feeding, N=3; (C) Trial for MDCA feeding, N=2.

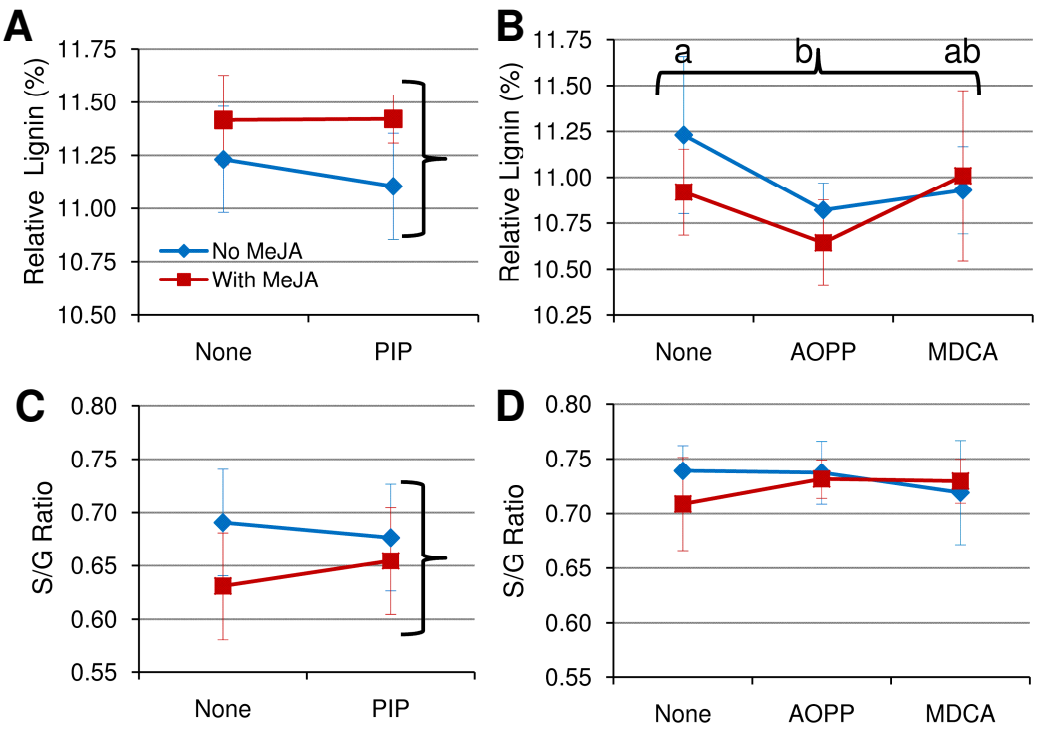

Figure 3.4 Changes in lignin content and S/G ratio (means+SD) in response to MeJA or inhibitor feeding in *Populus* cell suspension cultures.

(A) Lignin content in Experiment 1 samples; (B) Lignin content in Experiment 2 samples; (C) S/G ratio in Experiment 1 samples; (D) S/G ratio in Experiment 2 samples. Brackets indicate significant overall effects according to two-way ANOVA; samples with the same letter are not significantly different according to Hsu's MCB. For both analyses, Experiment 1 has N=3-4 and Experiment 2 has N=6-7.

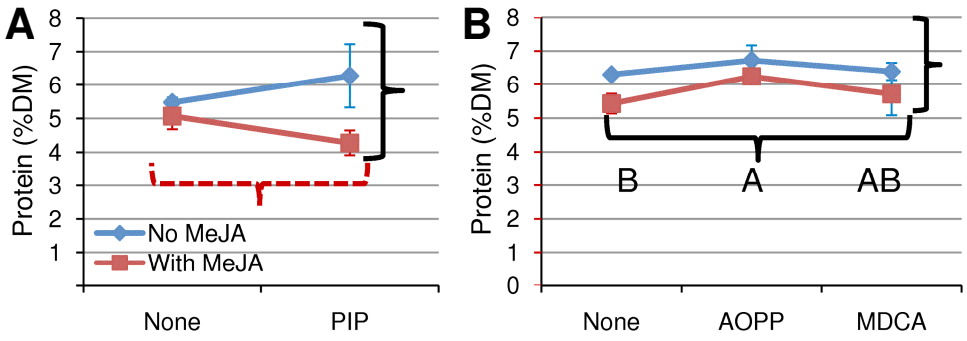

Figure 3.5: Treatment effects on protein levels in *Populus* cell suspension cultures after 48 h.

Data shown are means+SD. Solid brackets indicate significant overall effects according to two-way ANOVA, while dashed bracket indicates significant inhibitor effect within MeJA treated cells. Samples with the same letter are not significantly different overall according to Hsu's MCB test. (A) Experiment 1, N=3-4; (B) Experiment 2, N=6.

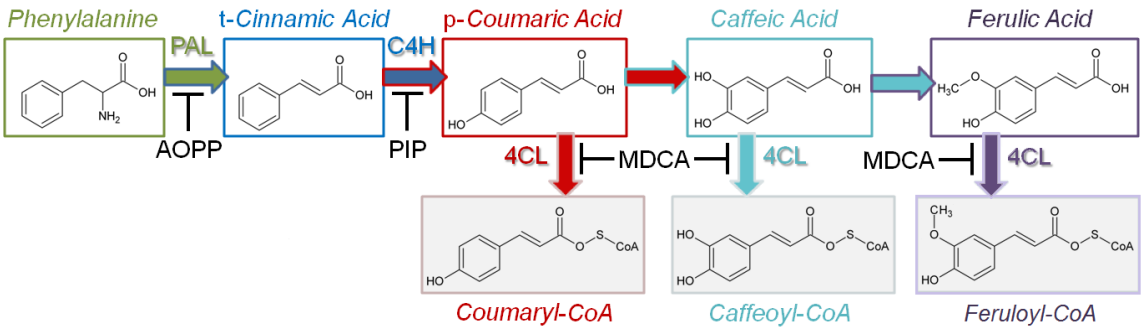

Figure 3.6: Enzyme targets of phenylpropanoid metabolic inhibitors.

Colored text identifies the enzymes of interest in the current work. Substrate and product names are in italics, with names and structures indicated by colors corresponding to histogram bars in Figures 3.7 & 3.8. Inhibitor acronyms are in black text.

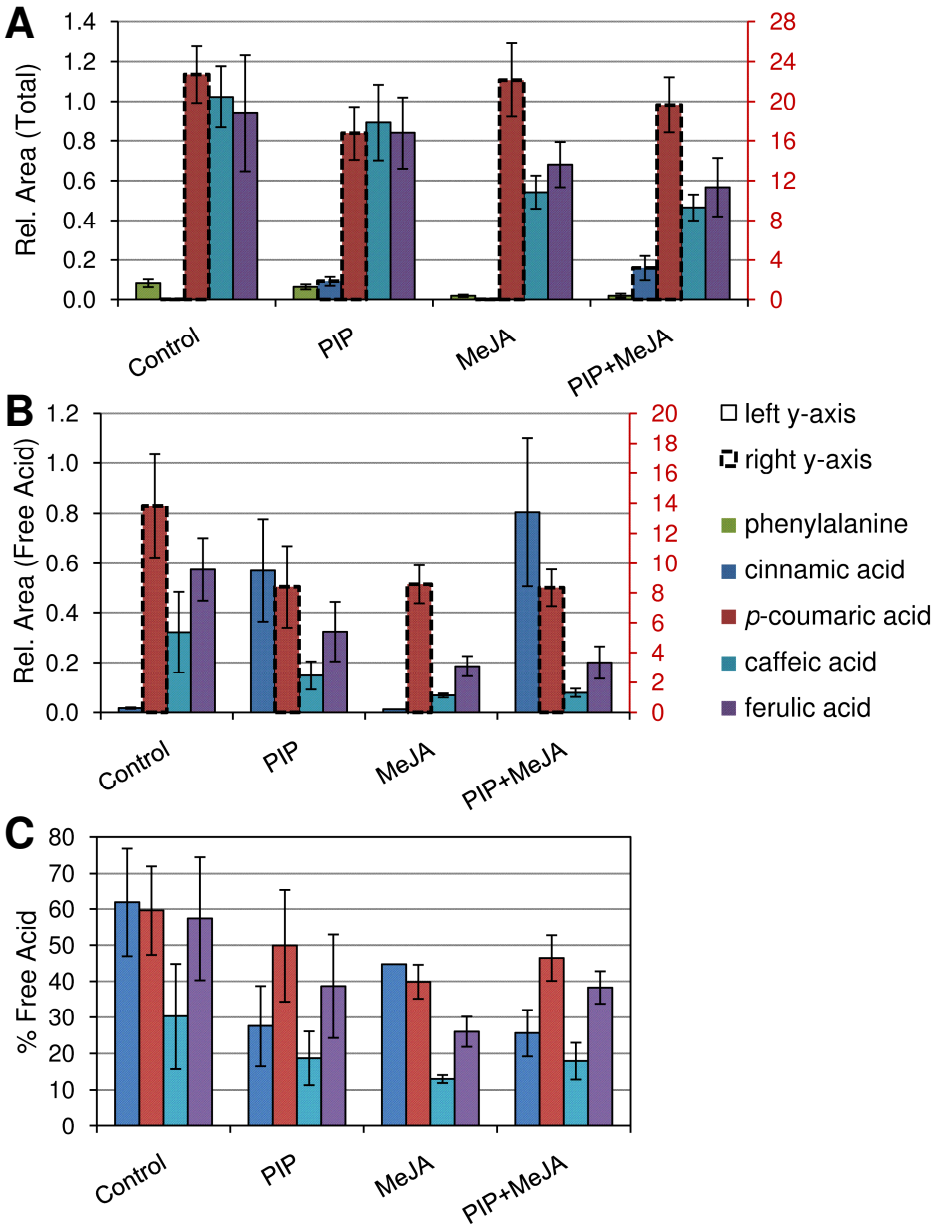

Figure 3.7: Relative levels (Mean+SD) of phenylalanine and hydroxycinnamate pools detected in *Populus* cell suspension cultures in Experiment 1.

Bars with dashed border are scaled to the y-axis on the right side of the graph. (A) Combined free-acid and glucose conjugates from B-DG fraction along with phenylalanine from UB fraction, N=6 or 7; (B) free-acid alone from B fraction, N=4; (C) percent of combined pool accounted for by the free-acid, N=4. Free cinnamic acid was detected only in one sample in the MeJA treatment group.

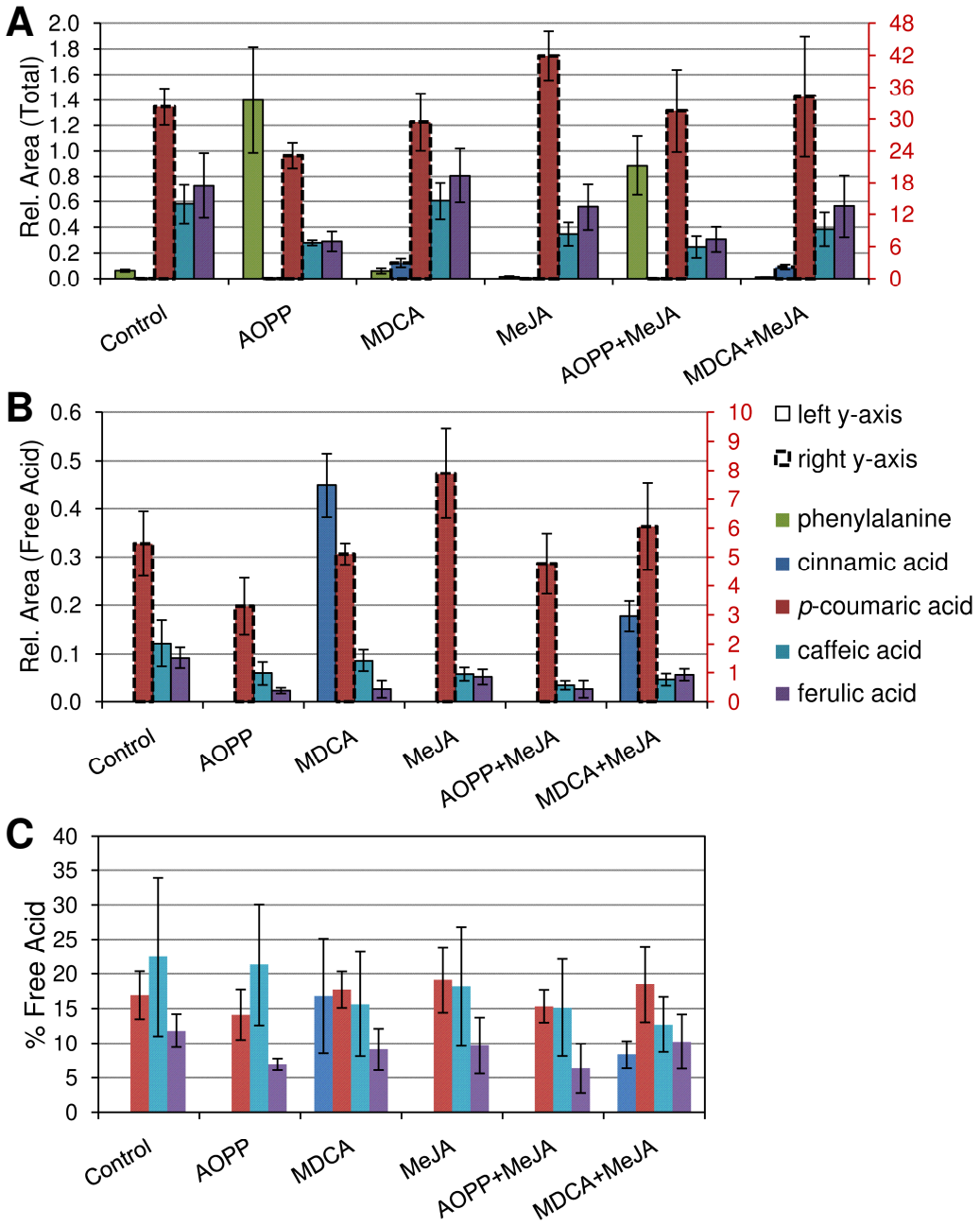

Figure 3.8: Relative levels (Mean+SD) of phenylalanine and hydroxycinnamate pools detected in *Populus* cell suspension cultures in Experiment 2.

Bars with dashed border are scaled to the y-axis on the right side of the graph; N=6 for all graphs. (A) Combined free-acid and glucose conjugates from B-DG fraction along with phenylalanine from UB fraction; (B) free-acid alone from B fraction; (C) percent of combined pool accounted for by the free-acid.

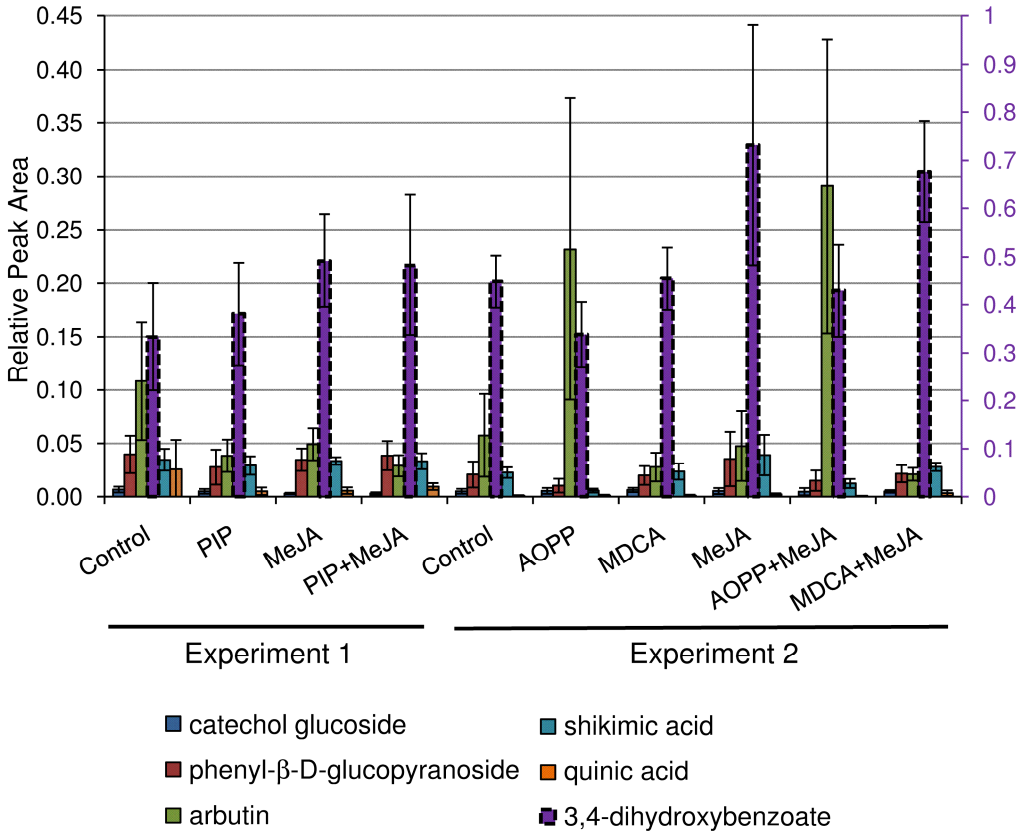

Figure 3.9: Responses of additional phenylpropanoid related compounds to feeding of elicitor and/or inhibitor.

Data are means±SD relative to *o*-anisic acid (3,4-dihydroxybenzoate) or adonitol (all other compounds).

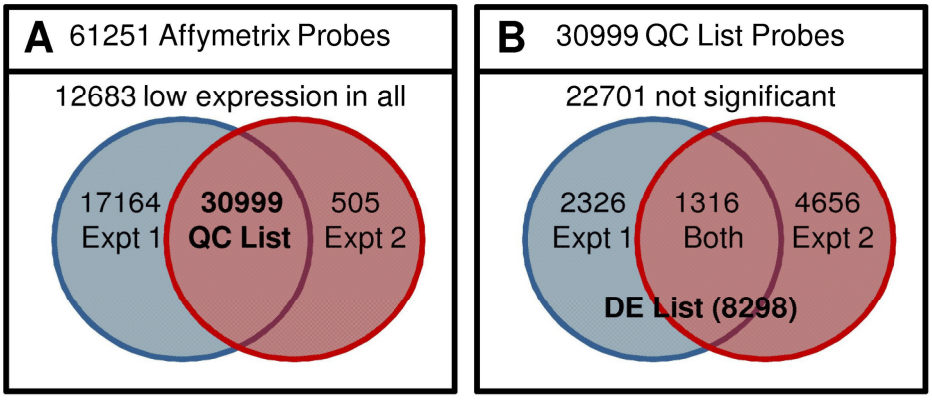

Figure 3.10: Venn diagrams outlining microarray analyses.

(A) Reduction of data from the total probe list involved removing non-*Populus* probes, as well as those that were below a minimal expression threshold (**Methods**). The intersection of the two lists ("QC List") was used for HCA and further statistical analysis. (B) Proportion of statistically significant probes from the QC List in each experimental dataset determined by separate two-way ANOVAs and SLIM (Wang et al. 2011). The union of the two lists ("DE List") indicates the total number of probes found statistically significant in at least one condition in one experiment.

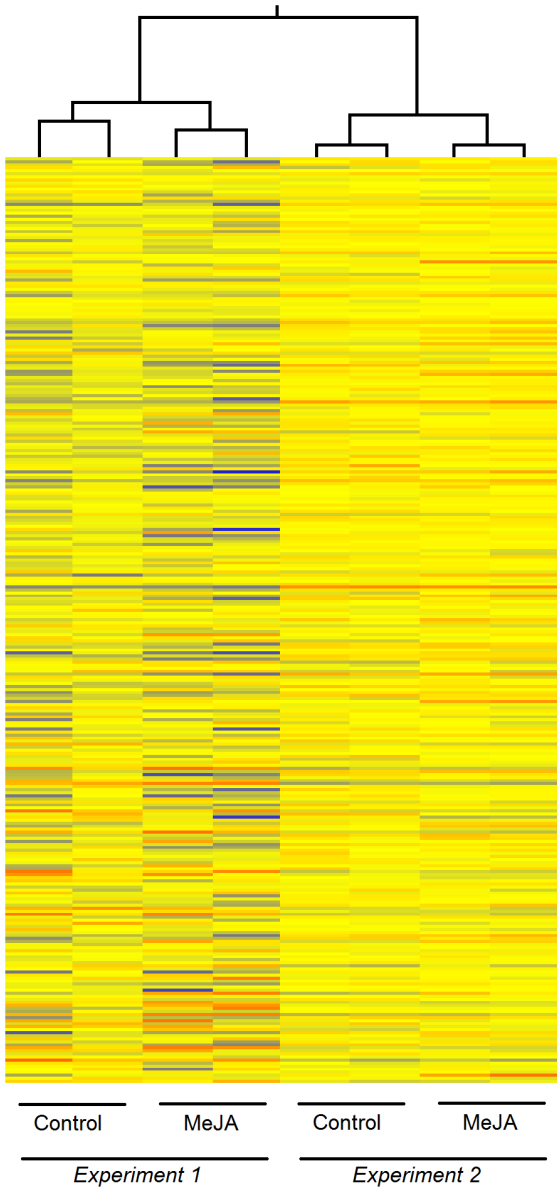

Figure 3.11: Hierarchical Clustering Analysis of Control and MeJA treated Cell Cultures from Experiment 1 and Experiment 2.

Pearson Centered was used as the distance metric in conjunction with complete linkage to test relationships among samples for all QC List probes.

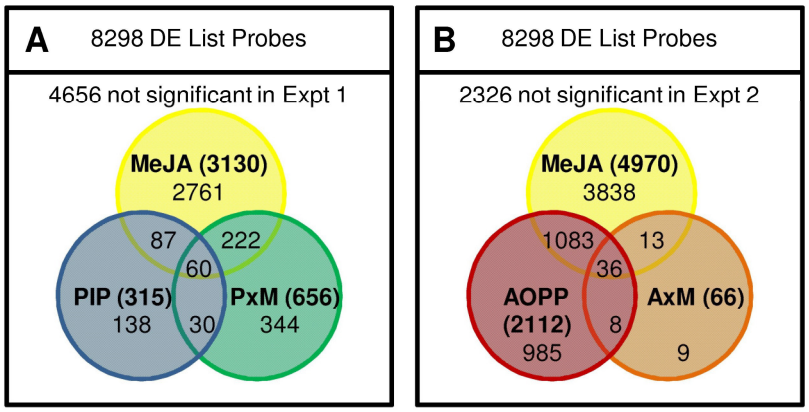

Figure 3.12: Venn diagrams outlining statistical analysis of microarray data.

(A) Experiment 1; (B) Experiment 2.

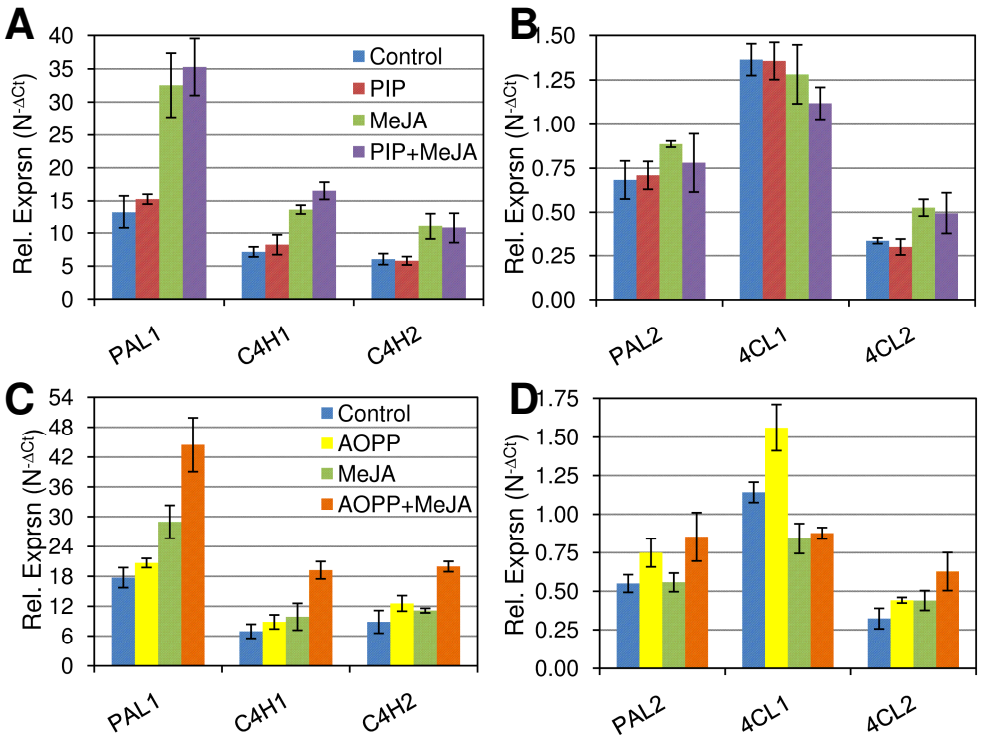

Figure 3.13: Relative expression (Mean+SD, N=3) of phenylpropanoid core pathway gene expression determined by QPCR.

(A) Experiment 1, higher expressers; (B) Experiment 1, lower expressers; (C) Experiment 2, higher expressers; (D) Experiment 2, lower expressers.

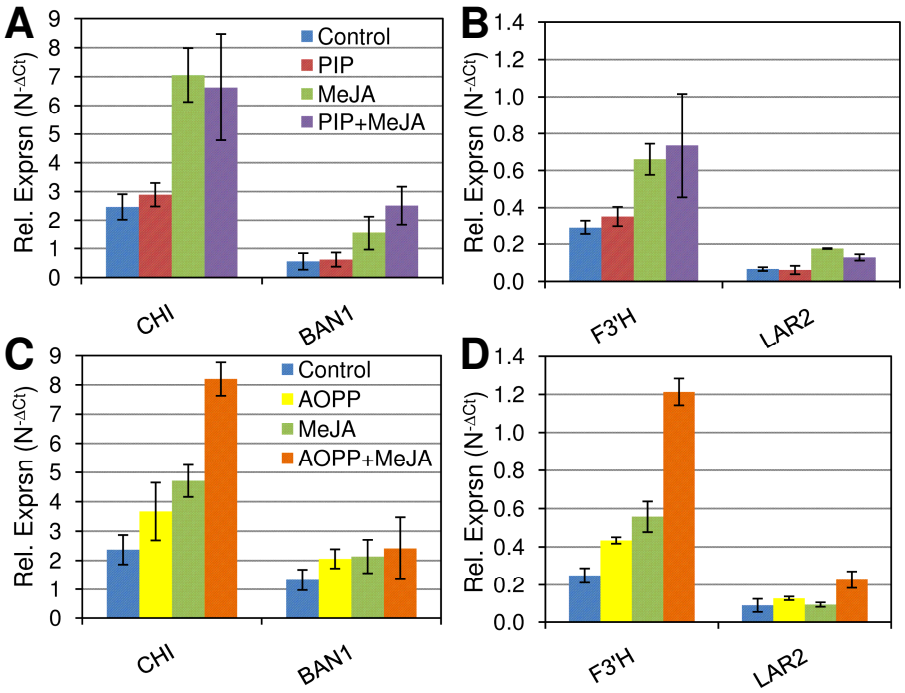

Figure 3.14: Relative expression (Mean+SD, N=3) of flavonoid pathway genes determined by QPCR.

(A) Experiment 1, higher expressers; (B) Experiment 1, lower expressers; (C) Experiment 2, higher expressers; (D) Experiment 2, lower expressers.

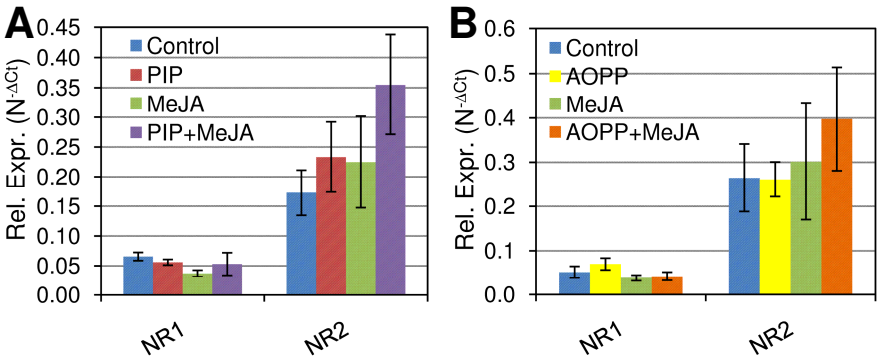

Figure 3.15: Relative expression (Mean+SD, N=3) of nitrate reductase genes according to QPCR.

(A) Experiment 1; (B) Experiment 2.

References

(2012) MetaLab. [http://128.192.158.63/x/MetaLab/]

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* 16: 3098-3109

Akagi T, Ikegami A, Suzuki Y, Yoshida J, Yamada M, Sato A, Yonemori K (2009) Expression balances of structural genes in shikimate and flavonoid biosynthesis cause a difference in proanthocyanidin accumulation in persimmon (*Diospyros kaki* Thunb.) fruit. *Planta* 230: 899-915

Amrhein N, Gerhardt J (1979) Superinduction of phenylalanine ammonia-lyase in gherkin hypocotyls caused by the inhibitor, L-α-aminooxy-β-phenylpropionic acid. *Biochimica et Biophysica Acta (BBA) - General Subjects* 583: 434-442

Amrhein N, Gödeke K-H (1977) α-aminooxy-β-phenylpropionic acid -- a potent inhibitor of L-phenylalanine ammonia-lyase *in vitro* and *in vivo*. *Plant Science Letters* 8: 313-317

Amrhein N, Gödeke K-H, Kefeli V (1976) The estimation of relative intracellular phenylalanine ammonia-lyase (PAL)-activities and the modulation *in vivo* and *in vitro* by competitive inhibitors. *Ber Deutsch Bot Ges Bd* 89: 247-259

An Y, Shen Y-B, Wu L-J, Zhang Z-X (2006) A change of phenolic acids content in poplar leaves induced by methyl salicylate and methyl jasmonate. *Journal of Forestry Research (Harbin)* 17: 107-110

Ananieva KI, Ananiev ED (1999) Effect of methyl ester of jasmonic acid and benzylaminopurine on growth and protein profile of excised cotyledons of *Cucurbita pepo* (Zucchini). *Biologia Plantarum* 42: 549-557

Arnold T, Appel H, Patel V, Stocum E, Kavalier A, Schultz J (2004) Carbohydrate translocation determines the phenolic content of *Populus* foliage: A test of the sink–source model of plant defense. *New Phytologist* 164: 157-164

Arnold T, Schultz J (2002) Induced sink strength as a prerequisite for induced tannin biosynthesis in developing leaves of *Populus*. *Oecologia* 130: 585-593

Babst B, Harding S, Tsai C-J (2010) Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in *Populus*. *Journal of Chemical Ecology* 36: 286-297

Babst B, Sjödin A, Jansson S, Orians C (2009) Local and systemic transcriptome responses to herbivory and jasmonic acid in *Populus*. *Tree Genetics & Genomes* 5: 459-474

Babst BA, Ferrieri RA, Gray DW, Lerdau M, Schlyer DJ, Schueller M, Thorpe MR, Orians CM (2005) Jasmonic acid induces rapid changes in carbon transport and partitioning in *Populus*. *New Phytologist* 167: 63-72

Bae H, Herman E, Bailey B, Bae H-J, Sicher R (2005a) Exogenous trehalose alters *Arabidopsis* transcripts involved in cell wall modification, abiotic stress, nitrogen metabolism, and plant defense. *Physiologia Plantarum* 125: 114-126

Bae H, Herman E, Sicher R (2005b) Exogenous trehalose promotes non-structural carbohydrate accumulation and induces chemical detoxification and stress response proteins in *Arabidopsis thaliana* grown in liquid culture. *Plant Science* 168: 1293-1301

Ballaré CL (2011) Jasmonate-induced defenses: A tale of intelligence, collaborators and rascals. *Trends in Plant Science* 16: 249-257

Beevers L, Hageman RH (1969) Nitrate reduction in higher plants. *Annual Review of Plant Physiology* 20: 495-522

Bergelson J, Purrington CB (1996) Surveying patterns in the cost of resistance in plants. *The American Naturalist* 148: 536-558

Bernard SM, Habash DZ (2009) The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytologist* 182: 608-620

Bernards MA, Båstrup-Spohr L (2008) Phenylpropanoid metabolism induced by wounding and insect herbivory. In: Schaller A (ed) *Induced Plant Resistance to Herbivory*. Springer Netherlands, pp 189-211

Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* 19: 148-162

Best M, Koenig K, McDonald K, Schueller M, Rogers A, Ferrieri RA (2011) Inhibition of trehalose breakdown increases new carbon partitioning into cellulosic biomass in *Nicotiana tabacum*. *Carbohydrate Research* 346: 595-601

Blount JW, Korth KL, Masoud SA, Rasmussen S, Lamb C, Dixon RA (2000) Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiology* 122: 107-116

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding *in vivo* benzenoid metabolism in Petunia petal tissue. *Plant Physiology* 135: 1993-2011

Boeckler GA, Gershenzon J, Unsicker SB (2011) Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry* 72: 1497-1509

Bogs J, Downey MO, Harvey JS, Ashton AR, Tanner GJ, Robinson SP (2005) Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. *Plant Physiology* 139: 652-663

Bolwell GP, Mavandad M, Millar DJ, Edwards KJ, Schuch W, Dixon RA (1988) Inhibition of mRNA levels and activities by *trans*-cinnamic acid in elicitor-induced bean cells. *Phytochemistry* 27: 2109-2117

Bornman J, Reuber S, Cen Y-P, Weissenbock G (1997) Ultraviolet radiation as a stress factor and the role of protective pigments. In: Lumsden P (ed) *Plants and UV-B: Responses to Environmental Change*. Cambridge University Press, Cambridge, United Kingdom, pp 157-168

Bowles D, Lim E, Poppenberger B, Vaistij F (2006) Glycosyltransferases of lipophilic small molecules. *Annual Review of Plant Biology* 57: 567-597

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248-254

Bryant JP, Chapin FS, III, Klein DR (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos* 40: 357-368

Cai C, Zhao X-Q, Zhu Y-G, Li B, Tong Y-P, Li Z-S (2007) Regulation of the high-affinity nitrate transport system in wheat roots by exogenous abscisic acid and glutamine. *Journal of Integrative Plant Biology* 49: 1719-1725

Camm EL, Towers GHN (1973) Phenylalanine ammonia lyase. *Phytochemistry* 12: 961-973

Cantón FR, Suárez MF, Cánovas FM (2005) Molecular aspects of nitrogen mobilization and recycling in trees. *Photosynthesis Research* 83: 265-278

Castro-Rodriguez V, Garcia-Gutierrez A, Canales J, Avila C, Kirby EG, Canovas FM (2011) The glutamine synthetase gene family in *Populus*. *BMC Plant Biology* 11

Chapple CCS, Walker MA, Ellis BE (1986) Plant tyrosine decarboxylase can be strongly inhibited by L-α-aminooxy-β-phenylpropionate. *Planta* 167: 101-105

Chen F, Liu C-J, Tschaplinski TJ, Zhao N (2009) Genomics of secondary metabolism in *Populus*: Interactions with biotic and abiotic environments. *Critical Reviews in Plant Sciences* 28: 375 - 392

Colón AM, Sengupta N, Rhodes D, Dudareva N, Morgan J (2010) A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of *Petunia hybrida*. *The Plant Journal* 62: 64-76

D'Onofrio C, Cox A, Davies C, Boss PK (2009) Induction of secondary metabolism in grape cell cultures by jasmonates. *Functional Plant Biology* 36: 323-338

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell* 7: 1085-1097

Dixon RA, Xie D-Y, Sharma SB (2005) Proanthocyanidins – a final frontier in flavonoid research? *New Phytologist* 165: 9-28

Donaldson J, Stevens M, Barnhill H, Lindroth R (2006) Age-related shifts in leaf chemistry of clonal aspen (*Populus tremuloides*). *Journal of Chemical Ecology* 32: 1415-1429

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) AgriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Research* 38: W64-W70

Eberhard J, Ehrler TT, Epple P, Felix G, Raesecke H-R, Amrhein N, Schmid J (1996) Cytosolic and plastidic chorismate mutase isozymes from *Arabidopsis thaliana*: molecular characterization and enzymatic properties. *The Plant Journal* 10: 815-821

Edwards R, Mavandad M, Dixon RA (1990) Metabolic fate of cinnamic acid in elicitor treated cell suspension cultures of *Phaseolus vulgaris*. *Phytochemistry* 29: 1867-1873

Endress R (1981) Conversion of phenylalanine into tyrosine by *Portulaca* callus. *Plant Physiology* 68: 272-274

Fatland BL, Ke J, Anderson MD, Mentzen WI, Cui LW, Allred CC, Johnston JL, Nikolau BJ, Wurtele ES (2002) Molecular characterization of a heteromeric ATP-citrate lyase that generates cytosolic acetyl-coenzyme A in *Arabidopsis*. *Plant Physiology* 130: 740-756

Fatland BL, Nikolau BJ, Wurtele ES (2005) Reverse genetic characterization of cytosolic acetyl-CoA generation by ATP-citrate lyase in *Arabidopsis*. *Plant Cell* 17: 182-203

Fraissinet-Tachet L, Baltz R, Chong J, Kauffmann S, Fritig B, Saindrenan P (1998) Two tobacco genes induced by infection, elicitor and salicylic acid encode glucosyltransferases acting on phenylpropanoids and benzoic acid derivatives, including salicylic acid. *FEBS Letters* 437: 319-323

Fritz C, Palacios-Rojas N, Feil R, Stitt M (2006) Regulation of secondary metabolism by the carbon–nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism. *The Plant Journal* 46: 533-548

Funk C, Brodelius PE (1990) Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia* Andr.: II. Effects of precursor feeding and metabolic inhibitors. *Plant Physiology* 94: 95-101

Gagné S, Lacampagne S, Claisse O, Gény L (2009) Leucoanthocyanidin reductase and anthocyanidin reductase gene expression and activity in flowers, young berries and skins of *Vitis vinifera* L. cv. Cabernet-Sauvignon during development. *Plant Physiology and Biochemistry* 47: 282-290

Goossens A, Häkkinen ST, Laakso I, Seppänen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, So¨derlund H, Zabeau M, Inzé D, Oksman-Caldentey K-M (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proceedings of the National Academy of Sciences* 100: 8595-8600

Greenaway W, English S, May J, Whatley FR (1991) Chemotaxonomy of section *Leuce* poplars by GC-MS of bud exudate. *Biochemical Systematics and Ecology* 19: 507-518

Greenaway W, English S, Whatley FR (1992) Relationships of *Populus* x *acuminata* and *Populus* x *generosa* with their parental species examined by gas chromatography - mass spectrometry of bud exudates. *Canadian Journal of Botany* 70: 212-221

Griesser M, Hoffmann T, Bellido ML, Rosati C, Fink B, Kurtzer R, Aharoni A, Muñoz-Blanco J, Schwab W (2008) Redirection of flavonoid biosynthesis through the down-regulation of an anthocyanidin glucosyltransferase in ripening strawberry fruit. *Plant Physiology* 146: 1528-1539

Gundlach H, Müller MJ, Kutchan TM, Zenk MH (1992) Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Proceedings of the National Academy of Sciences* 89: 2389-2393

Hagerman AE, Robbins CT, Weerasuriya Y, Wilson TC, McArthur C (1992) Tannin chemistry in relation to digestion. *Journal of Range Management* 45: 57-62

Hahlbrock K, Grisebach H (1979) Enzymic controls in the biosynthesis of lignin and flavonoids. *Annual Review of Plant Physiology* 30: 105-130

Hanik N, Gómez S, Best M, Schueller M, Orians C, Ferrieri R (2010) Partitioning of new carbon as ^{11}C in *Nicotiana tabacum* reveals insight into methyl jasmonate induced changes in metabolism. *Journal of Chemical Ecology* 36: 1058-1067

Harborne JB, Corner JJ (1961) Plant polyphenols. 4. Hydroxycinnamic acid-sugar derivatives. *Biochemistry Journal* 81: 242-250

Harding SA, Jarvie MM, Lindroth RL, Tsai C-J (2009) A comparative analysis of phenylpropanoid metabolism, N utilization, and carbon partitioning in fast- and slow-growing *Populus* hybrid clones. *Journal of Experimental Botany* 60: 3443-3452

Harding SA, Jiang H, Jeong ML, Casado FL, Lin H-W, Tsai C-J (2005) Functional genomics analysis of foliar condensed tannin and phenolic glycoside regulation in natural cottonwood hybrids. *Tree Physiology* 25: 1475-1486

Havir EA (1981) Modification of L-phenylalanine ammonia-lyase in soybean cell suspension cultures by 2-aminooxyacetate and L-2-aminooxy-3-phenylpropionate. *Planta* 152: 124-130

Hemming JDC, Lindroth RL (1995) Intraspecific variation in aspen phytochemistry: Effects on performance of gypsy moths and forest tent caterpillars. *Oecologia* 103: 79-88

Herrmann K (1995) The shikimate pathway: Early steps in the biosynthesis of aromatic compounds. *Plant Cell* 7: 907-919

Herrmann K, Weaver L (1999) The shikimate pathway. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 473-503

Hey SJ, Byrne E, Halford NG (2010) The interface between metabolic and stress signalling. *Annals of Botany* 105: 197-203

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* 16: 1446-1465

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry* 278: 95-103

Hsieh L-S, Yeh C-S, Pan H-C, Cheng C-Y, Yang C-C, Lee P-D (2010) Cloning and expression of a phenylalanine ammonia-lyase gene (*BoPAL2*) from *Bambusa oldhamii* in *Escherichia coli* and *Pichia pastoris*. *Protein Expression and Purification* 71: 224-230

Hsu JC (1981) Simultaneous confidence intervals for all distances from the "best". *The Annals of Statistics* 9: 1026-1034

Hu W-J, Kawaoka A, Tsai C-J, Lung J, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). *Proceedings of the National Academy of Sciences* 95: 5407-5412

Huang X, Stettmaier K, Michel C, Hutzler P, Mueller M, Durner J (2004) Nitric oxide is induced by wounding and influences jasmonic acid signaling in *Arabidopsis thaliana*. *Planta* 218: 938-946

Jones CG, Hartley SE (1999) A protein competition model of phenolic allocation. *Oikos* 86: 27-44

Joshi V, Joung J-G, Fei Z, Jander G (2010) Interdependence of threonine, methionine and isoleucine metabolism in plants: Accumulation and transcriptional regulation under abiotic stress. *Amino Acids* 39: 933-947

Julkunen-Tiitto R (1986) A chemotaxonomic survey of phenolics in leaves of northern Salicaceae species. *Phytochemistry* 25: 663-667

Kao Y-Y, Harding SA, Tsai C-J (2002) Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. *Plant Physiology* 130: 796-807

Keinänen M, Julkunen-Tiitto R, Mutikainen P, Walls M, Ovaska J, Vapaavuori E (1999) Trade-offs in phenolic metabolism of silver birch: Effects of fertilization, defoliation, and genotype. *Ecology* 80: 1970-1986

Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. *Nature* 389: 33-39

Kessler A, Baldwin IT (2002) Plant responses to insect herbivory: The emerging molecular analysis. *Annual Review of Plant Biology* 53: 299-328

Kittipongpatana N, Tangjai T, Srimanee A, Kittipongpatana OS (2007) Biotransformation of hydroquinone to arbutin by cell-suspension cultures of three Thai solanaceous plants. *Chiang Mai University Journal of Natural Sciences* 6: 219-229

Kleiner KW, Raffa KF, Dickson RE (1999) Partitioning of ^{14}C-labeled photosynthate to allelochemicals and primary metabolites in source and sink leaves of aspen: Evidence for secondary metabolite turnover. *Oecologia* 119: 408-418

Knudsen JT, Gershenzon J (2006) The chemical diversity of floral scent. In: Dudareva N, Pichersky E (eds) *Biology of Floral Scient.* CRC Press, Boca Raton, FL, pp 27-52

Koo AJK, Cooke TF, Howe GA (2011) Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. *Proceedings of the National Academy of Sciences* 108: 9298-9303

Koricheva J, Larsson S, Haukioja E, Keinänen M (1998) Regulation of woody plant secondary metabolism by resource availability: Hypothesis testing by means of meta-analysis. *Oikos* 83: 212-226

Lamotte O, Courtois C, Barnavon L, Pugin A, Wendehenne D (2005) Nitric oxide in plants: The biosynthesis and cell signalling properties of a fascinating molecule. *Planta* 221: 1-4

Lee Y, Yoon H, Paik Y, Liu J, Chung W-i, Choi G (2005) Reciprocal regulation of *Arabidopsis* UGT78D2 and BANYULS is critical for regulation of the metabolic flux of anthocyanidins to condensed tannins in developing seed coats. *Journal of Plant Biology* 48: 356-370

Letendre CH, Dickens G, Guroff G (1975) Phenylalanine hydroxylase from *Pseudomonas sp.* (ATCC 11299a). Purification, molecular weight, and influence of tyrosine metabolites on activation and hydroxylation. *Journal of Biological Chemistry* 250: 6672-6678

Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X (2007) Crystal structure of *Medicago truncatula* UGT85H2 – insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *Journal of Molecular Biology* 370: 951-963

Lim E-K, Bowles DJ (2004) A class of plant glycosyltransferases involved in cellular homeostasis. *EMBO Journal* 23: 2915-2922

Lim E-K, Doucet CJ, Li Y, Elias L, Worrall D, Spencer SP, Ross J, Bowles DJ (2002) The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *Journal of Biological Chemistry* 277: 586-592

Lindroth RL, Hwang S-Y (1996) Clonal variation in foliar chemistry of quaking aspen (*Populus tremuloides* Michx.). *Biochemical Systematics and Ecology* 24: 357-364

Louie GV, Bowman ME, Moffitt Michelle C, Baiga TJ, Moore Bradley S, Noel JP (2006) Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases. *Chemistry & Biology* 13: 1327-1338

Lu S, Zhou Y, Li L, Chiang VL (2006) Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. *Plant and Cell Physiology* 47: 905-914

Lu Y, Yeap Foo L (2000) Flavonoid and phenolic glycosides from *Salvia officinalis*. *Phytochemistry* 55: 263-267

Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y, Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C (2007) Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *The Plant Journal* 50: 678-695

Lutterbach R, Stöckigt J (1994) *In vivo* investigation of plant-cell metabolism by means of natural-abundance ^{13}C-NMR spectroscopy. *Helvetica Chimica Acta* 77: 2153-2161

Maeda H, Shasany AK, Schnepp J, Orlova I, Taguchi G, Cooper BR, Rhodes D, Pichersky E, Dudareva N (2010) RNAi suppression of *arogenate dehydratase1* reveals that phenylalanine is synthesized predominantly via the arogenate pathway in Petunia petals. *Plant Cell* 22: 832-849

Marques IA, Brodelius PE (1988) Elicitor-induced L-tyrosine decarboxylase from plant cell suspension cultures. II. Partial characterization. *Plant Physiology* 88: 52-55

Martens S, Preuß A, Matern U (2010) Multifunctional flavonoid dioxygenases: Flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* 71: 1040-1049

Matt P, Krapp A, Haake V, Mock H-P, Stitt M (2002) Decreased Rubisco activity leads to dramatic changes of nitrate metabolism, amino acid metabolism and the levels of phenylpropanoids and nicotine in tobacco antisense *RBCS* transformants. *The Plant Journal* 30: 663-677

Mavandad M, Edwards R, Liang X, Lamb CJ, Dixon RA (1990) Effects of *trans*-cinnamic acid on expression of the bean phenylalanine ammonia-lyase gene family. *Plant Physiology* 94: 671-680

Mellway RD, Tran LT, Prouse MB, Campbell MM, Constabel CP (2009) The wound-, pathogen-, and ultraviolet B-responsive *MYB134* gene encodes an R2R3 MYB transcription factor that regulates proanthocyanidin synthesis in Poplar. *Plant Physiology* 150: 924-941

Miersch O, Neumerkel J, Dippe M, Stenzel I, Wasternack C (2008) Hydroxylated jasmonates are commonly occurring metabolites of jasmonic acid and contribute to a partial switch-off in jasmonate signaling. *New Phytologist* 177: 114-127

Miller R, Owens SJ, Rørslett B (2011) Plants and colour: Flowers and pollination. *Optics & Laser Technology* 43: 282-294

Nagels L, Molderez M, Parmentier F (1981) UDPG: *p*-Coumarate glucosyltransferase activity in enzyme extracts from higher plants. *Phytochemistry* 20: 965-967

Nair PM, Vining LC (1965) Phenylalanine hydroxylase from spinach leaves. *Phytochemistry* 4: 401-411

Orlova I, Marshall-Colon A, Schnepp J, Wood B, Varbanova M, Fridman E, Blakeslee JJ, Peer WA, Murphy AS, Rhodes D, Pichersky E, Dudareva N (2006) Reduction of benzenoid synthesis in Petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *Plant Cell* 18: 3458-3475

Orr JD, Edwards R, Dixon RA (1993) Stress responses in Alfalfa (*Medicago sativa* L.) (XIV. Changes in the levels of phenylpropanoid pathway intermediates in relation to regulation of L-phenylalanine ammonia-lyase in elicitor-treated cell-suspension cultures). *Plant Physiology* 101: 847-856

Paquette S, Møller BL, Bak S (2003) On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62: 399-413

Pauwels L, Morreel K, De Witte E, Lammertyn F, Van Montagu M, Boerjan W, Inzé D, Goossens A (2008) Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured *Arabidopsis* cells. *Proceedings of the National Academy of Sciences* 105: 1380-1385

Payyavula R, Babst B, Nelsen M, Harding S, Tsai C-J (2009) Glycosylation-mediated phenylpropanoid partitioning in *Populus tremuloides* cell cultures. *BMC Plant Biology* 9: 151

Payyavula RS, Tay KHC, Tsai C-J, Harding SA (2011) The sucrose transporter family in *Populus*: The importance of a tonoplast PtaSUT4 to biomass and carbon partitioning. *The Plant Journal* 65: 757-770

Peters DJ, Constabel CP (2002) Molecular analysis of herbivore-induced condensed tannin synthesis: Cloning and expression of dihydroflavonol reductase from trembling aspen (*Populus tremuloides*). *The Plant Journal* 32: 701-712

Porter LJ, Hrstich LN, Chan BG (1986) The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin. *Phytochemistry* 25: 223-230

Pribat A, Noiriel A, Morse AM, Davis JM, Fouquet R, Loizeau K, Ravanel S, Frank W, Haas R, Reski R, Bedair M, Sumner LW, Hanson AD (2010) Nonflowering plants possess a unique folate-dependent phenylalanine hydroxylase that Is localized in chloroplasts. *Plant Cell* 22: 3410-3422

Qualley AV, Cooper BR, Dudareva N (2012) Profiling hydroxycinnamoyl-coenzyme A thioesters: Unlocking the back door of phenylpropanoid metabolism. *Analytical Biochemistry* 420: 182-184

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* 339: 62-66

Rasmussen S, Dixon RA (1999) Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. *Plant Cell* 11: 1537-1552

Razal RA, Ellis S, Singh S, Lewis NG, Towers GHN (1996) Nitrogen recycling in phenylpropanoid metabolism. *Phytochemistry* 41: 31-35

Rea PA (2007) Plant ATP-binding cassette transporters. *Annual Review of Plant Biology* 58: 347-375

Reed J (1986) Relationships among soluble phenolics, insoluble proanthocyanidins and fiber in East African browse species. *Journal of Range Management* 39: 5-7

Rehill B, Whitham T, Martinsen G, Schweitzer J, Bailey J, Lindroth R (2006) Developmental trajectories in cottonwood phytochemistry. *Journal of Chemical Ecology* 32: 2269-2285

Reichardt PB, Chapin FS, Bryant JP, Mattes BR, Clausen TP (1991) Carbon/nutrient balance as a predictor of plant defense in Alaskan balsam poplar: Potential importance of metabolite turnover. *Oecologia* 88: 401-406

Reymond P, Farmer EE (1998) Jasmonate and salicylate as global signals for defense gene expression. *Current Opinion in Plant Biology* 1: 404-411

Rohde A, Morreel K, Ralph J, Goeminne G, Hostyn V, De Rycke R, Kushnir S, Van Doorsselaere J, Joseleau J-P, Vuylsteke M, Van Driessche G, Van Beeumen J, Messens E, Boerjan W (2004) Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell* 16: 2749-2771

Rosler J, Krekel F, Amrhein N, Schmid J (1997) Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. *Plant Physiology* 113: 175-179

Roslin T, Salminen J-P (2008) Specialization pays off: Contrasting effects of two types of tannins on oak specialist and generalist moth species. *Oikos* 117: 1560-1568

Schalk M, Cabello-Hurtado F, Pierrel MA, Atanossova R, Saindrenan P, Werck-Reichhart D (1998) Piperonylic acid, a selective, mechanism-based inactivator of the *trans*-cinnamate 4-hydroxylase: A new tool to control the flux of metabolites in the phenylpropanoid pathway. *Plant Physiology* 118: 209-218

Schoch G, Nikov G, Alworth W, Werck-Reichhart D (2002) Chemical inactivation of the cinnamate 4-hydroxylase allows for the accumulation of salicylic acid in elicited cells. *Plant Physiology* 130: 1022-1031

Scioneaux A, Schmidt M, Moore M, Lindroth R, Wooley S, Hagerman A (2011) Qualitative variation in proanthocyanidin composition of *Populus* species and hybrids: Genetics is the key. *Journal of Chemical Ecology* 37: 57-70

Scriver CR, Clow CL (1980) Phenylketonuria and other phenylalanine hydroxylation mutants in man. *Annual Review of Genetics* 14: 179-202

Seto Y, Hamada S, Matsuura H, Matsushige M, Satou C, Takahashi K, Masuta C, Ito H, Matsui H, Nabeta K (2009) Purification and cDNA cloning of a wound inducible glucosyltransferase active toward 12-hydroxy jasmonic acid. *Phytochemistry* 70: 370-379

Simko I, Omer EA, Ewing EE, McMurry S, Koch JL, Davies PJ (1996) Tuberonic (12-OH-jasmonic) acid glucoside and its methyl ester in potato. *Phytochemistry* 43: 727-730

Sircar D, Mitra A (2009) Accumulation of *p*-hydroxybenzoic acid in hairy roots of *Daucus carota* 2: Confirming biosynthetic steps through feeding of inhibitors and precursors. *Journal of Plant Physiology* 166: 1370-1380

Soeno K, Goda H, Ishii T, Ogura T, Tachikawa T, Sasaki E, Yoshida S, Fujioka S, Asami T, Shimada Y (2010) Auxin biosynthesis inhibitors, identified by a genomics-based approach, provide insights into auxin biosynthesis. *Plant and Cell Physiology* 51: 524-536

Stafford HA (1988) Proanthocyanidins and the lignin connection. *Phytochemistry* 27: 1-6

Strauss SY, Rudgers JA, Lau JA, Irwin RE (2002) Direct and ecological costs of resistance to herbivory. *Trends in Ecology & Evolution* 17: 278-285

Sun J-Q, Jiang H-L, Li C-Y (2011) Systemin/Jasmonate-mediated systemic defense signaling in tomato. *Molecular Plant* 4: 607-615

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* 6: e21800

Suzuki H, Hayase H, Nakayama A, Yamaguchi I, Asami T, Nakajima M (2007) Identification and characterization of an *Ipomoea nil* glucosyltransferase which metabolizes some phytohormones. *Biochemical and Biophysical Research Communications* 361: 980-986

Suzuki H, Reddy M, Naoumkina M, Aziz N, May G, Huhman D, Sumner L, Blount J, Mendes P, Dixon R (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. *Planta* 220: 696-707

Swain T (1979) Tannins and lignins. In: Rosenthal G, Jansen D (eds) *Herbivores: their interaction with secondary plant metabolites*. Academic Press, New York, NY, pp 657-682

Sweetlove LJ, Beard KFM, Nunes-Nesi A, Fernie AR, Ratcliffe RG (2010) Not just a circle: Flux modes in the plant TCA cycle. *Trends in Plant Science* 15: 462-470

Sykes R, Yung M, Novaes E, Kirst M, Peter G, Davis M (2010) High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. In: Mielenz JR (ed) *Biofuels*. Humana Press, pp 169-183

Tahvanainen J, Helle E, Julkunen-Tiitto R, Lavola A (1985) Phenolic compounds of willow bark as deterrents against feeding by mountain hare. *Oecologia* 65: 319-323

Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR (2003) Proanthocyanidin biosynthesis in plants. *Journal of Biological Chemistry* 278: 31647-31656

Tanner GJ, Kristiansen KN (1993) Synthesis of 3,4-*cis*-[^{3}H]leucocyanidin and enzymatic reduction to catechin. *Analytical Biochemistry* 209: 274-277

Thornton B (2004) Inhibition of nitrate influx by glutamine in *Lolium perenne* depends upon the contribution of the HATS to the total influx. *Journal of Experimental Botany* 55: 761-769

Tian S, Nakamura K, Cui T, Kayahara H (2005) High-performance liquid chromatographic determination of phenolic compounds in rice. *Journal of Chromatography A* 1063: 121-128

Tsai C-J, El Kayal W, Harding SA (2006a) *Populus*, the new model system for investigating phenylpropanoid complexity. *International Journal of Applied Science and Engineering* 4: 221-233

Tsai C-J, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006b) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytologist* 172: 47-62

Tsai C, Cseke L, Harding S (2003) Isolation and purification of RNA. In: Cseke L, Kaufman P, Podila G, Tsai C (eds) *Handbook of Molecular and Cellular Methods in Biology and Medicine*. CRC Press, Boca Raton, FL, pp 25-44

Tsai CJ, Ranjan P, DiFazio SP, Tuskan GA, Johnson V (2011) Poplar genome microarrays. In: Joshi CP, DiFazio SP, Kole C (eds) *Genetics, Genomics and Breeding of Crop Plants: Poplar*. Science Publishers, Enfield, NH, pp 112-127

Tzin V, Galili G (2010) New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Molecular Plant* 3: 956-972

Vaistij FE, Lim E-K, Edwards R, Bowles DJ (2009) Glycosylation of secondary metabolites and xenobiotics. In: Osbourn AE, Lanzotti V (eds) *Plant-derived Natural Products: Synthesis, Function, and Application*. Springer, New York, NY, pp 209-228

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. *Plant Physiology* 153: 895-905

Vidmar JJ, Zhuo D, Siddiqi MY, Schjoerring JK, Touraine B, Glass ADM (2000) Regulation of high-affinity nitrate transporter genes and high-affinity nitrate influx by nitrogen pools in roots of barley. *Plant Physiology* 123: 307-318

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: Characterization of a supergene family. *Trends in Plant Science* 5: 380-386

Wang H-Q, Tuominen LK, Tsai C-J (2011) SLIM: A sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* 27: 225-231

Watts KT, Mijts BN, Lee PC, Manning AJ, Schmidt-Dannert C (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chemistry & Biology* 13: 1317-1326

Wink M, R.A. Leigh DS, Callow JA (1997) Compartmentation of secondary metabolites and xenobiotics in plant vacuoles. *Advances in Botanical Research*. Academic Press, pp 141-169

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* 126: 485-493

Wu J, Zhang S, Huang J, Xiao Q, Li Q, Long L, Huang L (2003) New aliphatic alcohol and (*Z*)-4-coumaric acid glycosides from *Acanthus ilicifolius*. *Chemical & Pharmaceutical Bulletin (Tokyo)* 51: 1201-1203

Wünsche H, Baldwin IT, Wu J (2011) S-Nitrosoglutathione reductase (GSNOR) mediates the biosynthesis of jasmonic acid and ethylene induced by feeding of the insect herbivore *Manduca sexta* and is important for jasmonate-elicited responses in *Nicotiana attenuata*. *Journal of Experimental Botany* 62: 4605-4616

Xiao Y, Gao S, Di P, Chen J, Chen W, Zhang L (2009) Methyl jasmonate dramatically enhances the accumulation of phenolic acids in *Salvia miltiorrhiza* hairy root cultures. *Physiologia Plantarum* 137: 1-9

Xie D-Y, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. *Science* 299: 396-399

Yamada T, Matsuda F, Kasai K, Fukuoka S, Kitamura K, Tozawa Y, Miyagawa H, Wakasa K (2008) Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan. *Plant Cell* 20: 1316-1329

Yamamoto K-i, Kobayashi N, Yoshitama K, Teramoto S, Komamine A (2001) Isolation and purification of tyrosine hydroxylase from callus cultures of *Portulaca grandiflora*. *Plant and Cell Physiology* 42: 969-975

Yan C-Y, Yu R-M, Zhang Z, Kong L-Y (2007) Biotransformation of 4-hydroxybenzen derivatives by hairy root cultures of *Polygonum multiflorum* Thunb. *Journal of Integrative Plant Biology* 49: 207-212

Yazaki K (2005) Transporters of secondary metabolites. *Current Opinion in Plant Biology* 8: 301-307

Yin R, Messner B, Faus-Kessler T, Hoffmann T, Schwab W, Hajirezaei M-R, von Saint Paul V, Heller W, Schäffner AR (2012) Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-*O*-glycosylation. *Journal of Experimental Botany* Advance Access Publication available at doi: 10.1093/jxb/err416

Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *The Plant Journal* 66: 182-193

Yoshihara T, Omer EA, Koshino H, Sakamura S, Kikuta Y, Koda Y (1989) Structure of a tuber-inducing stimulus from potato leaves (*Solanum tuberosum* L.). *Agricultural and Biological Chemistry* 53: 2835-2837

Zenk MH (1967) Pathways of salicyl alcohol and salicin formation in *Salix purpurea* L. *Phytochemistry* 6: 245-252

Zhao J, Davis LC, Verpoorte R (2005) Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnology Advances* 23: 283-333

CHAPTER 4.

COTTONWOODS AND CHESTNUTS: ECOSYSTEM MODELING AS A SCIENTIFIC TOOL TO HELP ADDRESS PUBLIC CONCERNS SURROUNDING THE FIELD RELEASE OF GENETICALLY MODIFIED TREES

Summary

The scientific underpinnings and impacts of biotechnology are well-understood from the perspectives of molecular biology, physiology, and population genetics; however, the use of this technology in a field context remains socially controversial. A commonly cited concern of individuals opposing the use of transgenic organisms is the difficulty of predicting indirect environmental effects of field releases *a priori*, a concern that is difficult to counter using traditional, reductionistic experimental approaches. Nevertheless, scientists would do well to work towards methods of understanding likely follow-on impacts of particular transgenic traits in order to help gain public acceptance. I provide a proof of concept for the use of systems-based ecological modeling as a risk assessment tool for trait-based, indirect ecological effects of transgenic organisms. In particular, I consider three different scenarios involving the use of transgenic trees, each embodying a different level of concern from an environmental ethics perspective. Generating a conceptual ecosystem model for each scenario as well as simulation results from one of these models, I give evidence for how a holistic approach can allow users to

"see" indirect effects and generate new hypotheses for field research. While this method requires additional development and does not yield an all-encompassing determination of the risk factors involved in transgenic field releases, it provides an additional framework from which scientists can help to address public concerns regarding biotechnology.

Background

The widespread use of genetically modified organisms has been a globally contentious issue over the past two decades, for a variety of reasons. Individuals unfamiliar with the technologies involved have expressed concerns over the integrity and safety of genetically transformed organisms, exemplified in their use of terms such as "Frankenfoods" (Hellsten 2003). Meanwhile, as scientists validly use research data to address criticisms that genetically modified plants are inherently dangerous to eat (see, for example, Lemaux 2008), they may gloss over issues associated with biotechnology that are not exclusively scientific, such as whether individuals have a right to choose what sorts of foods they put into their bodies (Franken 2000). Scientific defenses in favor of biotechnology that focus exclusively on technical issues can lead to skepticism on the part of non-specialists (de Melo-Martin and Meghani 2008), particularly when scientists themselves do not speak with one voice regarding technical, ecological, and social concerns or uncertainties inherent to new advances (Myhr 2010). Documented events such as transgene flow into wild plant populations (Kwit et al. 2011; Zapiola et al. 2008), increasing pesticide use due to burgeoning populations of non-target pests in BT maize (Lu et al. 2010), corporate legal actions on farmers who re-use patented seed without

paying licensing fees (Monsanto v. Schmeiser 2004; Monsanto v. Bowman 2009), and NGO legal actions against US government efforts to deregulate specific transgenic crops (Geertson Seed Farms v. Johanns 2008; Center for Food Safety v. Vilsack 2011) are a few examples of the validity such concerns. Far from existing as a purely scientific issue, the implementation of biotechnology in an open-field context is fraught with both technical and value-oriented questions that scientists and non-specialists, businesses, governments, and non-governmental organizations around the world have been struggling to answer for decades (see also Cox 2008; Gavora and Lister 1989; Merges 1987-1988; Myhr 2010; National Research Council 2008; Regal 1994; Schmidt 2008; Schurman 2004; Selig 2008 for detailed treatments of a variety of contentious issues surrounding biotechnology) .

As the originators of the technologies allowing novel forms of genetic modification, scientists have an important role to play in developing solutions to questions provoked by the implementation of biotechnology. Among professional scientists, scientific ethics are conventionally presented as a form of professional ethics. Researchers are expected as members of the scientific profession to maintain honesty in regards to their data and how they are represented, ensure that the research one carries out does not cross into the realm of unnecessary harm (or at least provides a reasonable amount of benefit for the harm done, in the case of research involving humans or other vertebrates), and to disclose competing interests when publishing the results of one's work. However, viewing scientific ethics as equivalent to professional ethics can be seen as somewhat myopic. Scientists, like any other group of working people, are not merely professionals but also citizens and (particularly relevant for biologists) members of the

living world. Some have argued that research is a moral responsibility when uncertainties are significant (Tannert et al. 2007), and this responsibility seems even more keen when the uncertainties arise from one's own work. When contextualized in this manner, it seems reasonable to suggest that on some level, biotechnologists are not *only* responsible for conducting their research in ethical ways, but that they also must shoulder some level of responsibility for the implications of that research. To provide an analogy, adults are typically considered responsible for their offspring, at least for a particular period of time in those individuals' youth. Similarly to how children are treated by their parents, scientists actively take responsibility for their new hypotheses and technologies as advocates. Without doing so, it would be difficult to maintain funding for their work. However, taking responsibility for one's work should not be limited to advocacy or the envisioning of best-case scenarios. Like parents, scientists developing new technologies also benefit their intellectual offspring in the long run if they are able to envision the newcomer's possible excesses and rein them in. This stewardship role can benefit the flourishing of the technology and the broader community alike by helping to figure out how to tip the balance of potential harms versus potential benefits in favor of the latter as strongly as possible.

It is with this view in mind that I have developed the goal of helping to address public concerns surrounding the ecological effects of field release of genetically modified plants using a scientific approach. A critical and valid concern expressed by non-scientists and ecologists alike has been the possibility of indirect environmental effects arising from the use of transgenic organisms (National Research Council 2008; Wagner et al. 2001). Non-specialists have expressed this concern by arguing that biotechnologists

are “tinkering with nature” or “playing God,” without fully understanding the follow-on impacts of their work in an environmental context (Wagner et al. 2001). Although such assertions may rankle, they are not entirely ill-placed, since scientists are typically expected to specialize in a relatively narrow area of research. Therefore, as a student of tree functional genomics, I choose in this work to take the arguments of such critics seriously by moving beyond the expected boundaries of my field and attempting to comprehend how the living beings potentially generated from my work are likely to interact with other parts of the environment, including other living organisms. A major challenge to developing such an understanding is the insufficiency of direct interactions for explaining of the overall effects of an organism in a particular ecosystem (see Higashi and Patten 1989 for one attempt; reviewed by Strauss 1991). Predicting indirect effects can be difficult using traditional field-based methods, in part because science usually follows a reductionistic approach to experimentation, and assumptions that indirectness is unimportant can have potentially drastic conservation consequences (Bergstrom et al. 2009). My work therefore requires the capacity to account for and predict not only direct interactions, but also indirect ones.

In this chapter, I will outline how applying systems ecology to specific ecological scenarios can help respond to concerns about the difficult-to-predict effects of biotechnology’s use in the field. I will integrate environmental ethics principles with systems ecology tools to draw conceptual sketches and ecosystem models examining environmental impacts of three different scenarios involving the release of transgenic trees. These scenarios differ in their level of environmental ethics concern, demonstrating the flexibility of the methods while highlighting the idea that science is

necessary, but not sufficient, to make ethical judgments about the implementation of biotechnology (de Melo-Martin and Meghani 2008). Finally, I will show results from initial simulations for one model, the results from which suggest a possible hypothesis for field research.

Methodology

Scenario One: Transgenic Escape

The first scenario I consider is an unmanaged forest ecosystem into which a metabolically engineered line of *Populus* has escaped. Transgenic escape is a documented phenomenon among plants in agricultural use. Escapes may occur due to cross-fertilization with non-transgenic plants of the same species in other agricultural fields, cross-hybridization with wild relatives, or simply volunteer growth of seeds that were lost during harvest (Ellstrand 2003; Lu and Snow 2005; Smyth et al. 2002). Current APHIS regulations stipulate the need for "barrier" zones when regulated transgenic plants are grown in the field to avoid pollen contamination of non-transgenic genotypes (7 CFR 340 Strauss et al. 2010; USGPO 2010); however, pollen drift from trees has been documented to occur for hundreds of kilometers (Williams 2010). This clearly increases the likelihood of transgenic escape for trees *Populus*, a genus which also readily hybridizes, clonally reproduces, and contains species having some of the widest native ranges of all trees (Hoenicka and Fladung 2006; Strauss et al. 2004). Thus, non-specialists concerned with the use of biotechnology would be likely to have strong reservations about the use of metabolically engineered *Populus*, even if documented gene

flow risks are relatively low (DiFazio et al. 2004; Lu et al. 2006). In order to meet APHIS standards, transgenic *Populus* would not be grown near other species of *Populus* with which it is known to hybridize. In this scenario, I therefore include wild *Populus* in the unmanaged ecosystem, but assume that hybridization does not occur. In other words, escape is assumed to occur via drift of volunteer seed from a distant plantation. Current APHIS regulations do not permit transgenic trees to grow to reproductive age in the field for initial testing (7 CFR 340 Strauss et al. 2010; USGPO 2010), making this scenario less realistic if the transgenic genotype has not yet been deregulated.

Scenario Two: Plantation/Unmanaged Forest Interface

The second scenario increases both realism and complexity in that it considers the interface between an unmanaged forest and a poplar plantation containing a mix of wild type and metabolically engineered trees that do not hybridize. The unmanaged forest also contains the same wild type *Populus* species which cannot hybridize with the transgenic line. Transgenic escape is assumed not to occur, although drift of leaf litter may occur in both directions between the plantation and the forest. This scenario is clearly more complex than the first in that the ecosystem is divided into two discrete subsections, which partially interact with each other. It is also more realistic in regards to current APHIS regulations. Thus, if non-specialists are concerned about genetically modified trees due to the possibility of transgenic escape, this scenario is one that assumes those concerns have been fully addressed. Non-specialists might also be concerned about difficult-to-predict or indirect ecological effects, however, and this scenario provides a way of addressing that concern separately from transgenic escape.

Scenario Three: Restoration of American Chestnut

The third scenario considers the use of biotechnology for the purpose of restoring an endangered species. Efforts to develop blight-resistant American chestnut (*Castanea dentata*) using transgenic methods are currently underway, with the ultimate goal of intentionally releasing resistant, transgenic trees into wild forest ecosystems within the species' former range (Merkle et al. 2007). This scenario is an ethically interesting case, in that many who oppose biotechnology do so on the grounds that its use (a) represents an increasing commodification of nature and exploitation of biology to provide exclusive benefits to humans (Majumder et al. 2008), and (b) tends to systemically benefit large multinational corporations, which by their very nature have interests focused on producing an economic benefit to themselves before (and potentially counter to) producing environmental or social benefits for the broader human and ecological worlds (Schurman 2004). Restoration of American chestnut, however, represents an instance where the goal is preventing a species from becoming extinct: while human nostalgia may be the cause for the restoration work, ultimately the survival of American chestnut is a benefit to the trees themselves. In addition, although it is possible that a successful restoration of the species could ultimately lead to economic benefits for corporations having the primary goal of making a profit, it is difficult to argue that those leading the efforts for biotechnologically-based restoration efforts are only out to satisfy shareholders or claim ownership of the whole species. Restoration efforts for American chestnut have largely been led by academic researchers and nonprofit organizations. Another unique aspect of this scenario is the high likelihood that current APHIS standards regarding biological containment would require an exemption for the case of species restoration:

transgenic escape is essentially the goal of such efforts, rather than an unwelcome side-effect.

Although the ethical considerations in this case may be considered less contentious for this scenario than for the first, ecosystem-level risk assessment studies would likely be helpful in restoration efforts. The fungus causing chestnut blight, *Cryphonectria parasitica*, is devastating to American and European chestnut, but it may also cause damage to trees such as *Eucalyptus* (Old and Kobayashi 1988) and *Quercus* (Radocz and Tarcali 2005; Radocz et al. 2010). Therefore, in the third scenario I consider an unmanaged forest ecosystem that includes transgenic, blight-resistant American chestnut, oaks which are less susceptible to the blight fungus than wild-type American chestnut but more susceptible than the transgenic chestnut, and tree species that do not host chestnut blight.

Systems Ecology as a Theoretical Approach

To address these three scenarios, I take a systems-theory based approach towards ecosystem modeling. This approach has been used to characterize ecosystems from a holistic perspective (e.g., Dame and Patten 1981), and models developed using systems theory methods are thought to reveal higher-level characteristics of ecosystems including the impacts of indirect effects (Higashi and Patten 1986), which have subsequently been supported by empirical research (reviewed by Strauss 1991). This ability to reveal indirect effects makes systems-based modeling an attractive tool for understanding the impacts of transgenic trees in the three scenarios, because most traditional, field-based methods are limited to investigation of direct effects identified *a priori*.

In practice, the development of ecosystems models within the systems theory paradigm involves defining the boundary of the ecosystem, partitioning the ecosystem into a comprehensive set of compartments, defining the currency in which conservative (i.e., loss-free), direct transactional flows take place, identifying which compartments are linked by transactions (including transactions from outside the ecosystem boundary), and quantifying or estimating the magnitude of those transactions (Fath et al. 2007; Hannon 1973; Patten 1978). Defining the magnitude of each compartment as a standing stock by quantification or estimation is also necessary, and determination of controlling factors on transactions can add a level of realism (and complexity) where such factors might otherwise be ignored. Conceptual models with comprehensive estimates or quantified values for standing stocks and transactions may be subjected to the assumption that the system reaches steady state given a long enough time frame, allowing for mathematically solving the systems of differential (using advanced calculus) or difference (using algebraic methods) equations. The solutions to these systems of equations provide information about expected relative quantities of compartment standing stocks and transactional flows based on the initial conditions provided. Additional mathematical manipulations can reveal throughflows for particular compartments (Finn 1976), the amount of cycling of currency within the ecosystem (Finn 1976), and non-transactional relations such as ecological mutualism (Fath and Patten 1998; Patten 1991). Relations in particular are interesting here, since they are essentially indirect effects between two compartments and can be both quantified in terms of magnitude and qualified in terms of whether one compartment has a positive, negative, or neutral influence on any other compartment. Where direct computation of the systems of equations is difficult or

intractable, computational simulations can directly calculate stock values at a series of time increments using the equations to provide a faster approach to identifying these higher-order factors (e.g., Kazançi 2007).

Conceptual Development of Ecosystem Models

The three scenarios were developed into conceptual ecosystem models by thinking about the qualitative aspects described above. Boundaries for each (sub)system are defined such that the forests and plantation are assumed to have a fixed geographic area, and do not directly border on other ecosystems outside the scenario. Ecosystems were partitioned into nonliving compartments, such as source-dependent detritus and soil, and into living compartments depending on food source (e.g. parasites, primary producers, herbivores, omnivores, carnivores, detritivores). Compartments specific to transgenic trees were included; for the first two scenarios these were distinguished separately from other *Populus* compartments. If flow rates would be expected to differ among transgenic sources of materials and other compartments for reasons directly related to the engineering goals for the transgenic organism, additional compartments downstream of the transgenic trees were added, with separate flows. Currency of the models was assumed to be carbon in all three scenarios, and known carbon transactions were represented with structural connections. When the existence of a particular flow was in doubt, it was included in the model on the basis that it is easier to set a given flow to zero than to add a new one later on, thus allowing simulations to be tested without restructuring the model itself. Controls, magnitude of stocks, and magnitude of transactions were not directly considered at the conceptual stage. Conceptual models of

each ecosystem's structure were visually represented in STELLA® v.8.1.1 software for Windows (isee Systems).

Simulation of the Transgenic Escape Scenario

Initial model simulations were run in EcoNet 2.1 beta (Kazançi 2007) using the model structure developed for the transgenic escape scenario. Quantitative estimates were made for magnitudes of initial stock sizes, inputs and outputs from the ecosystem, and internal flow rates between compartments within the ecosystem. Flows were assumed to be donor controlled, except for flows leaving from the Soil Organic Matter compartment, which were assumed to be donor-recipient controlled. Outputs were all donor-controlled, while inputs were all donor-recipient controlled (i.e. simply recipient, since the "stock" beyond the ecosystem boundary is not quantified). Flows, inputs, and outputs were all quantified as proportions relative to the stock size of the corresponding controlling compartment(s). Inputs (i.e., photosynthetic rates) and outputs (i.e., metabolic or entropic losses of carbon) were estimated with reference to previously published literature for plant (Loomis and Amthor 1999) and animal (Nagy 1987) metabolic efficiencies, adjusting rates based on the proportion of photosynthetic tissue in different plant compartments and proportions of different taxa in different herbivore, omnivore, carnivore, and detritivore compartments (derived from Kimmins 1997 and Odum 1971). Initial stock sizes for all compartments were estimated from information summarized by Odum (1971) and assumed a 10 ha ecosystem. While breakdown of plant detritus (Madritch et al. 2006) and herbivory rates (Kimmins 1997; Odum 1971) were estimated based on published literature, flow rate estimates were essentially "dummy

numbers" for other internal flows. Verifying the initial results presented from these simulations would therefore require collection of field data, additional literature searches, or global sensitivity analysis of the model.

Baseline simulations assumed that the Transgenic *Populus* compartment did not differ in flow rate from Native *Populus*, and that the ratio of Transgenic to Native *Populus* in the ecosystem was 1:100,000. Experimental simulations assumed that Transgenic *Populus* experienced a 10% rate increase in input via photosynthesis, and 25% rate increases in flows to the compartments for Plant Parasites, Herbivores that eat *Populus*, and Omnivores that eat *Populus*. In addition, Transgenic *Populus* Detritus had 25% rate increases in flows to Soil Organic Matter and to Detritivores. These rates are based on the assumption of a growth-defense tradeoff; that is, a transgenic line of *Populus* metabolically engineered to increase growth should have reduced levels of compounds to deter plant parasites, herbivores, and omnivores, leading to those organisms' increased consumption of this *Populus* genotype under field conditions. Reduced recalcitrance of litter to decay would also lead to increases in detritivores' consumption of transgenic detritus and more rapid transfer of organic matter into soil. Simulations to test the effects of feedbacks to primary producers changed the magnitudes of flows from the compartment SoilOrganics to PlParasites, WildPop, PopEscape, OtherWoody, and OtherHerb, then ran the baseline and experimental simulations as described above. Full EcoNet code and parameters are provided in the Simulation Code.

Results

Conceptual Model One: Transgenic Escape

Compartments, inputs, outputs, internal flows, and controls for this model are listed in Table 4.1. A visual representation of the Transgenic Escape model is presented in Figure 4.1. Conceptual model details in EcoNet code format can also be found in the Simulation Code.

Conceptual Model Two: Plantation/Forest Interface

The visual model of the Plantation/Forest Interface scenario is presented in Figure 4.2. The model's basic structure is analogous to two interconnected versions of the Transgenic Escape scenario, one representing the Plantation Forest and one representing the Unmanaged Forest. The compartment for Other Woody plants, and all the flows entering and leaving it, have been removed from the Plantation Forest. Similarly, the compartment for Escaped *Populus* and its flows have been removed from the Unmanaged Forest. In addition, its name has been changed to TrangenPop and its detritus compartment DetritusPT. Flows connecting the two subsystems are described in Table 4.2. Although not further addressed in this paper, the size of both *Populus* compartments in the Plantation Forest could be set to undergo an automatic reduction at regular intervals during simulation to test the effects of different harvest rotation lengths.

Conceptual Model Three: Restoration of American Chestnut

The visual model for this final scenario is shown in Figure 4.3, with details for compartment names, flows, and controls are listed in Table 4.3. The basic structure of

this model is similar to that for Model 1. However, primary producers have been divided into blight-susceptible and blight-resistant categories here, with herbaceous plants assumed to be resistant to the fungus. Fungal populations are divided into separate compartments by host, with dispersal made accountable by including flows among the three populations. This structure allows users to employ simulations to test the effects on fungal propagation that may result from different rates of host susceptibility across the three plant compartments subject to infection; presumably the Transgenic Chestnut compartment would have the lowest flow rate to its corresponding fungal compartment in any non-baseline simulations. Flow rates to the pathogen are controlled by both donor and recipient compartments, which allows for simulations testing the impacts of initial inoculum level and initial distribution of woody plant carbon mass across species, transgenic status, and susceptibility type. While no Plant Parasite or Soil Organic compartments exist in this model, it has been supplemented with a new compartment for Fungivores to account for organisms which may feed on the blight fungus. The possibility of preferential feeding on or avoidance of the focal species for herbivores and omnivores is removed from consideration here, but separate compartments for "clean" detritus and "infected" detritus exist, the latter containing dead fungus and plant litter with a small probability of re-populating the main fungal compartments. A particular advantage of STELLA simulation for this model is that flows can be controlled by complex, probabilistic "back end" scenarios. This could allow for increased rates of transfer from living susceptible matter to infective detritus dependent upon the size of the corresponding fungal compartment, representing plant mortality.

Transgenic Escape Simulations

Baseline simulations were generated for the transgenic escape model by setting all comparable input and flow rates equal for wild and transgenic *Populus* and their respective detritus compartments (Simulation Code). Under these conditions, most compartments showed a steady decline in carbon mass over the hundred-year simulation time (Figure 4.4), indicating that the estimated values for the model do not generate a steady state system. However, since manipulating a system input (transgenic *Populus* photosynthesis) is part of the transgenic simulation, utilizing a steady state system as a baseline would not necessarily provide a more objective test of the impacts of transgenic escape than the current one.

By increasing photosynthesis, herbivory, and detrital decay rates of transgenic *Populus* we generated a contrasting set of simulation data (Figure 4.5). Under these new conditions, the decline seen under baseline conditions is mitigated in all compartments other than non-transgenic primary producers and detritus from wild *Populus*, indicating broad indirect effects of metabolically engineered *Populus* on the unmanaged ecosystem. These effects had a clear time-lag before taking effect. Organisms interacting directly with *Populus* moved from decreasing to increasing stock sizes starting around 42 years, while other compartments affected moved from decreasing to increasing stock sizes between 45 and 75 years. Again, we emphasize relative patterns between Figures 4.4 and 4.5 rather than the specific quantitative results for each compartment, due to the known uncertainties for internal flow rates in the simulations.

The baseline simulation was repeated for a second test, this time including small feedbacks to all primary producer compartments. The overall impact on the ecosystem

relative to the baseline with no feedbacks to primary producers was generally a slower trend towards loss of carbon from the compartments (Figure 4.6). This slower loss of carbon came at a large initial expense to the soil organic matter compartment, but this compartment showed slow increases after the second year of the simulation compared to a trend towards slight losses in the baseline scenario without feedbacks to primary producers.

Finally, we repeated the transgenic escape simulation with feedbacks to primary producers. The most dramatically visible pattern is a breakdown of the ecosystem around year 65 of the simulation (Figure 4.7). This was attributable to the feedback from soil to primary producers being controlled at the recipient level as well as at the donor level; repeating the simulation using donor control for these flows eliminated this breakdown (data not shown). As in Figure 4.5, the negative trends of the corresponding baseline simulation were reversed under a time lag when the metabolically engineered *Populus* was present. In keeping with the effects of feedbacks to primary producers observed in Figure 4.6 relative to Figure 4.4, the initial decreases in carbon loss proceeded more slowly and became positive sooner, with compartments interacting directly with transgenic *Populus* shifting to carbon accumulation beginning around 39 years. Compartments not interacting directly with transgenic *Populus* shifted between 42 years and 49 years, with the caveat that the compartment for non-*Populus* consuming herbivores had not yet shifted towards carbon accumulation at the time point where the simulation broke down.

Discussion

The construction of the three conceptual models here provides examples for the initial stages of implementing a scientific approach to help address ethical concerns surrounding the use of genetically engineered trees under field conditions. Regardless of the level of concern an individual may hold regarding specific uses of biotechnology (transgenic escape, contained transgenic influence on a nearby unmanaged ecosystem, or intentional planting of transgenics to restore a native species), a model based on material or energy flows can be developed to conceptualize ecosystem-level effects of a transgenic tree. By populating the model with literature-based data and performance goals of genetic engineering, the conceptual models can then be used for ecosystem simulations.

The results of the transgenic escape simulation suggest that organisms interacting directly with *Populus* are not the only ones which will be affected by transgenic lines of the species if it has higher growth, herbivory, and litter decay rates. Simulations including feedbacks to primary producers from soil organic matter suggest that such feedbacks could influence the extent and timeline of impact of metabolically engineered *Populus* in an unmanaged forest. It is worth noting, however, that under the assumptions of these simulations the existence of metabolically engineered *Populu*s had a negligible impact the biomass of other primary producers.

The current approach is subject to additional development. Numerous flows in the transgenic escape simulation could be better tuned with additional reference to existing literature or by making field measurements using methods such as radioisotope tracing. An alternative approach could be to employ general sensitivity analysis for the simulation (such as the approach developed in Arogo Ogejo et al. 2010), which allows

the user to determine ranges of particular flows over which the patterns observed in Figures 4.4-4.7 remain relatively stable. The benefit of this type of analysis is that it would likely help to narrow down considerations both of which flows need more rigorous quantification, and of the level of precision required for those flows to generate useful ecosystem simulation data. Other aspects that could be included using STELLA software are the effects of environmental factors less strongly related to trophic processes, such as temperature, precipitation, or day length. Rapid changes in carbon mass can also potentially be accounted for. In the case of Model 2, for example, the Plantation Forest subsystem could be programmed for different harvest regimes, where large amounts of *Populus* biomass (wild and transgenic) are removed from the system at regular intervals. Given the time lags observed for ecosystem-level changes in response to altered flow rates for transgenic *Populus* (Figures 4.5 & 4.7), simulating harvest rotations could help rapidly determine if rotation length might act as a mitigating factor for any ecosystem changes that the new line of *Populus* precipitates in the ecosystems where it is found.

Also worth noting is the fact that this form of modeling, like any other, is not all-encompassing. Systems models of the type presented here are not spatially explicit, although the creation of distinct subsystems in the Forest Interface model suggests a way of working around the need for spatial modeling. By limiting itself to flows of carbon, differential use efficiencies of other resources, such as nitrogen or light, are disregarded. This factor could, for example, underestimate the ability of a rapidly-growing *Populu*s genotype to outcompete and perhaps competitively exclude a more slowly-growing genotype. The models presented here assume no gene flow between transgenic and wild-type *Populus*, an assumption which is unrealistic at the present time but will likely

become feasible in the future (Hoenicka and Fladung 2006; Strauss et al. 2004), nor does it account for the possibility of evolution within species or coevolutionary species interactions over time. Existing methods such as theoretical or geographical population genetics analyses of gene flow (e.g., DiFazio et al. 2004), greenhouse studies to understand competition between transgenic and wild type lines, and field work to understand persistence of novel genes in unmanaged populations (e.g., Lu et al. 2006) will still be necessary. The goal of this research is not to create an all-encompassing risk analysis tool, but to add to the number of methods available to help ensure that the still-young technology of plant genetic engineering will provide a greater level of benefits than social, environmental, and economic costs.

While the simulations are preliminary, the results here provide questions worthy of further consideration. Do primary producers take up carbon directly from soil? If so, what is the magnitude of this transaction? Will simulated results suggesting that the indirect effects of transgenic *Populus* are mediated in part by feedbacks to primary producers hold under field conditions? This work thus re-emphasizes that science can be responsive to public concerns about new technologies without being inhibited or made any less rigorous. Instead, specific ethical concerns are translated into ecologically-based models incorporating ideas about transgenic escape, indirect effects, and survival of an endangered species, while acknowledging a broader social context in which existing regulation schemes are seen as inadequate by multiple stakeholders and social friction over the economic implications of living organisms being treated as patentable intellectual property is ongoing. Simulations from one conceptual model clearly lead here to new hypotheses for field research for both basic (Does a feedback exist?) and

applied purposes (How will the feedback influence the effects of the transgenic *Populus*?), the latter of which would likely hold interest for both scientists and non-scientists concerned about the possibility of indirect ecological effects following from transgenic escape. The work therefore provides a basis for further refinement of a hypothesis-generation and decision-making tool to the scientific community while avoiding an approach to the implementation of biotechnology that disregards the views of non-scientists.

In general, the results of simulations that have been rigorously verified via references to the literature, collection of field data, and sensitivity analyses will not account for all possible environmental impacts of a given transgenic line. All models contain a certain amount of oversimplification and can therefore never be perfectly predictive of the biological world. This work, in other words, is only a partial way of addressing specific concerns about unintended ecological effects, with clear relevance to organisms living in an unmanaged context and individuals concerned with the ethics of using transgenic plants in the field. Because the approach used here does not account for evolution or possible additional competitive advantages of the metabolically-engineered *Populus* relative to native genotypes, for example, these simulations may overestimate how soon after introduction any ecological changes become noticeable. Therefore, the simulation results here could be relatively conservative in one sense, even while being unrealistically extreme in terms of the raw biomass values generated for the transgenic *Populus*. In addition, this tool could be used with equal ease for plant biologists involved with molecular breeding or more traditional approaches to plant breeding as a possible way to identify factors of potential ecological concern, since it is the gross impacts of

specific traits on growth, interactions with other species, and metabolism rather than molecular-scale processes *per se* for which these models account. A systems approach therefore also has potential for broad applicability for research on the ecosystem-level effects of introduction of new species or traits.

Because scientists are ultimately playing a professional social role and have a variety of other obligations in their daily lives, I hold that professional research ethics are only one part of the ethical principles which scientists should consider in their work. The clearest way that scientists can and should shoulder a level of responsibility for the effects of their work is by positing critical questions that they may not be able to answer themselves but could pass to other scientists (or non-scientists) with more suitable expertise. Since asking questions is necessary for ongoing scientific research, this approach is aligned with a scientific ethic. It is also well aligned with -- if not critical to the implementation and maintenance of -- democracy, market economics, and sustainability, because additional research is ultimately what helps ensure that the new technologies will have a greater benefit to cost ratio over the long term.

Placed within the context of a society based on the values of democratic decision-making and a free-market economy, public acceptance of new technologies is not a side issue, but a primary concern. I hold that gaining public acceptance is not simply an issue of advertising or scientific education, nor should it entail going so far as to obscure the presence of a new technology. When considering non-specialists, though, these seem to be the current approaches towards implementing agricultural biotechnology in the United States today. In contrast, a truly democratic approach to gaining public acceptance regarding any issue, whether it be economic, technological, or otherwise, inherently

involves an ongoing process of active listening and negotiation. The approach I take here is representative of one way in which scientists can actively listen to others' concerns regarding biotechnology.

Acknowledgements

This work was financially supported by a Graduate Research Assistantship and a Dissertation Completion Fellowship award from the University of Georgia Graduate School. Special thanks to Bernie Patten and Bob Teskey for acting as Capstone Readers for this paper.

Simulation Code

Simulation code is shown in EcoNet format; # preceeds comments, * indicates a source or sink external to the ecosystem, r indicates a Donor-Recipient controlled flow, and c indicates a Donor controlled flow proportion. Flows are listed as proportions of the corresponding stocks, rather than absolute quantities.

```
#MODEL 1 IS BUILT ON A YEARLY RATE BASIS FOR C BIOMASS (kg)
#PER 10 HA.
#Ran as Fixed time-step 4th order Runge-Kutta; 100 t, 0.001 dt.

#INITIAL STOCKS
WildPop = 87500       #Odum 1971 + inference
```

```
PopEscape = .875      #W:E ratio is 100000:1
OtherWoody = 87500    #Odum 1971 + inference
OtherHerb = 25000     #Odum 1971 + inference
PlParasites = 400     #primary:PP ratio is about 500:1
DetritusPW = 15000    #half of tree detritus
DetritusPE = .15      #from same ratio as for DetritusPW:WildPop
DetritusO = 30000     #half of tree detritus, plus 150,000 for
                      #herbaceous and secondary detritus
SoilOrganics = 30000  #40% of organic dead portion of soil, where
                      #detritus is remainder & 45000 kg/10 ha
HerbivEatsP = 33.3    #Odum 1971 + inference
HerbivAvoidsP = 66.7  #Odum 1971 + inference
OmnivEatsP = 0.8      #Odum 1971 + inference
OmnivAvoidsP = 1.7    #Odum 1971 + inference
Carniv = 0.1          #Odum 1971 + inference
OtherParasites = 1    #secondary:P ratio is about 100:1 not
                      #counting detritivores
Detritv = 1100        #Odum 1971 + inference

#PHOTOSYNTHESIS INPUTS
* -> WildPop r=2.835       #Loomis 1998
* -> PopEscape r=2.835     #Loomis 1998; (110% 3.119)
* -> OtherWoody r=2.835    #Loomis 1998
* -> OtherHerb r=4.200     #Loomis 1998
* -> PlParasites r=0.788   #Loomis 1998
```

```
#FLOWS WITHIN SYSTEM
#SoilOrganics -> PlParasites r=0.005  #Cut for no feedback
#SoilOrganics -> WildPop r=0.01       #Cut for no feedback
#SoilOrganics -> PopEscape r=0.01     #Cut for no feedback
#SoilOrganics -> OtherWoody r=0.01    #Cut for no feedback
#SoilOrganics -> OtherHerb r=0.01     #Cut for no feedback
SoilOrganics -> Detritv r=0.01

WildPop -> PlParasites c=0.005
PopEscape -> PlParasites c=0.005      #(125% 0.00625)
OtherWoody -> PlParasites c=0.004
OtherHerb -> PlParasites c=0.02

WildPop -> HerbivEatsP c=0.0225
PopEscape -> HerbivEatsP c=0.0225     #(125% 0.02813)
OtherWoody -> HerbivEatsP c=0.0075
OtherHerb -> HerbivEatsP c=0.5

OtherWoody -> HerbivAvoidsP c=0.0225
OtherHerb -> HerbivAvoidsP c=0.7

WildPop -> OmnivEatsP c=0.01688
PopEscape -> OmnivEatsP c=0.01688     #(125% 0.02813)
OtherWoody -> OmnivEatsP c=0.0056
OtherHerb -> OmnivEatsP c=0.375
```

```
OtherWoody -> OmnivAvoidsP c=0.01688
OtherHerb -> OmnivAvoidsP c=0.525

WildPop -> DetritusPW c=0.35     #Lose 1/3 at end of season for
                                 #leaves, plus some for dead branches
PopEscape -> DetritusPE c=0.35   #Lose 1/3 at end of season for
                                 #leaves, plus some for dead branches
OtherWoody -> DetritusO c=0.35   #Lose 1/3 at end of season for
                                 #leaves, plus some for dead branches
OtherHerb -> DetritusO c=0.95    #Lose 95% at end of season, rest
                                 #is seed and storage biomass

DetritusPW -> Detritv c=1.05      #Madritch et al. 2006
DetritusPW -> SoilOrganics c=0.15
DetritusPE -> Detritv c=1.05      #Madritch et al. 2006 (125%
                                  #1.3125)
DetritusPE -> SoilOrganics c=0.15 #(125% 0.1875)
DetritusO -> Detritv c=3.5
DetritusO -> SoilOrganics c=0.25

Detritv -> DetritusO c=53.46
Detritv -> OmnivEatsP c=0.25
Detritv -> OmnivAvoidsP c=0.5
Detritv -> OtherParasites c=0.05
Detritv -> Carniv c=0.4
```

```
PlParasites -> OmnivEatsP c=0.025

PlParasites -> OmnivAvoidsP c=0.025

PlParasites -> HerbivEatsP c=0.01

PlParasites -> HerbivAvoidsP c=0.01

PlParasites -> Carniv c=0.02

PlParasites -> DetritusO c=0.95

PlParasites -> OtherParasites c=0.01


OtherParasites -> OmnivEatsP c=0.02

OtherParasites -> OmnivAvoidsP c=0.04

OtherParasites -> Carniv c=0.04

OtherParasites -> DetritusO c=8.511


HerbivEatsP -> OtherParasites c=0.05

HerbivAvoidsP -> OtherParasites c=0.05

HerbivEatsP -> Carniv c=0.15

HerbivAvoidsP -> Carniv c=0.15

HerbivEatsP -> OmnivEatsP c=0.1

HerbivAvoidsP -> OmnivEatsP c=0.1

HerbivEatsP -> OmnivAvoidsP c=0.15

HerbivAvoidsP -> OmnivAvoidsP c=0.15

HerbivEatsP -> DetritusO c=44.88

HerbivAvoidsP -> DetritusO c=44.88


OmnivEatsP -> OtherParasites c=0.05
```

```
OmnivAvoidsP -> OtherParasites c=0.05

OmnivEatsP -> Carniv c=0.15

OmnivAvoidsP -> Carniv c=0.15

OmnivEatsP -> DetritusO c=35.84

OmnivAvoidsP -> DetritusO c=35.84

OmnivEatsP -> OmnivAvoidsP c=0.15

OmnivAvoidsP -> OmnivEatsP c=0.1


Carniv -> OtherParasites c=0.05

Carniv -> OmnivEatsP c=0.05

Carniv -> OmnivAvoidsP c=0.05

Carniv -> DetritusO c=26.83


#ENTROPIC & METABOLIC LOSS OF CARBON

WildPop -> * c=2.563       #Loomis 1998

PopEscape -> * c=2.563     #Loomis 1998

OtherWoody -> * c=2.563    #Loomis 1998

OtherHerb -> * c=2.533     #Loomis 1998

PlParasites -> * c= 42.274 #Loomis 1998 + Nagy 1987

DetritusPW -> * c=0.05     #Madritch et al. 2006

DetritusPE -> * c=0.05     #Madritch et al. 2006

DetritusO -> * c=0.05

SoilOrganics -> * c=0.005

HerbivEatsP -> * c=174.1         #Nagy 1987

HerbivAvoidsP -> * c=174.1       #Nagy 1987

OmnivEatsP -> * c=173.8          #Nagy 1987
```

```
OmnivAvoidsP -> * c=173.8        #Nagy 1987

Carniv -> * c=173.5              #Nagy 1987

OtherParasites -> * c=80.00      #Nagy 1987

Detritv -> * c=125               #Nagy 1987
```

Tables

Table 4.1: Names of compartments, inputs, outputs, and flows for the Transgenic Escape ecosystem conceptual model, with controls on inputs and arriving flows

Compartment names correlate with their visual representations in Figure 4.1.

Compartment	Inputs	Outputs	Departing Flows	Arriving Flows	Controls
WildPop	PhSnth1	PWEntropy	PP1, Litterfall1, HvP1, OvP1	PlantUptake1	Donor-Recipient
PopEscape	PhSnth2	PEEntropy	PP2, Litterfall2, HvP2, OvP2	PlantUptake2	Donor-Recipient
OtherHerb	PhSnth3	OHEntropy	PP3, Litterfall3, HvP3, OvP3, HvO3, OvO3	PlantUptake3	Donor-Recipient
OtherWoody	PhSnth4	OWEntropy	PP4, Litterfall4, HvP4, OvP4, HvO4, OvO4	PlantUptake4	Donor-Recipient
DetritusPW	none	DPWEntropy	Dv1, Breakdown1	Litterfall1	Donor
DetritusPE	none	DPEEntropy	Dv2, Breakdown2	Litterfall2	Donor
DetritusO	none	DOEntropy	Dv3, Breakdown3	Litterfall3, Litterfall4, ParaD1, ParaD2, DetD, HvD1, HvD2, OvD1, OvD2, CvD	Donor
SoilOrganics	none	SOEntropy	PlantUptake1, PlantUptake2, PlantUptake3, PlantUptake4, SoilProcessors, MycelialUptake	Breakdown1, Breakdown2, Breakdown3	Donor
PlParasites	PhSnth5	PPEntropy	ParaD1, ParaHv1, ParaHv2, ParaOv1, ParaOv2, ParaCv1, ParaPs1	PP1, PP2, PP3, PP4, MycelialUptake	MycelialUptake is Donor-Recipient; all others Donor
OtherParasites	none	OPEntropy	ParaD2, ParaEaten1, ParaEaten2, ParaEaten3	ParaPs1, ParaDet, HvPara1, HvPara2, OvPara1, OvPara2, CvPara	Donor
HerbivEatsP	none	HEPEntropy	HvD1, HvOv1, HvCv1, HvPara1, HvOv3	HvP1, HvP2, HvP3, HvP4, ParaHv1	Donor

Table 4.1, Continued:

Compartment	Inputs	Outputs	Departing Flows	Arriving Flows	Controls
HerbivAvoidsP	none	HAPEntropy	HvD2, HvPara2, HvCv2, HvOv2, HvOv4	HvO3, HvO4, ParaHv2	Donor
OmnivEatsP	none	OEPEntropy	OvD1, OvCv1, OvOv1, OvPara1	OvP1, OvP2, OvP3, OvP4, ParaOv1, OvDet2, HvOv1, HvOv2, OvOv2, CvOv1, ParaEaten1	Donor
OmnivAvoidsP	none	OAPEntropy	OvD2, OvOv2, OvPara2, OvCv2	OvO3, OvO4, ParaOv2, OvDet1, HvOv3, HvOv4, OvOv1, CvOv2	Donor
Carniv	none	CvEntropy	CvOv1, CvOv2, CvPara, CvD	ParaCv1, CvDet, HvCv1, HvCv2, OvCv1, OvCv2, ParaEaten3	Donor
Detritiv	none	DvEntropy	DetD, ParaDet, OvDet1, OvDet2, CvDet	Dv1, Dv2, Dv3, SoilProcessors	SoilProcessors is Donor-Recipient; all others Donor

Table 4.2: Names of cross-subsystem flows and their donor and recipient compartments in the Plantation/Forest Interface conceptual model

Donor	Flow	Recipient	Represents
WildPop	PWLitterfall2U	DetritusPW 2	Wind-carried leaf litter from plantation to forest
TransgenPop	PopTg2U	DetritusPT 2	Wind-carried leaf litter from plantation to forest
PlParasites	PlPa2U	PlParasites 2	Net migration from plantation to forest
OtherParasites	OP2U	OtherParasites 2	Net migration from plantation to forest
HerbivEatsP	HvEP2U	HerbivEatsP 2	Net migration from plantation to forest
HerbivAvoidsP	HvAP2U	HerbivAvoidsP 2	Net migration from plantation to forest
OmnivEatsP	OvEP2U	OmnivEatsP 2	Net migration from plantation to forest
OmnivAvoidsP	OvAP2U	OmnivAvoidsP 2	Net migration from plantation to forest
Carniv	Cv2U	Carniv 2	Net migration from plantation to forest
Detritiv	Dv2U	Detritiv 2	Net migration from plantation to forest
WildPop 2	PWLitterfall2P	DetritusPW	Wind-carried leaf litter from forest to plantation
OtherWoody 2	OWLitterfall2P	DetritusO	Wind-carried leaf litter from forest to plantation
PlParasites 2	PlPa2P	PlParasites	Net migration from forest to plantation
OtherParasites 2	OP2P	OtherParasites	Net migration from forest to plantation
HerbivEatsP 2	HvEP2P	HerbivEatsP	Net migration from forest to plantation
HerbivAvoidsP 2	HvAP2P	HerbivAvoidsP	Net migration from forest to plantation
OmnivEatsP 2	OvEP2P	OmnivEatsP	Net migration from forest to plantation
OmnivAvoidsP 2	OvAP2P	OmnivAvoidsP	Net migration from forest to plantation
Carniv 2	Cv2P	Carniv	Net migration from forest to plantation
Detritiv 2	Dv2P	Detritiv	Net migration from forest to plantation

Table 4.3: Names of compartments, inputs, outputs, and flows for the Transgenic Chestnut restoration ecosystem conceptual model, with controls on inputs and arriving flows

Compartment names correlate with their visual representations in Figure 4.3.

Compartment	Inputs	Outputs	Departing Flows	Arriving Flows	Controls
WildChestnut	PhSnth1	WCEntropy	WCInfect, Litter1, InfectLitter1, WCtoHv, WCtoOv	none	Donor-Recipient
TGChestnut	PhSnth2	TCEntropy	TCInfect, Litter2, InfectLitter2, TCtoHv, TCtoOv	none	Donor-Recipient
Herbaceous	PhSnth3	HerbEntropy	Litter3, HbtoHv, HbtoOv	none	Donor-Recipient
WoodyResistant	PhSnth4	WREntropy	Litter4, WRtoHv, WRtoOv	none	Donor-Recipient
WoodySusceptible	PhSnth5	WSEntropy	WSInfect, Litter5, InfectLitter5, WStoHv, WStoOv	none	Donor-Recipient
DetritusInfected	none	DIEntropy	InfDetProcess, IDtoP1, IDtoP2, IDtoP3	InfectLitter1, InfectLitter2, InfectLittter5, PWSDead, PTCDead, PWCDead	Donor
DetritusClean	none	DCEntropy	DetProcessors	Litter1, Litter2, Litter3, Litter4, Litter5, HvtoDetr, OvtoDetr, CvtoDetr, DvtoDetr, FvtoDetr, ParatoDetr	Donor
PathogenWC	PhSnth5	PWCEntropy	P1toFv, P1disp2, P1disp3, PWCDead	WCInfect, P2disp1, P3disp1, IDtoP1	WCInfect is Donor-Recipient; all others Donor
PathogenTC	none	PTCEntropy	P2toFv, P2disp1, P2disp3, PTCdead	TCInfect, P1disp2, P3disp2, IDtoP2	TCInfect is Donor-Recipient; all others Donor
PathogenWS	none	PWSEntropy	P3toFv, P3disp2, P3disp1, PWSDead	WSInfect, P1disp3, P2disp3, IDtoP3	WSInfect is Donor-Recipient; all others Donor
OtherParasites	none	ParaEntropy	ParatoDetr, ParatoOv, ParatoCv	FvtoPara, DvtoPara, HvtoPara, OvtoPara, CvtoPara	Donor

Table 4.3, Continued:

Compartment	Inputs	Outputs	Departing Flows	Arriving Flows	Controls
Fungiv	none	FvEntropy	FvtoDetr, FvtoCv, FvtoOv, FvtoPara	P1toFv, P2toFv, P3toFv	Donor
Herbiv	none	HvEntropy	HvtoDetr, HvtoOv, HvtoCv, HvtoPara	WctoHv, TctoHv, WstoHV, HbtoHv, WRtoHv	Donor
Omniv	none	OvEntropy	OvtoDetr, OvtoCv, OvtoPara	HvtoOv, CvtoOv, ParatoOv, DvtoOv, FvtoOv, WctoOV, TctoOv, WstoOv, HbtoOv, WRtoOv	Donor
Carniv	none	CvEntropy	CvtoDetr, CvtoOv, CvtoPara	OvtoCv, HvtoCv, ParatoCv, DvtoCv, FvtoCv	Donor
Detritiv	none	DvEntropy	DvtoDetr, DvtoOv, DvtoCv, DvtoPara	DetProcessors, InfDetProcess	Donor-Recipient

Figures

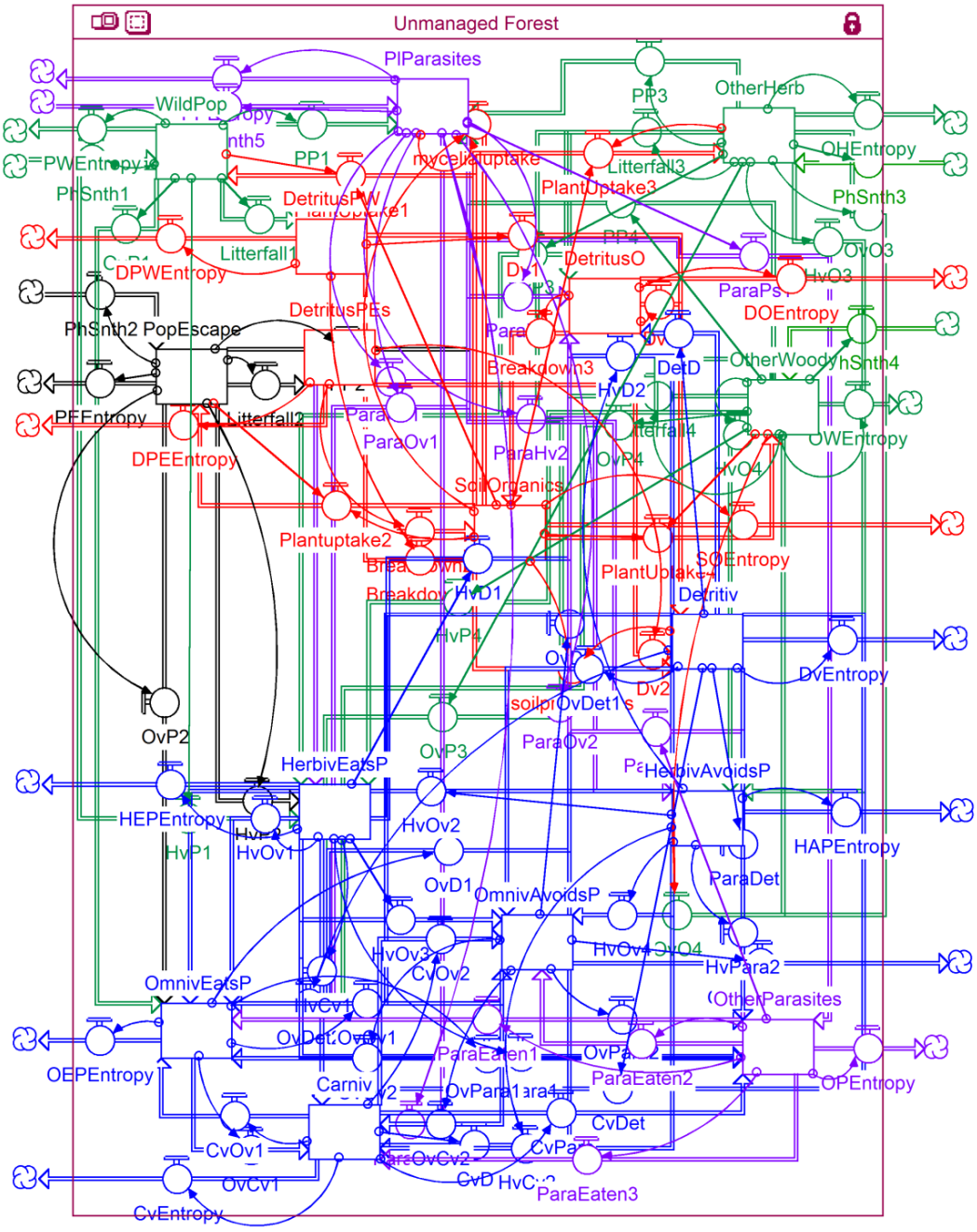

Figure 4.1: Transgenic Escape ecosystem conceptual model represented in STELLA format

Full figure on the following page. The large mauve box represents the ecosystem boundary. Small boxes represent compartments, "pipes" (double-lined arrows) represent

inputs, outputs and transactional flows, and arrows represent controls. Green items represent stocks, inputs, outputs, and internal flows departing from primary producer compartments. Red items represent the same for nonliving compartments, purple for parasitic compartments, and blue for consumer compartments. Black items highlight same for the Transgenic *Populus* compartment.

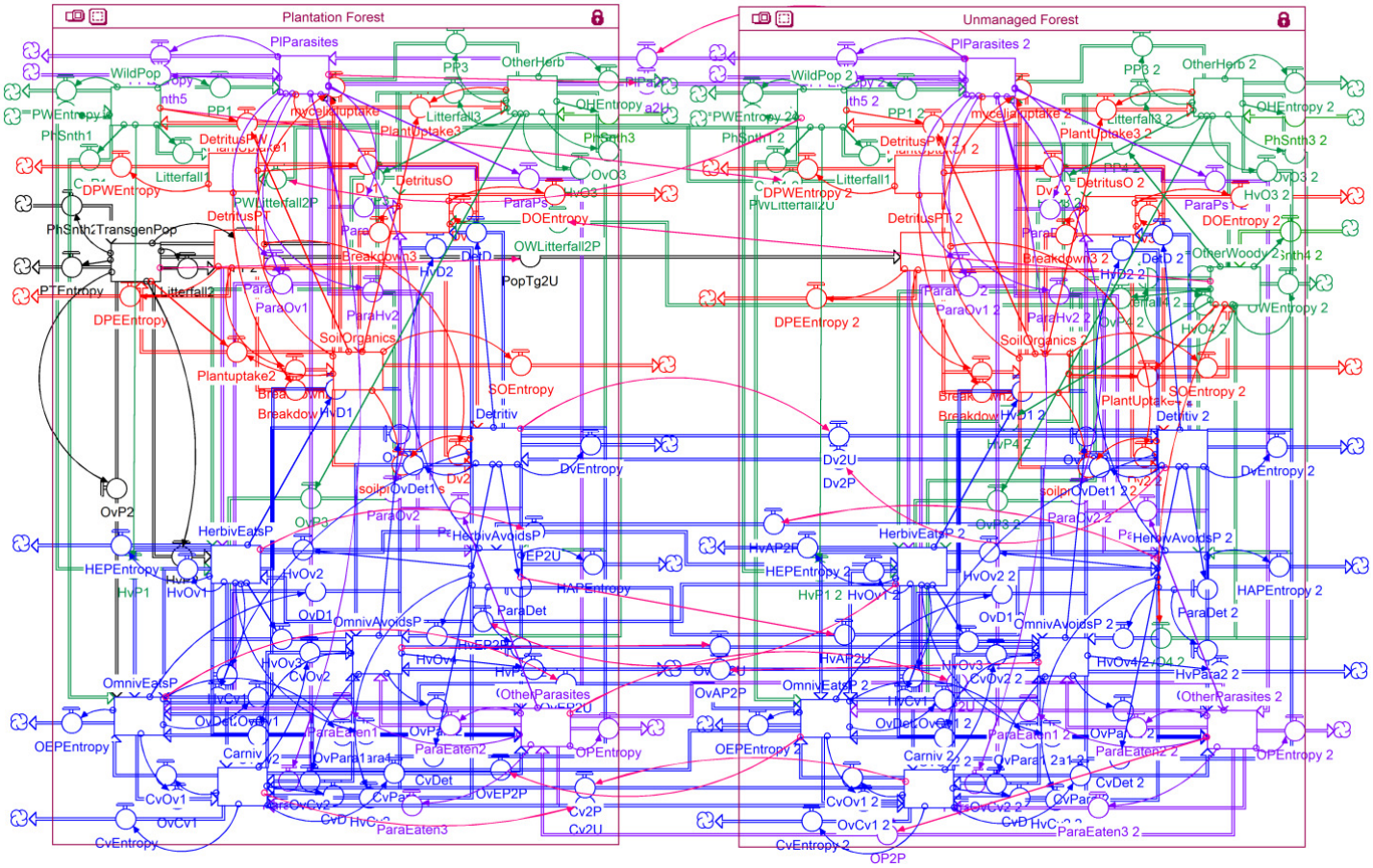

Figure 4.2: Plantation/Forest Interface conceptual model represented in STELLA format

All symbols and colors are used as in Figure 4.1. Plantation Forest compartment and flow names are identical to those in Figure 4.1, while Unmanaged Forest names have "2" appended to each.

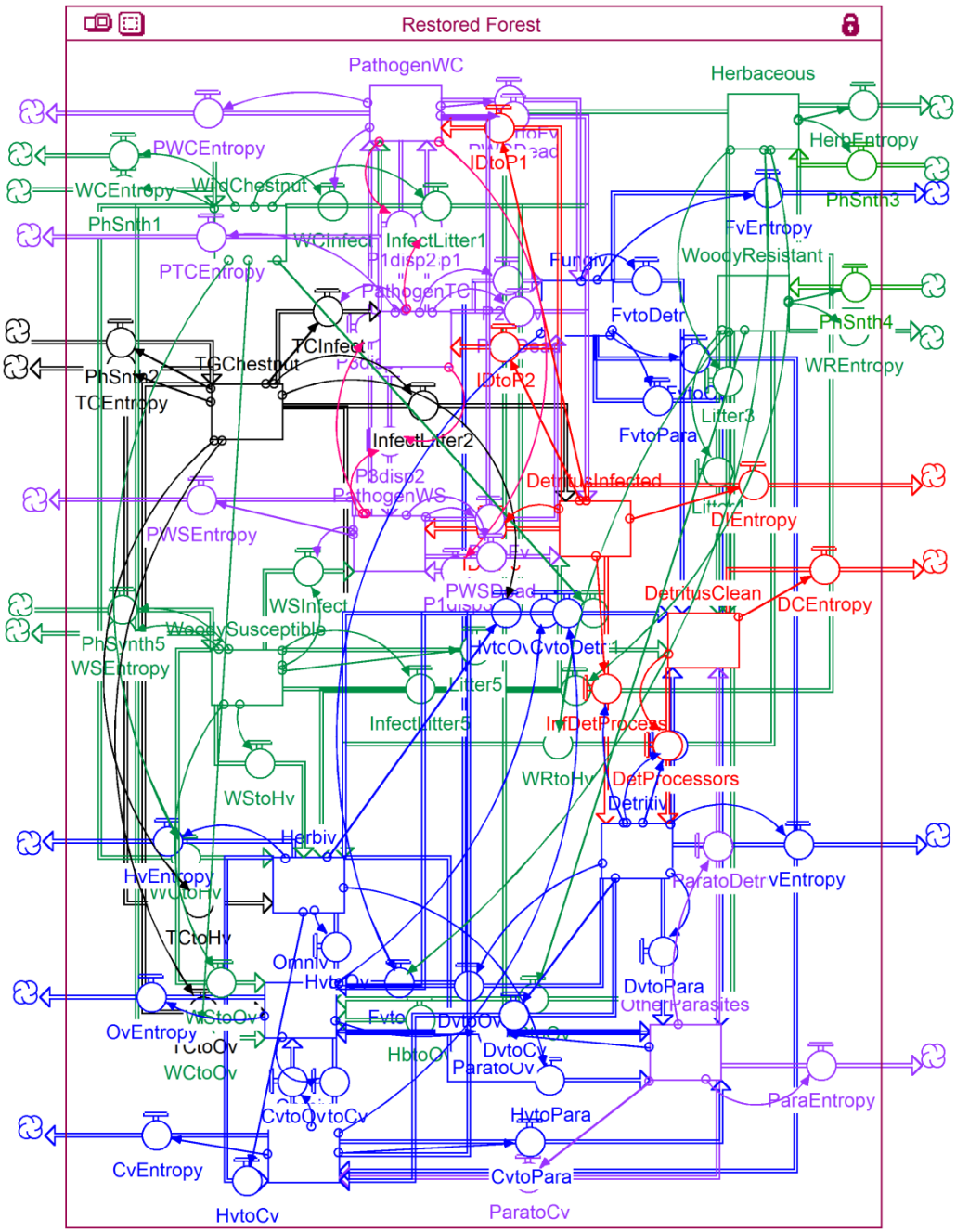

Figure 4.3: Transgenic Chestnut restored ecosystem conceptual model represented in STELLA format

Symbols and colors are used as in Figure 4.1, except that black items highlight stock, input, output, and flows for the Transgenic Chestnut compartment.

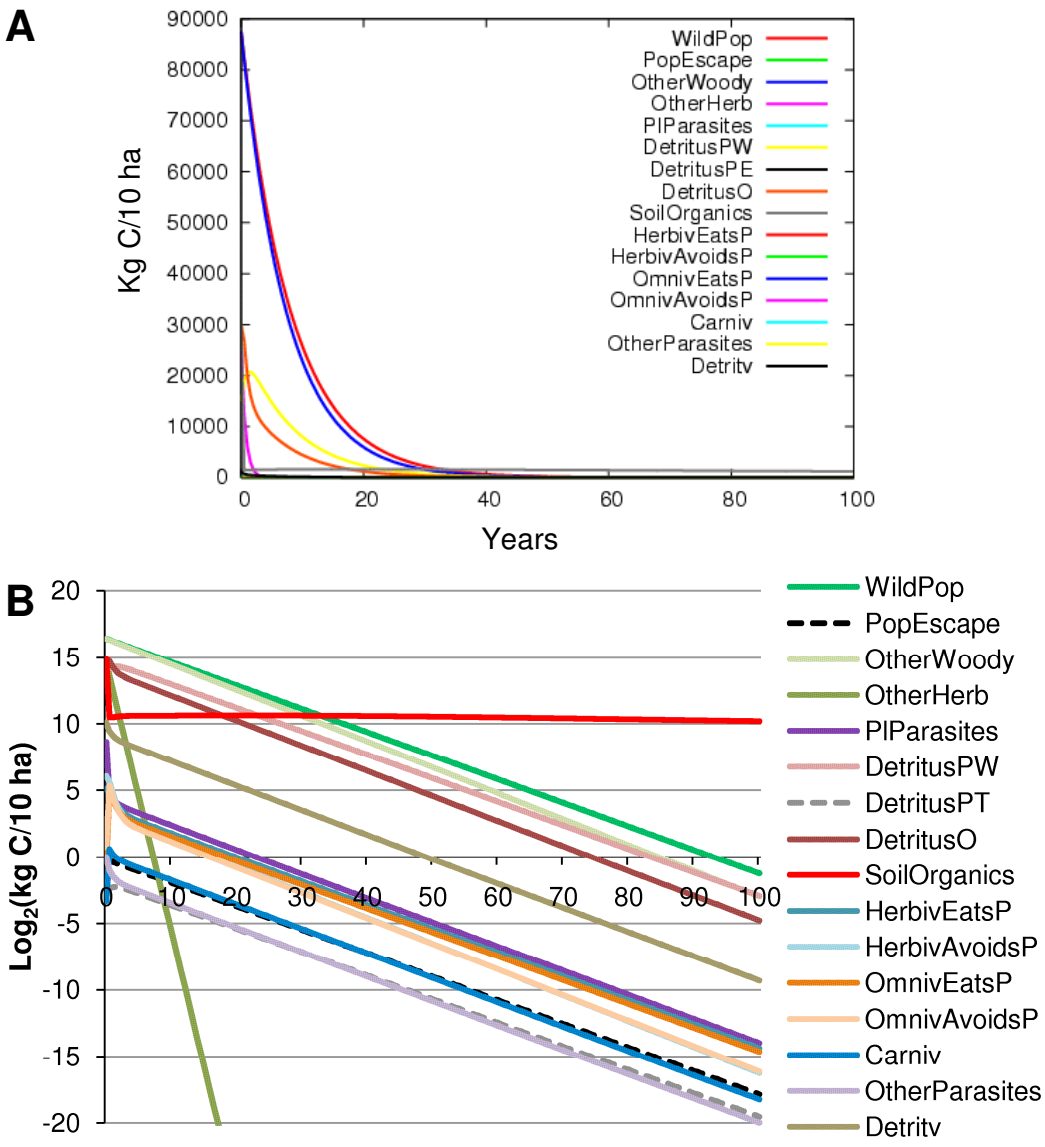

Figure 4.4: Stock values (in carbon biomass per 10 ha) for components in an unmanaged forest ecosystem over 100 years under baseline assumptions

(A) Raw simulation data; (B) Log_2-transformed data. Wild and engineered *Populus* are assumed to be equivalent in all input, output, and flow rates.

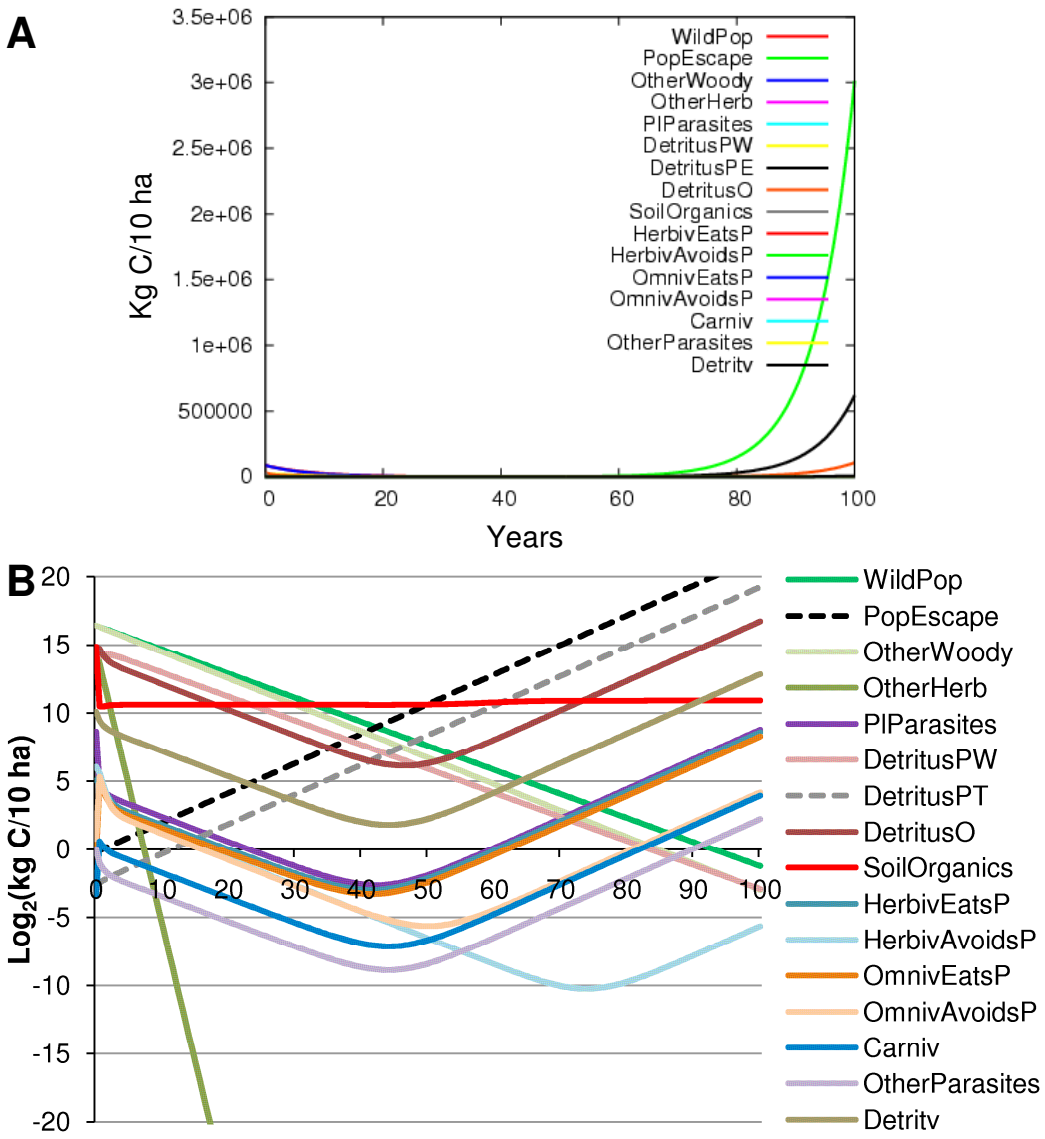

Figure 4.5: Stock values (in carbon biomass per 10 ha) over time in years for components in an unmanaged forest ecosystem into which metabolically engineered *Populus* has escaped

(A) Raw data; (B) Log_2-transformed data. Metabolically engineered *Populus* is assumed to have 10% higher photosynthetic inputs, 25% higher herbivory flow rates, and 25% higher breakdown of its detritus than wild *Populus*.

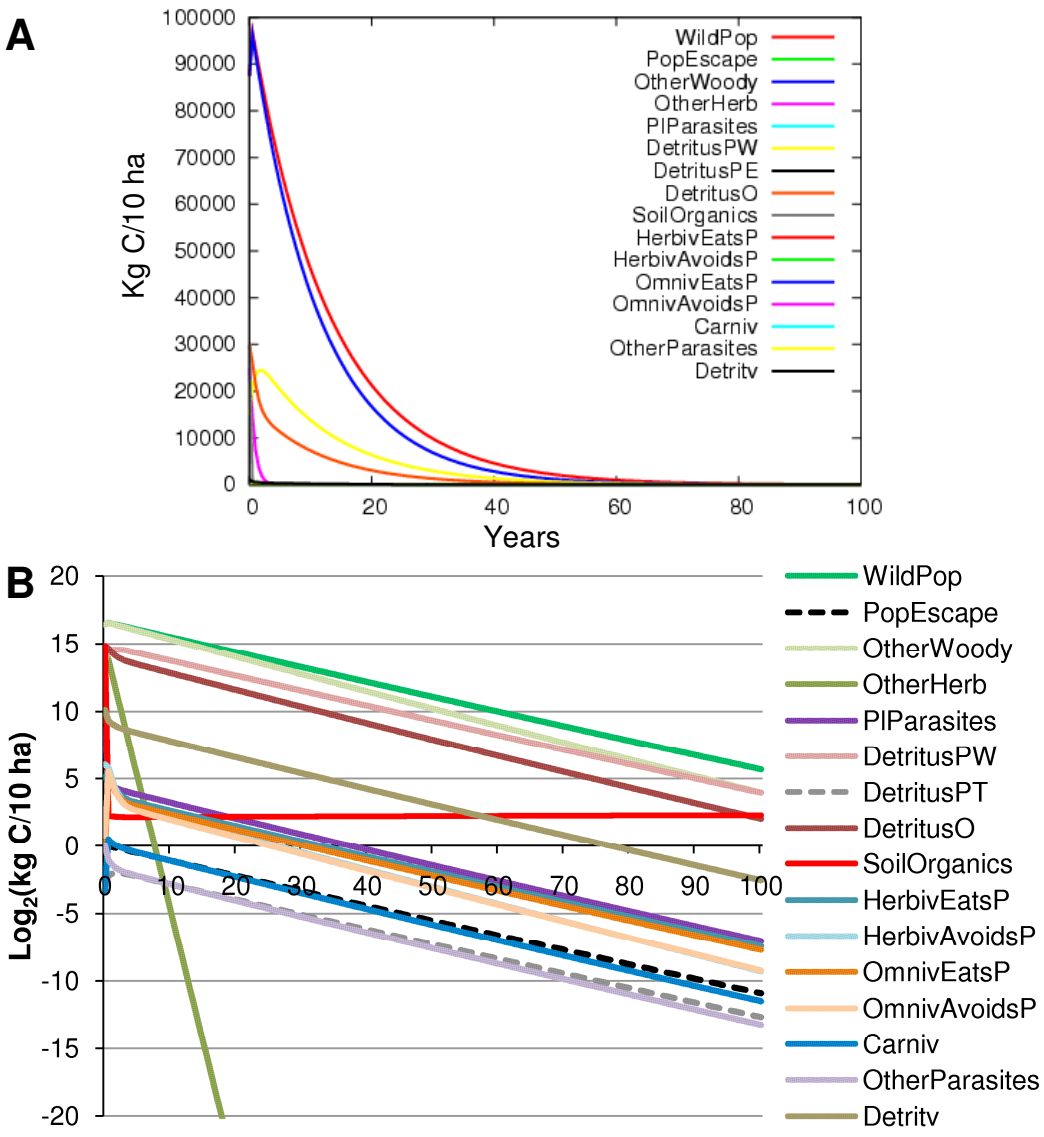

Figure 4.6: Stock values (in carbon biomass per 10 ha) over time in years for components in an unmanaged forest ecosystem under baseline assumptions plus feedbacks from soil to primary producers

(A) Raw data; (B) Log_2-transformed data. Wild and engineered *Populus* are assumed to be equivalent in all input, output, and flow rates.

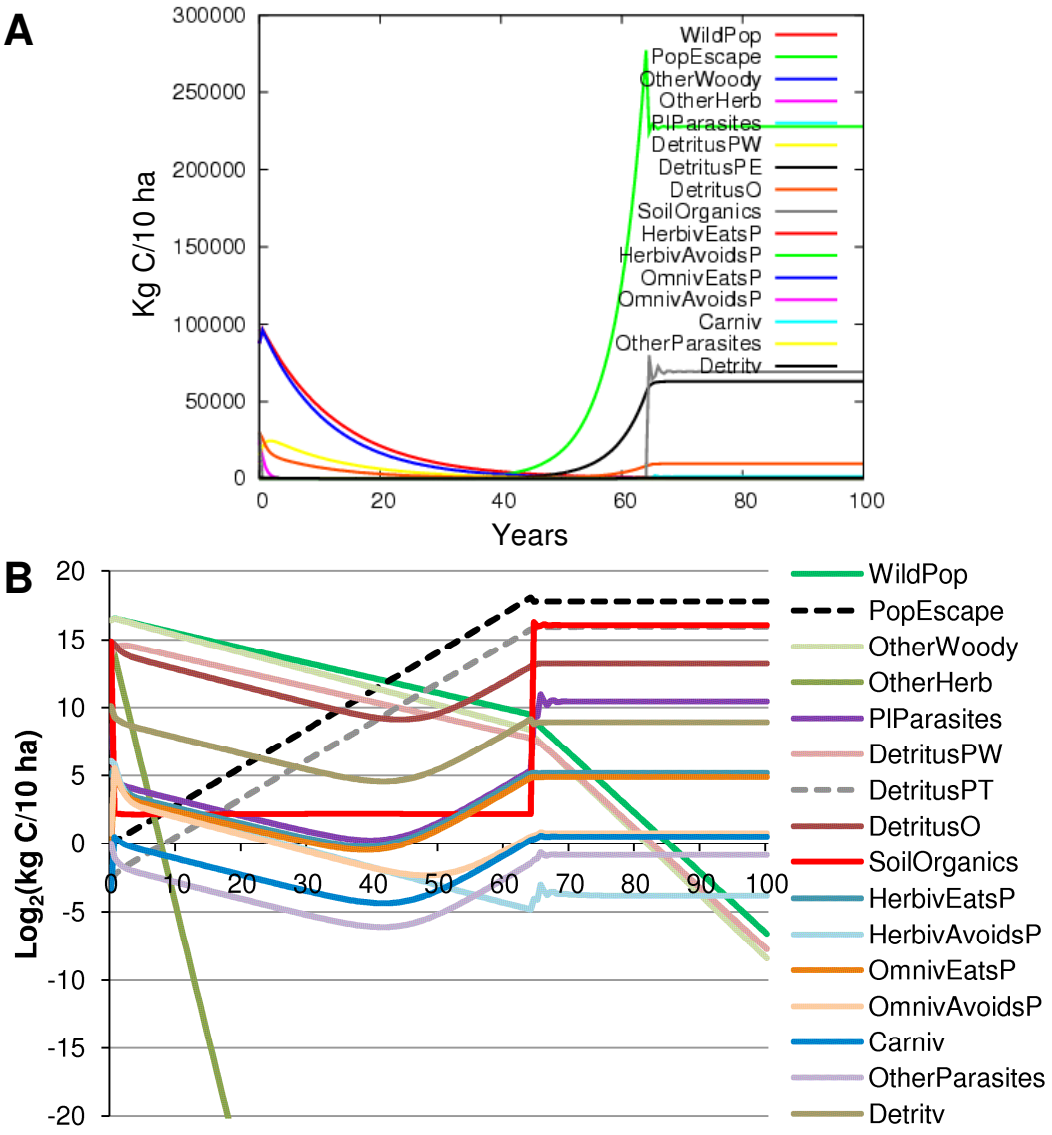

Figure 4.7: Stock values (in carbon biomass per 10 ha) over time in years for components in an unmanaged forest ecosystem into which metabolically engineered *Populus* has escaped, including feedbacks from soil to primary producers

(A) Raw data; (B) Log_2-transformed data. Data are log_2-transformed values for kg of carbon per 10 ha.

References

(2004) Monsanto v. Schmeiser. 2004 SCC 34

(2008) Geertson Seed Farms v. Johanns. 9th Cir.

(2009) Monsanto v. Bowman. S.D. Ind.

(2011) Center for Food Safety v. Vilsack. 9th Cir.

Arogo Ogejo J, Senger RS, Zhang RH (2010) Global sensitivity analysis of a process-based model for ammonia emissions from manure storage and treatment structures. *Atmospheric Environment* 44: 3621-3629

Bergstrom DM, Lucieer A, Kiefer K, Wasley J, Belbin L, Pedersen TK, Chown SL (2009) Indirect effects of invasive species removal devastate World Heritage Island. *Journal of Applied Ecology* 46: 73-81

Cox SE (2008) Genetically modified organisms: Who should pay the price for pollen drift contamination? *Drake Journal of Agricultural Law* 13: 401-418

Dame RF, Patten BC (1981) Analysis of energy flows in an intertidal oyster reef. *Marine Ecology Progress Series* 5: 115-124

de Melo-Martin I, Meghani Z (2008) Beyond risk. *EMBO Reports* 9: 302-306

DiFazio SP, Slavov GT, Burczyk J, Leonardi S, Strauss SH (2004) Gene flow from tree plantations and implications for transgenic risk assessment. In: Walter C, Carson M (eds) *Plantation Forest Biotechnology for the 21st Century*. Research Signpost, Kerala, India, pp 405-422

Ellstrand NC (2003) Current knowledge of gene flow in plants: Implications for transgene flow. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358: 1163-1170

Fath BD, Patten BC (1998) Network synergism: Emergence of positive relations in ecological systems. *Ecological Modelling* 107: 127-143

Fath BD, Scharler UM, Ulanowicz RE, Hannon B (2007) Ecological network analysis: Network construction. *Ecological Modelling* 208: 49-55

Finn JT (1976) Measures of ecosystem structure and function derived from analysis of flows. *Journal of Theoretical Biology* 56: 363-380

Franken M (2000) Fear of Frankenfoods: A better labeling standard for genetically modified foods. *Minnesota Intellectual Property Review* 1: 153-181

Gavora JS, Lister EE (1989) Practical and ethical considerations of agricultural research assistance for the Third World. *Journal of Agricultural and Environmental Ethics* 2: 307-322

Hannon B (1973) The structure of ecosystems. *Journal of Theoretical Biology* 41: 535-546

Hellsten I (2003) Focus on metaphors: The case of "Frankenfood" on the web. *Journal of Computer-Mediated Communication* 8: doi: 10.1111/j.1083-6101.2003.tb00218.x

Higashi M, Patten BC (1986) Further aspects of the analysis of indirect effects in ecosystems. *Ecological Modelling* 31: 69-77

Higashi M, Patten BC (1989) Dominance of indirect causality in ecosystems. *The American Naturalist* 133: 288-302

Hoenicka H, Fladung M (2006) Biosafety in *Populus spp*. and other forest trees: From non-native species to taxa derived from traditional breeding and genetic engineering. *Trees* 20: 131-144

Kazançi C (2007) EcoNet: A new software for ecological modeling, simulation and network analysis. *Ecological Modelling* 208: 3-8

Kimmins JP (1997) *Forest Ecology: A Foundation for Sustainable Management*. Prentice-Hall, Inc., Upper Saddle River, NJ

Kwit C, Moon HS, Warwick SI, Stewart Jr CN (2011) Transgene introgression in crop relatives: Molecular evidence and mitigation strategies. *Trends in Biotechnology* 29: 284-293

Lemaux PG (2008) Genetically engineered plants and foods: A scientist's analysis of the issues (Part I). *Annual Review of Plant Biology* 59: 771-812

Loomis RS, Amthor JS (1999) Yield potential, plant assimulatory capacity, and metabolic efficiencies. *Crop Science* 39: 1584-1596

Lu B-R, Snow AA (2005) Gene flow from genetically modified rice and its environmental consequences. *BioScience* 55: 669-678

Lu M-Z, Chen X-L, Hu J-J (2006) Empirical assessment of gene flow from transgenic poplar plantation. *The 9th International Symposium on the Biosafety of Genetically Modified Organisms, Jeju Island, Korea, 24-29 September 2006: Biosafety Research and Environmental Risk Assessment*. Saskatoon, pp 131-136

Lu Y, Wu K, Jiang Y, Xia B, Li P, Feng H, Wyckhuys KAG, Guo Y (2010) Mirid bug outbreaks in multiple crops correlated with wide-scale adoption of Bt cotton in China. *Science* 328: 1151-1154

Madritch M, Donaldson J, Lindroth R (2006) Genetic identity of *Populus tremuloides* litter influences decomposition and nutrient release in a mixed forest stand. *Ecosystems* 9: 528-537

Majumder MA, Byrne MM, Bongmba E, Rothenberg LS, Dubler NN (2008) Ethical challenges of patenting "nature": Legal and economic accounts of altered nature as property. In: Lustig BA, Brody BA, McKenny GP (eds) *Altering Nature*. Springer Netherlands, pp 199-273

Merges RP (1987-1988) Intellectual property in higher life forms: The patent system and controversial technologies. *Maryland Law Review* 47: 1051-1075

Merkle S, Andrade G, Nairn C, Powell W, Maynard C (2007) Restoration of threatened species: a noble cause for transgenic trees. *Tree Genetics & Genomes* 3: 111-118

Myhr AI (2010) The challenge of scientific uncertainty and disunity in risk assessment and management of GM crops. *Environmental Values* 19: 7-31

Nagy KA (1987) Field metabolic rate and food requirement scaling in mammals and birds. *Ecological Monographs* 57: 111-128

National Research Council (2008) *Genetically Engineered Organisms, Wildlife, and Habitat: A Workshop Summary*. The National Academies Press, Washington, D.C.

Odum EP (1971) *Fundamentals of Ecology*. W.B. Saunders Company, Philadelphia, PA

Old KM, Kobayashi T (1988) Eucalypts are susceptible to the chestnut blight fungus, *Cryphonectria parasitica*. *Australian Journal of Botany* 36: 599-603

Patten BC (1978) Systems approach to the concept of environment. *Ohio Journal of Science* 78: 206-222

Patten BC (1991) Network ecology: Indirect determination of the life-environment relationship in ecosystems. In: Higashi M, Burns TP (eds) *Theoretical Ecosystem Ecology: The Network Perspective*. Cambridge University Press, London, pp 288-351

Radocz L, Tarcali G (2005) Identification of natural infection of *Quercus spp.* by chestnut blight fungus (*Cryphonectria parasitica*). *Acta Horticulturae* 693: 617-620

Radocz L, Tarcali G, Jenei A (2010) In vitro susceptibility of different oak species to the blight fungus (*Cryphonectria parasitica*). *Acta Horticulturae* 866: 455-458

Regal PJ (1994) Scientific principles for ecologically based risk assessment of transgenic organisms. *Molecular Ecology* 3: 5-13

Schmidt DE (2008) Postcard from the reality-based universe: "Wish you were all here!" A meditation on the relationship between science, intellectual property law, and the rights of indigenous populations in plant genetic resources. *Environmental Law* 38: 315-365

Schurman R (2004) Fighting "Frankenfoods": Industry opportunity structures and the efficacy of the anti-biotech movement in western Europe. *Social Problems* 51: 243-268

Selig MF (2008) Getting more from the giving tree: Regulating the use of biotechnology in forest management. *Virginia Environmental Law Journal* 26: 577

Smyth S, Khachatourians GG, Phillips PWB (2002) Liabilities and economics of transgenic crops. *Nature Biotechnology* 20: 537-541

Strauss SH, Brunner AM, Busov VB, Ma C, Meilan R (2004) Ten lessons from 15 years of transgenic *Populus* research. *Forestry* 77: 455-465

Strauss SH, Kershen DL, Bouton JH, Redick TP, Tan H, Sedjo RA (2010) Far-reaching deleterious impacts of regulations on research and environmental studies of recombinant DNA-modified perennial biofuel crops in the United States. *BioScience* 60: 729-741

Strauss SY (1991) Indirect effects in community ecology: Their definition, study and importance. *Trends in Ecology & Evolution* 6: 206-210

Tannert C, Elvers H-D, Jandrig B (2007) The ethics of uncertainty. *EMBO Reports* 8: 892-896

USGPO (2010) Introduction of Organisms and Products Altered or Produced Through Genetic Engineering Which are Plant Pests or Which There is Reason to Beleive are Plant Pests. 7 CFR 340. U.S. Government Printing Office, pp 458-475

Wagner W, Kronberger N, Gaskell G, Allansdottir A, Allum N, de Cheveigne S, Dahinden U, Diego C, Montali L, Mortensen AT, Pfenning U, Rusanen T, Seger N (2001) Nature in disorder: The troubled public of biotechnology. In: Gaskell G, Bauer MW (eds) *Biotechnology 1996-2000: The Years of Controversy*. NMSI Trading Ltd, London

Williams CG (2010) Long-distance pine pollen still germinates after meso-scale dispersal. *American Journal of Botany* 97: 846-855

Zapiola ML, Campbell CK, Butler MD, Mallory-Smith CA (2008) Escape and establishment of transgenic glyphosate-resistant creeping bentgrass *Agrostis stolonifera* in Oregon, USA: A 4-year study. *Journal of Applied Ecology* 45: 486-494

CHAPTER 5.

CONCLUSIONS

The work presented in this dissertation has aimed to understand phenylpropanoid metabolism in *Populus* at multiple biological scales. Each chapter on its own has generated new insights, at the levels of gene family organization within and across plant genomes, metabolic and molecular regulation in response to multiple forms of perturbation in cell cultures, and modeling assessment of ecosystem-scale effects due to changes in organismal metabolism, respectively. These findings can also be linked to generate a more comprehensive understanding of *Populus* phenylpropanoid metabolism, particularly with respect to understanding metabolic diversity and propagation of these metabolites' effects across biological scales. As is often the case in scientific research, the findings also led to new questions for future studies on the topic. Initial technical groundwork has already been laid towards answering some of these questions. In this manner, the work described in this dissertation has contributed toward the understanding of phenylpropanoid biology within multiple subfields and laid groundwork for integrating this understanding in a cross-scale manner.

Phylogenomics of BAHD Acyltransferases and Phenylpropanoid Metabolism

Initial gene expression analysis, some of which is described in **APPENDIX A**, combined with literature-informed rational biochemistry provided support for the involvement of the BAHD acyltransferase gene family in *Populus* phenylpropanoid metabolism. Based on findings in **CHAPTER 2**, the unusually large number of BAHD genes in the *Populus* genome relative to the other dicots surveyed, particularly the two *Populus*-dominant subclades known to be associated with synthesis of flavonoid and benzoate derivatives, could very well be linked to the reported diversity of phenylpropanoids in this genus. Within *Populus*, differential retention of BAHD paralogues following duplication events was evident, with an overrepresentation of genes derived from local/tandem events and underrepresentation of those derived from the recent genome duplication, relative to the genome-wide patterns. It appears that expansion of the BAHD acyltransferase family has equipped *Populus* with an augmented "toolkit" for modifying the phenylpropanoid skeletons via acylation. Closer examination suggested that such expanded toolkits do not merely contain more copies of the same gene, but can include divergent expression patterns, as evidenced by the highly homologous *Populus CHATL* paralogues. Since phenylpropanoids can serve as both acyl donors and as substrates for BAHD enzymes, acylation has doubly important consequences for this suite of metabolites. That all five genomes surveyed showed evidence of taxon-specific subfamily expansion suggests that evolutionary pressures favor different "BAHD toolkits" in different plant taxa depending on a variety of internal and external constraints. The results hint at the complexity of phenylpropanoid

metabolism well beyond the core biosynthetic pathway, which likely involves homeostatic regulation and interfaces with other cellular metabolisms and processes.

A natural follow-up to understand the evolution of the BAHD acyltransferases is functional characterization of recently duplicated genes in *Populus*. A prominent set of candidates for such work is the *CHATL* cluster, based on their divergent expression across genotypes, tissues and stress treatments (**CHAPTER 2**). Transgenic manipulation of candidate genes via RNAi (RNA interference) is one such approach for functional characterization, as outlined in **APPENDIX A**. Particularly relevant RNAi vectors designed included one to silence both *CHATL1 & 2* and another to silence the *CHATL3 & 6* paralogues. Use of these vectors to generate transgenic *Populus* could help determine the extent to which differences in *CHATL* expression translate into different functional consequences *in vivo*. A second approach to understanding the type(s) and process(es) of functional divergence among *CHATL* paralogues is outlined in **APPENDIX B**, in which the construction of *CHATL2* and *CHATL3* protein expression vectors is described. Recombinant proteins from these constructs could be used for *in vitro* biochemical assays to help determine whether or not biochemical function of these paralogues has diverged along with gene expression. Some of the tools for future investigation have therefore already been assembled and are ready for the next phase of research, extending the present investigation that was primarily genomic and bioinformatic in scope. Integration of *in vitro* biochemical data and *in vivo* metabolic data from transgenic plants should improve our ability to infer functional evolution of large gene families like BAHD across a broader range of genome-sequenced species.

Metabolic and Transcriptomic Investigation of Phenylpropanoid Perturbation

The omics approach taken in **CHAPTER 3** has allowed transcriptional and metabolic investigation of phenylpropanoid regulation and homeostasis in response to two different forms of perturbation in *Populus* cell suspension cultures. The elicitor methyl jasmonate operates primarily via transcriptional activation to stimulate phenylpropanoid biosynthesis while altering multiple other processes. The three enzyme inhibitors lead to direct metabolic perturbations that appear to propagate to the level of transcription in order for the cells to reestablish homeostasis. Thus, the two forms of perturbation were directed initially at different levels of biological organization, but both influenced metabolism and gene expression within and beyond the phenylpropanoid pathway. In addition to highlighting the effects of perturbation on phenylpropanoid partitioning to branch pathways, metabolic and gene expression data supported previously suggested links between phenylpropanoid metabolism and carbon and nitrogen homeostasis. Particularly interesting is the insight that reducing flux into the phenylpropanoid pathway seems to be linked with changes in amino acid levels but less so with changes in levels of citric acid cycle components, sugars, or starch. Whether the weaker metabolic effects on central carbon pools as observed is a broad phenomenon in plants, whether it is specific to the fast-growing, phenylpropanoid-rich species within the Salicaceae, or if it may be an artefact of the heterotrophic nature of the cell culture system has yet to be determined. In contrast, amino acid and phenylpropanoid metabolism likely represents an irreconcilable metabolic-level tradeoff in that phenylalanine (and debatably tyrosine) committed for synthesis of phenylpropanoids cannot be used for protein biosynthesis. The efficient recycling of ammonium released

from PAL (and possibly TAL) catalysis into amino acid pools during cinnamate (or *p*-coumarate) synthesis may represent an adaptive feature of phenylpropanoid-rich species that has evolved to tolerate ecological constraints such as limited nitrogen availability. Nevertheless, the shifts in amino acid levels observed in PAL-limited cells suggest that phenylpropanoid biosynthesis still comes at a cost to protein accumulation.

One major caveat to these interpretations is the limited applicability of heterotrophic cell culture systems to whole-plant ecophysiological conditions. In particular, carbon itself was fed at an excess in the cell cultures in the form of sucrose, so the kinds of constraints on energy metabolism and carbon assimilation that may be observed in whole-plant systems, whether under field or greenhouse conditions, were simply irrelevant in these cells. For example, Rubisco represents the largest protein pool in photosynthetic tissues, so limitations on protein synthesis due to high levels of phenylpropanoid metabolism would reduce the future potential of a plant's carbon assimilation capacity. Therefore, a relevant question arising from the findings in **CHAPTER 3** is whether similar results are likely to occur in photoautotrophic cell cultures not supplemented with a carbon source or in intact *Populus* trees. This question is also prompted by the lack of phenolic glycoside (salicinoid) accumulation in the heterotrophic cultures, as these compounds represent a major carbon sink in intact *Populus*. For this reason, a similar analysis to that accomplished here on elicitor treated *Populus* lines with downregulated expression of core phenylpropanoid genes could be particularly informative. Although not addressed in this dissertation, one side project in the Tsai lab that has begun to bear small fruit involves the development of a short term genetic assay system for *Populus* plantlets grown *in vitro*, potentially including

photoautotrophic plantlets. A transient gene silencing system would bear many similarities with enzyme inhibitor treatments, with the additional advantage of reducing the likelihood of direct effects on non-targeted enzymes or enzyme isoforms. Bringing this kind of tool to full utility would allow for more rapid experimentation on the questions stimulated by the results generated from the cell culture work and would reduce the preparation time required for transgenic experiments in *Populus* by about 90%.

Simulation of Ecological Effects Related to Reduced Phenylpropanoid Biosynthesis

The work in **CHAPTER 4**, although preliminary, shows a new conceptual approach for assessing possible environmental effects of novel tree genotypes in managed and unmanaged ecosystems. By developing a simulation model of hypothetical effects of altering condensed tannin accumulation in *Populus*, I was able to highlight whole-ecosystem effects on carbon cycling from changes in a major phenylpropanoid end product previously demonstrated to influence community and ecosystem-level processes. In noting the possible propagation of the effects of altering phenylpropanoid metabolism *in planta* to the ecological scale, the work represents a mathematically-based approach to understanding how introduced genotypes produced via transgenic or breeding technologies, or introduced by translocation to a new habitat, may affect the environment. Most current work in transgenic assessment focuses primarily on the likelihood and effects of gene flow among transgenic organisms and their close relatives. While researchers have also investigated more broadly the effects of genotype, including transgenic genotypes, on other species known to interact with them and specific

ecosystem processes, there has not been much comprehensive exploration of the propagation of direct effects of genotype at more broadly to the whole-ecosystem scale. Such work is both novel in its holistic approach and practical in its potential outcome. It can also help address a major concern among those resistant to biotechnology -- that our scientific understanding of the effects introducing new genotypes into an ecosystem is still lacking.

In the proof-of-concept simulation, carbon flows from the environment into metabolically engineered *Populus*, from engineered *Populus* to herbivores, and from engineered detritus to soil and detritivroes were assumed to be altered relative to wild type *Populus*. These alterations exemplified a "metabolic engineering goal phenotype" arising from theoretically reducing condensed tannin accumulation in order to optimize the genotype for biomass production. These changes led to simulation results showing shifts in most carbon pools in the ecosystem on a one-century time scale, indicating that tree genotypes with different direct effects on only some organisms in the environment also have indirect effects propagating to other ecosystem components. Determining the extent to which these indirect effects actually play out in the field, as opposed to *in silico*, will first require a global sensitivity analysis of the full simulation model. Such an analysis should reveal which direct carbon flows have the greatest effects overall on the ecosystem, and which can vary widely with little overall influence. This determination can be followed by more precise quantification of the specific flows having the greatest influence on ecosystem carbon dynamics. The precision needed for these flows in order to understand the ecosystem as a whole should also be evident from initial sensitivity analysis, and may not necessarily be as great as is currently achievable using standard

instrumentation. A more important factor will likely be gaining an understanding the variability in carbon flows over time and under varying environmental conditions.

A comprehensive "ground-truthing" of the simulation model would provide baseline data to work towards development and implementation of a novel design-oriented approach in which *a priori* environmental assessment of transgenic organism "concepts" is possible even before labwork is initiated to generate them. Testing of alternative concept "extended phenotypes" (carbon flows to and from the transgenic organism) within simulated ecosystems expected as release sites or generating concern as possible escape sites can help identify ecosystem-generalized quantitative parameter windows for traits acceptable from the environmental perspective. As long as actual transgenic lines exhibit carbon-flow traits within these windows, movement of the organism to field testing in the ecosystems so simulated could potentially be expedited. Similarly, if initial models assessing possible environmental effects of a transgenic genotype have been run before the generation of that genotype, measuring actual changes in direct carbon flows to and from the new genotype can feed back into the simulation for additional assessment prior to field trials. In this manner, model development acts as a foundation for iterative cycles of *in silico* simulation and sensitivity analysis along with experimental flow rate data collection. As seen for **APPENDIX A** and **APPENDIX B**, this work may not be fully complete, but the conceptual groundwork now exists and the necessary steps to advance the work are clear.

Synthesis and Conclusion

The understanding of *Populus* phenylpropanoid metabolism need not be limited to one biological scale; more holistic insights about this phenomenon can be achieved through the use of multiple lenses that focus on different aspects of biology. While work in **CHAPTER 2** focused on the genome scale and used primarily bioinformatic tools, the findings were informed by transcript profiling data. Follow-up questions regarding functional divergence of duplicated genes will require physiological and biochemical approaches. The data also offered a glimpse of multiple genetic factors potentially underlying the structural diversity of phenylpropanoid compounds. **CHAPTER 3** explored phenylpropanoid complexity in terms of the interconnections within its branch pathways and extending to primary metabolism, employing transcript profiling, metabolic profiling, and gross-scale analysis of major carbon and nitrogen pools at the multi-omics scale. These analyses showed linkages between the accumulation of condensed tannins and other phenylpropanoid pools, shikimate pathway intermediates, amino acid metabolism, primary carbon metabolism, and fatty acid metabolism. Such linkages have implications for both "internal" physiological processes such as growth or allocation of acquired resources and for more ecophysiological processes like defense responses or carbon-nitrogen metabolism. The focus was widened to phenylpropanoids' complex effects at ecosystem scale in **CHAPTER 4**, employing known effects of altered condensed tannin levels on ecosystem processes as impetus for developing simulations to understand how changing growth-defense tradeoffs via metabolic engineering may translate into changes in carbon flows and pools among different ecosystem components. This work feeds back to the whole-plant level in that further development of this

approach may help inform biotechnologists which strategies and traits for metabolic engineering have the lowest environmental cost-to-benefit ratios and whether field management practices may help mitigate any foreseen costs of altered traits. In summary, the study of phenylpropanoid metabolism in *Populus* at multiple biological scales does not merely provide insights at the scale being studied. Here, the work at three different scales usually addressed by different subfields of biology has also produced synergistic connections linking biological understanding of phenylpropanoid complexity from the molecular to the ecosystem level, connections that are only made possible by considering a single phenomenon from multiple perspectives.

APPENDICES

APPENDIX A.

EXPRESSION OF BAHD ACYLTRANSFERASE, CYTOCHROME P450 SUBFAMILY 92A, AND LACCASE GENES IN *POPULUS* AND ASSEMBLY OF RNAi CONSTRUCTS

Background

While the diversity of BAHD acyltransferases and their possible roles in *Populus* phenylpropanoid metabolism have already been addressed in detail (**CHAPTER 2**), two additional gene families warrant interest due to their possible roles in phenylpropanoid metabolism. Some laccases and cytochomes P450 in subfamily 92A were both identified prior to 2007 in the Tsai lab, along with several BAHD acyltransferases, as showing transcriptional coregulation with phenolic glycosides (PGs). PGs are a category of defense secondary metabolites exclusive to the Salicaceae likely to derive from the phenylpropanoid pathway in general (Tsai et al. 2006), and in particular, from cinnamate and/or benzoate (Babst et al. 2010; Zenk 1967). Since the biosynthesis and regulation of PGs is still poorly understood, further investigation of these three gene families was a high priority for the Tsai lab at the start of my dissertation work.

Laccases are *p*-diphenol:dioxygen oxidoreductases capable of polymerizing monolignols and other phenolic compounds *in vitro* (reviewed by Gavnholt and Larsen 2002). While fungal laccases have established roles in lignin degradation, plant laccases have long been suggested to play a role in lignin biosynthesis (Freudenberg 1959;

Gavnholt and Larsen 2002), with much of the evidence favoring this interpretation indirect. Evidence for direct involvement of laccases in lignification has been gleaned in recent years based on biochemical analysis (Bao et al. 1993) and studies of two laccase mutants in *Arabidopsis* (Berthet et al. 2011). The laccase family likely plays multiple physiological roles in plants, as other *Arabidopsis* laccase mutants exhibit changes in the timing of flowering and response to root dehydration stress, while some exhibit no visible phenotype and may therefore be limited to metabolic roles (Cai et al. 2006). Laccases also seem to be involved in the regulation of other phenylpropanoid branch pathways. A laccase in *Arabidopsis* is capable of polymerizing epicatechin to form proanthocyanidin (Cai et al. 2006; Pourcel et al. 2005), an early step in the condensed tannin pathway. While downregulation of individual laccases in *Populus tremula* x *alba* did not influence their lignin content or syringyl-to-guaiacyl lignin (S/G) ratio, total soluble phenolics were increased in stem tissue (Ranocha et al. 2002). Of particular interest was the increased accumulation of salicortin, salireposide, and tremulacin in two lines expressing only residual levels of *LAC3* (Ranocha et al. 2002), suggesting that laccases may be indirectly involved in regulating PG levels in *Populus*.

Cytochrome P450 monooxygenases (CYPs) play diverse metabolic roles ranging from detoxification of xenobiotics to biosynthesis of hormones and secondary metabolites. These enzymes catalyze oxidative reactions of a wide variety of organic substrates, including secondary metabolites, and some CYPs can perform sequential modifications on a single substrate (Schuler and Werck-Reichhart 2003). Phylogenomic analysis has previously established 11 CYP clans in land plants (Nelson and Werck-Reichhart 2011). The CYP71 clan is the largest and is thought to have the greatest

functional diversity. It includes several key biosynthetic enzymes of particular relevance for phenylpropanoid metabolism. For example, the CYP73 family encodes cinnamate-4-hydroxylases (C4H) acting in the core phenylpropanoid pathway; lignin biosynthesis involves members of the CYP84 family, encoding coniferaldehyde-5-hydroxylases/ferulate-5-hyroxylases (CAld5H or F5H), and the CYP98 family, encoding coumaroyl-shikimate 3'-hydroxylases (C3'H; Ehlting et al. 2006). The CYP75B subfamily of flavanoid-3'-hydroxylases (F3'H) and the CYP75A subfamily of flavonoid 3′, 5′-hydroxylases (F3'5'H) are involved in flavonoid biosynthesis (Ayabe and Akashi 2006). Microarray data mining in the Tsai lab has previously identified several CYP92A members of the CYP71 clan that exhibited co-regulation with PGs. The CYP92 family appears to be paralogous to the CYP75 family (Nelson and Werck-Reichhart 2011), and several CYP92A members have been linked to defense responses associated with chemical and bacterial stimuli in maize (Persans et al. 2001), chemical stimulus in Capsicum (Dai et al. 2007), and bacterial or fungal pathogen response in tobacco (Czernic et al. 1996; Ralston et al. 2001). A pea CYP92A6 was shown to be involved in brassinosteroid biosynthesis during etiolated hypocotyl growth (Kang et al. 2001). However, the CYP92 family is absent from *Arabidopsis*, which is known to synthesize brassinosteroids (Nelson et al. 2004). Combined with known activity of CYP85 enzymes in BR synthesis, this suggests the reported activity of pea CYP92A6 in BR synthesis is likely a lineage-specific shift to a new metabolic function rather than the primary function of the subfamily in angiosperms (Nelson and Werck-Reichhart 2011; Nomura and Bishop 2006). Based on our previously observed coregulation patterns, whether or

not the CYP92A subfamily may also have evolved in a lineage-specific manner in the Salicaceae to support PG biosynthesis is an interesting question to consider.

BAHD acyltransferases likely play an important role in contributing to the chemical diversity of PGs. BAHD enzyme activity in the Salicaceae could, for example, catalyze acylation of an acetyl moiety to salicin or salicortin to give rise to 2'-*O*-acetylsalicin or 2'-*O*-acetylsalicortin, respectively. Similarly, acylation of a benzoyl moiety to salicin or salicortin can lead to tremuloidin or tremulacin, respectively. Recent work has demonstrated that a benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase (BPBT) in *Petunia* is critical for the synthesis of floral volatile compounds phenylethyl benzoate and benzylbenzoate (Boatright et al. 2004; Orlova et al. 2006). Benzylbenzoate has notable structural similarity to complex PGs such as salicortin (Babst et al. 2010), and is known to act as a metabolic source of benzoic acid in *Petunia* (Orlova et al. 2006).

At the onset of my dissertation research, one of the goals of the work was to help identify candidate genes in *Populus* from three gene families, selected based on rational biochemistry and known involvement with phenylpropanoid metabolism, that may be involved with PG metabolism. Early approaches involved microarray data mining and QPCR to determine whether the expression of any BAHD acyltransferase, CYP subfamily 92A, and laccase genes was coordinated with shifts in PG accumulation under a variety of stress conditions and tissue types. A genomic level characterization of the BAHD acyltransferase family was presented in **CHAPTER 2**, but possible links between particular BAHD acyltransferases and PG metabolism have not yet been addressed.

Another original objective was to generate transgenic *Populus tremula* x *alba* with RNAi-silenced expression of candidate genes to investigate their involvement in PG metabolism. Included in the candidate gene set was the *CHATL* paralogues previously identified for their divergent expression patterns (**CHAPTER 2**). Transgenic silencing of different *CHATL* paralogues, in conjunction with stress experiments, is one approach to help determine the roles of these genes *in vivo* and to assess whether and how differences in expression translate to the physiological function(s). Although the *Populus* transformation work did not continue, I present here the initial gene expression data, selection of candidate genes and construction of RNAi vectors. These RNAi vectors can be used by new lab members to clarify the roles of these genes *in planta*.

Methods

Tissues and RNA Extraction

For the nitrogen stress experiment, RNA isolated from young and mature leaves of *P. fremontii x angustifolia* genotypes 1979 and 3200 and *P. tremuloides* genotype 271 grown under nitrogen-replete or nitrogen-limited hydroponic conditions were provided by Scott Harding and Edward Anino. RNA samples from 1979 and 3200 were the same as those used for microarray analysis presented in **CHAPTER 2**. In addition, the CTAB protocol (Tsai et al. 2003) was used to extract RNA from leaf apex, LPI 0 & 1, LPI 3 & 4, internodes between LPI1 and LPI4 ("Internodes 1-4"), and internodes between LPI7 and LPI10 ("Internodes 7-10") sourced from three hydroponically-grown *Populus*

tremuloides genotype 271 and male flowers, female flowers, and root tips sourced from wild-grown *P. tremuloides* as described in **CHAPTER 2**.

Synthesis of cDNA & QPCR

RNA was used for synthesis of cDNA and subsequent QPCR analysis as previously described (**CHAPTER 2**) with the exception that no DNase treatment was conducted prior to cDNA synthesis in the *P. tremuloides* nitrogen stress experiment. Primer information not already described therein is available in the Tsai lab. Housekeeping genes included elongation factor 1β (EF1β), cyclophilin (Cyp), and ubiquitin-conjugating enzyme E2 (UBCc).

Cloning of Gene Fragments into pGFPm-T vector and subcloning into RNAi vectors

Candidate gene fragments for RNAi construction were PCR-amplified using cDNA synthesized from *P. tremuloides* RNA (line 271, a mixture of either LPI 3 and root tip tissues or LPI 0 and LPI 1) and RedTaq DNA polymerase (Sigma). Primers for laccase and BAHD acyltransferase cloning were identical to those used for QPCR analysis, while CYP92A primers were separately designed to amplify longer regions. Samples were checked for successful amplification by electrophoresis on a 1.5% agarose gel in TAE buffer, then either column- or gel-purified depending on whether any non-target amplicons were observed. Purified gene fragments were ligated to the pGFPm-T cloning vector (Luo et al. 2008) previously digested with *Xcm* I, transformed into TOP10 or TOP10F' *E. coli*, and grown on ampicillin selection medium. Up to three colonies, confirmed to be positive according to colony PCR, were used for plasmid extraction and

Sanger sequencing confirmation. The insert orientation was noted upon analysis of sequencing data.

Confirmed clones were subjected to two separate restriction digestion reactions, one for the "sense" fragment relative to the SP6 promoter in the pGFPm-T vector (Frag 1, *Bam*H I & *Spe* I) and the other for the "antisense" fragment (Frag 2, *Asc* I & *Swa* I). Frag 1 was ligated to RNAi vector pFGC5941 pre-digested with compatible *Bam*H I & *Xba* I, then transformed into *E. coli*. Antibiotic-resistant clones were subjected to directional colony-PCR confirmation and Sanger sequencing. Successful pFGC5941:Frag 1 constructs were digested with *Asc* I and *Swa* I, then ligated to Frag 2, transformed into *E. coli*, and the transformants subjected to the above confirmation procedures. Early work also utilized pGSA1285 as the RNAi vector, although chloramphenicol selection of *Agrobacterium* RNAi transformants was unsuccessful.

Plasmid Transformation into *Agrobacterium*

Confirmed RNAi constructs were transformed into chemically competent *Agrobacterium tumefasciens* pMP90 by the freeze-thaw method (Holsters et al. 1978). Transformed *Agrobacterium* was plated onto kanamycin selection medium and grown for two days at 28^{o}C, followed by colony PCR confirmation.

Results

Rationale For Gene Family Choices

Initial exploration of gene expression in laccase, CYP92A, and BAHD acyltransferase families was based on microarray data collected from a variety of *Populus* genotypes and stress conditions. Candidate genes from these families were identified on the basis of their coregulation with PG accumulations or their reported PG association in the literature. Although direct stress responses of these genes may be a confounding factor, their co-regulation with PG levels across multiple stress treatments was used as a criterion to further reduce the candidate gene list. QPCR was used to verify the expression patterns observed in microarray data, focusing on the effects of nitrogen stress in young leaf (LPI3), root tip, and mature internode tissues (between LPI7-8) of *Populus tremuloides* line 271, as well as young (LPI1) and mature (LPI5) leaves of *Populus fremontii* × *angustifolia* genotypes 1979 and 3200. Nitrogen stress has been shown to induce variable PG responses in these *Populus* genotypes in the Tsai lab (Scott Harding, personal communication), and baseline concentrations of PGs and CTs differ strongly between the genotypes as well (Harding et al. 2009).

Laccase Gene Expression Under Nitrogen Limitation

Six *Lac* primer pairs were designed to analyze *Lac1a-d* (four paralogs), *Lac2/3* (two paralogs), *Lac90a-d* (four paralogs), *Lac110a*, *Lac110b*, and *Lac110c* expression. Consistent with their widely hypothesized role in lignification, nearly all laccase genes were most strongly expressed in the internode tissues of *P. tremuloides*; the exception was *Lac110a*, which was most strongly expressed in root tips (Figure A.1). *Lac1(a-d)*,

Lac2/3, and *Lac110b* showed the highest expression among laccases in internode tissues, but did not exhibit a clear response to nitrogen stress. *Lac90(a-d)* and *Lac110c*, while having notably lower expression in internodes than the other laccases, were both significantly upregulated by under nitrogen stressed conditions (1.8 fold with p=0.0116 and 1.4 fold with p=0.0414, respectively). The root-specific *Lac110a* was upregulated 3.7-fold under nitrogen limitation (p=0.0002). The low-expressing *Lac90(a-d)* was also upregulated 2.8-fold in root tips under nitrogen limitation (p=0.0001). Laccase gene expression in young leaves was very low overall, although *Lac1a* and *Lac110b* were expressed by at least an order of magnitude greater than any other laccases in this tissue. All but *Lac2/3* exhibited a 2-fold increase in leaf expression under nitrogen stress. The overall expression data was more consistent with a role of laccases in lignification rather than PG biosynthesis.

A second set of nitrogen stress experiments was also analyzed, involving two leaf ages and two *Populus fremontii* x *angustifolia* hybrid genotypes. The 1979 and 3200 genotypes differ in their foliar PG response under nitrogen limitation, with increased levels in 1979 and decreased levels in 3200 at the time points for which tissues were collected for gene expression analysis (Scott Harding, personal communication). This allows dissection of gene expression patterns that avoids conflating leaf-specific gene expression with PG coregulation. Laccase genes showed little indication of co-regulation with PG levels in *P. fremontii* x *angustifolia* hybrids, broadly consistent with patterns seen in *P. tremuloides*, which also ccumulates more PGs in leaves than in woody tissues (Figure A.2). *Lac1(a-d)* was the most strongly expressed of the laccases in both hybrid lines and was stimulated by nitrogen limitation regardless of leaf age or genotype. Its

expression was somewhat lower in LPI5 than in LPI1 for genotype 1979, but this pattern was reversed in genotype 3200. Expression of all other laccases was low in leaves of both genotypes. However, all laccase genes examined showed a striking induction (ranging from 4.7 to 53.3 fold) in LPI5 of genotype 3200 under nitrogen limitation (Figure A.2). Interestingly, in LPI6 of the same genotype, soluble PG accumulation was also reduced about 4.4 fold under nitrogen limitation (Harding, personal communication), indicating a possible inverse relationship between soluble PGs and laccase expression consistent with previous observations in antisense-*LAC3 Populus* hybrid lines (Ranocha et al. 2002). Overall, the results suggest that laccase induction may be more directly related to nitrogen status itself than to PG accumulation.

CYP92A Gene Expression Under Nitrogen Limitation

QPCR was conducted for two sets of paralogous CYP92A members, *CYP92A17/19* (likely to detect *17v1*, *17v2*, and *19v1*, but not the truncated *19v2*) and *CYP92A24/25*. In *P. tremuloides*, the *CYP92A* genes examined showed stronger expression in leaves than in stems and roots, and with consistently higher baseline levels of expression than for laccase genes in the corresponding tissues (Figure A.3). Expression of *CYP92A24* was more than twice as high as that of *CYP92A17* in leaves, about 75% greater in internodes, and about 25% greater in root tips. Both set of genes showed clear upregulation in nitrogen-stressed leaf tissues, with *CYP92A17* also increased by more than 70% ($p=0.0004$) and *CYP92A24* showing a trend towards upregulation in root tips ($p=0.0651$). In leaves of *P. fremontii* x *angustifolia* hybrids, *CYP92A24* had similar baseline expression as seen for leaves of *P. tremuloides*, but

CYP92A17 had much lower expression than in *P. tremuloides* (Figure A.4). Similar shifts in expression due to leaf age by genotype as seen for laccases, with reduced baseline expression in LPI5 of genotype 1979 but increased baseline expression in LPI5 of genotype 3200 relative to that seen in the corresponding LPI1 tissues, were observed for both CYP92A genes. However, this pattern was not nearly as distinct as responses related to nitrogen stress. Expression of *CYP92A24* was strongly increased in LPI1 while showing little change in LPI5 of genotype 1979 under nitrogen limitation; its expression was reduced in genotype 3200, especially in LPI5. Similar patterns were observed for *CYP92A17*. The contrasting responses in CYP92A gene expression between the two hybrid genotypes under nitrogen limitation closely resembled foliar PG accumulation patterns, suggesting *CYP92A* are reasonable candidates for involvement in PG metabolism.

BAHD Acyltransferase Gene Expression Under Nitrogen Limitation

Expression of the CHATL subfamily BAHD acyltransferases *CHATL1/2*, *CHATL3/6* and *CHATL4/5* was examined by QPCR in *P. tremuloides*. *CHATL1*/2 and *CHATL3*/6 expression were highest in leaves overall, whereas *CHATL4/5* was detected mainly in root tips (Figure A.5). Expression of *CHATL3/6* was higher than that of *CHATL1/2*, with upregulation in leaf tissues under nitrogen stress of about 2.3 fold and 1.5 fold, respectively. Although these genes were more weakly expressed in root tips, statistically significant increases in expression under nitrogen limitation were detected there as well, with increases of 53% and 70%, respectively.

The leaf-expressing *CHATL1/2* and *CHATL3/6* were selected for further analysis, along with additional BAHD family genes, in the *P. fremontii* x *angustifolia* hybrids that exhibited different PG responses under nitrogen limitation. This broader suite of BAHD acyltransferase candidates was selected from previous microarray analysis for verification by QPCR. For ease of presentation, these genes were divided into two types, well-expressed and weakly expressed, shown in Figures A.6 and A.7, respectively. The well-expressed group included Clade Ia malonyltransferase-like members *MATL1/2/3* and *MATL7/8*, with the remaining members from Clade Va, including anthraniloyl-CoA:methanol acyltransferase-like *AMATL1*, acetyl/benzoyl transferase-like *ABTL12*, and both pairs of *CHATL* genes. Genotype-dependent developmental shifts in expression were observed under nitrogen-replete conditions, again similar to patterns seen for laccase and CYP92A genes. For instance, the most strongly expressed genes for both genotypes were *CHATL3/6*, followed by *MATL1/2/3* and *CHATL1/2*, which were all present at higher levels in LPI1 than LPI5 for genotype 1979 while being present at lower levels in LPI1 than LPI5 for genotype 3200 (Figure A.6A). Under nitrogen limitation, two patterns of expression changes were observed for genotype 1979: (1) strong upregulation in LPI1 and no change or subtle downregulation in LPI5 (*MATL1/2/3*, *MATL7/8*, *CHATL3/6*), and (2) upregulation in both tissues (*AMATL1*, *ABTL12*, *CHATL1/2*; Figure A.6A). Within genotype 3200, the first group of genes described for genotype 1979 above (*MATL1/2/3*, *MATL7/8*, *CHATL3/6*) exhibited a strong down-regulation in LPI5 but no change in LPI1 (Figure A.6B). *ABTL12* from the second group exhibited upregulation in both leaves as observed for genotype 1979, but with a greater magnitude. *AMATL1* also showed a stronger upregulation in LPI1 in genotype 3200

under nitrogen stress than in the same leaf in genotype 1979, although in LPI5 expression of this gene showed no significant response. *CHATL1/2* had a similar expression pattern to *AMATL1* in genotype 3200, although the nitrogen limitation response in LPI1 was not statistically significant. These data indicate likely coregulation with PG levels for *MATL1/2/3*, *MATL7/8*, and *CHATL3/6*. *AMATL1* and *ABTL12* seem to be associated with nitrogen stress response, while *CHATL1/2* had a somewhat ambiguous, genotype-dependent stress response.

The low-expressing group spanned all six of the clades sampled within the BAHD superfamily, representing a broad range of putative functionalities. They included two *MATL* genes, acyltransferase-like gene *ATL4*, five alcohol acyltransferase-like genes (*AATL*s, using four primer pairs), coniferyl alcohol acetyltransferase-like gene *CFATL2*, *ABTL3*, *CHATL4/5*, and six of the seven *Populus* hydroxycinnamoyltransferases (*HCT*s; Figure A.7). Some genes in this low-expressing group were only expressed above the detection limit in genotype 1979 (*AATL2*, *AATL18/19*) or in genotype 3200 (*AATL14*, *ABTL3*, *HCT5/7*), but not both. Overall, none of these genes showed expression patterns between the two genotypes indicative of coregulation with PG levels (Figure A.7). Three additional genes, *MATL5*, *ABTL13*, and *HCT2* were also examined by QPCR but were found to express at very low levels (data not shown).

BAHD Acyltransferase Gene Expression Across *Populus tremuloides* Tissues

An extensive examination of BAHD acyltransferase gene expression was also carried out across *Populus tremuloides* tissues. Five of the 22 gene sets tested were classified as "well-expressed" (Figure A.8), including two sets in Clade Ia (*MATL1/2/3*

and *MATL5*), one gene in Clade IIIb (*CFATL2*), and two of the CHATL sets in Clade Va (*CHATL1/2* and *CHATL3/6*). These genes were biased strongly towards higher expression in apical or expanding leaves (LPI 0-1). Twelve of the 22 gene sets were designated as "weakly expressed" (Figure A.9), including two genes in Clade Ia, one in Clade Ib, one in Clade IIIa, and four sets each in Clades Va and Vb. Among these, *HCT1/6* was the only gene set expressed in all eight tissues. Its highest expression was in internodes 7-10 and elongating roots, a pattern that is in line with their reported association with lignification. *ABTL12* was readily detected in LPI8, internodes 7-10, and elongating roots. Several genes with tissue-preferential expression were also noted, including *AATL13* in elongating roots, *ABTL13* in male flowers, and *HCT3/4* in LPI 8. Finally, no or very low expression (i.e., <0.01 mean relative expression) was detected for five gene sets that may be considered not expressed in the tissues examined. Most of these genes were in Clade IIIa (primer sets *AATL2*, *AATL4*, *AATL8*, *AATL18/19*), although one was in Clade Va (*ABTL3*).

Candidate Genes

Based on the above data and on existing microarray data in the Tsai lab (not shown), we prioritized candidate genes as described in Table A.1. Primary criteria were high overall expression relative to other genes within the same functional group and responsiveness to nitrogen stress in a pattern consistent with PG accumulation. The rank order indicates likelihood of involvement with PG metabolism based on all available data, with the top four considered excellent candidates for further work. Two additional

CYP92A genes were included on the basis of microarray data from the nitrogen stress experiments in genotypes 1979 and 3200 (data not shown).

RNAi Constructs

Eight RNAi constructs were initially generated using the pGSA1285 vector (ABRC; reference information available at Gendler et al. 2008). However, after difficulty with chloramphenicol screening during transformation of the plasmids into *Agrobacterium*, the pFGC5941 vector was used instead to employ kanamycin screening as an alternative (ABRC; reference information available at Gendler et al. 2008). A list of candidate genes and RNAi constructs in the pGSA1285 vector (in *E. coli* TOP10 host) that have been sequence-confirmed are noted in Table A.2. For *MATL1/2/3*, the construct was labeled previously as *MaTL8/9/10*, and for *CYP92A20*, constructs targeting both *v1* and *v2* are available.

A total of nine RNAi constructs were fully assembled in the pFGC5941 vector, sequence-confirmed, and transformed into *Agrobacterium tumefasciens* pMP90 (Table A.2). All glycerol stocks for BAHD acyltransferases carried the updated nomenclature described in **CHAPTER 2**. An empty vector control for pFGC5941 in *Agrobacterium* is also available. It appears that *CYP92A20v1* and *CYP92A20v2* were annotated as separate gene models in the v1.1 *Populus* genome, but correspond to a single gene in the updated v2 *Populus* genome. For this reason, only one construct (*CYP92A20v2*) was prepared in pFGC5941. A construct containing the first fragment of *CYP92A20v1* in pFGC5941 is available for completion if the construct for *v2* does not provide effective gene silencing

in planta. A construct for *LAC2/3* was also initiated but not completed; two independent bacterial transformations are available in *E. coli* TOP10.

Discussion and Future Directions

Gene expression analysis in nitrogen stressed *Populus* lines and across an array of tissues in *P. tremuloides* provides circumstantial evidence for coregulation of some CYP92A and BAHD acyltransferase genes with PG levels. Despite previous work indicating that silencing of *Lac3* can influence PG levels (Ranocha et al. 2002), gene expression analysis provided little evidence for coregulation of *Populus* laccases with PGs. However, this does not entirely rule out a role for laccases in PG metabolism; their putative role in lignin polymerization might alter the pool size of phenylpropanoid intermediates upstream of monolignols. In this manner, metabolic competition between PG and lignin biosynthesis might be influenced by laccases in *Populus*.

RNAi constructs for the top four candidates for involvement with PG metabolism have been assembled in two different vectors. RNAi constructs have also been prepared for five additional candidates in the pFGC5941 backbone. Although no attempts at plant transformation were fully successful despite over a year of work (one "transformant" of the empty vector pFGC5941 was regenerated to the shoot stage), the available *Agrobacterium* stocks can facilitate future work to assess the functions of these candidate genes *in planta*.

One particularly interesting approach would be to focus on silencing of *CHATL1/2* and *CHATL3/6*. While *CHATL3/*6 appears to be a better candidate than

CHATL1/2 for involvement in PG metabolism, all four genes are closely related paralogues. While *CHATL1-3* are located on a separate linkage group from *CHATL6*, phylogenetically the genes cluster into two distinct pairs that are sister subclades to each other (**CHAPTER 2**). Furthermore, the two gene pairs show evidence of distinct expression patterns indicative of functional divergence despite their high homology. Examination of growth and metabolic consequences of RNAi silencing specifically targeting *CHATL1/2* or *CHATL3/6* may therefore provide insights on the evolution of metabolic gene function, regardless of their PG involvement. A biochemical approach to understand divergence of CHATL function is outlined in **APPENDIX B**. The combined transgenic and biochemical approaches could generate additional knowledge about the process of functional divergence at the biochemical and physiological levels, supporting existing data at the transcript level.

Tables

Table A.1: Ranking of Candidate Genes Coregulated With Phenolic Glycoside Accumulation in *Populus*.

*(Ranocha et al. 2002)

Rank	Gene(s)	Expression Pattern	Basis for Selection
1	CYP92A24/25	N responsive (down in 3200); strongest in leaves	Strong expresser, putative PG coregulation
2	CYP92A17/19	N responsive (down in 3200); strongest in leaves	Strong expresser, putative PG coregulation
3	CHATL3/6	N responsive (down in 3200); strongest in young leaves	Strong expresser, putative PG coregulation
4	MATL1/2/3	N responsive (down in 3200); strongest in young leaves	Strong expresser, putative PG coregulation
5	LAC2/3	N responsive; strongest in xylem	Previous literature indicates role in PG levels*
6	CHATL1/2	N responsive; strongest in leaves	Strong expresser, some indication of PG coregulation
7	CYP92A18	N responsive (down in 3200)	Microarray, similarity to other CYP92As
8	CYP92A20	N responsive	Microarray, similarity to other CYP92As
9	LAC1a/1b/1c/1d	N responsive; strongest in xylem; most strongly expressed in leaves among laccases	Strongly expressed in leaves, some indication of coregulation
10	MATL5	Strongest in young leaves	Strongly expressed in young leaves in *P. tremuloides*

Table A.2: RNAi Constructs Available for Transgenic Silencing Experiments of Candidate Genes Linked with Phenolic Metabolism in *Populus*.

*Labeled previously as MaTL8/9/10.

Gene(s)	Vectors Used	Status of pFGC	Status of pGSA
CHATL1/2	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence-confirmed in *E. coli*
CHATL3/6	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence -confirmed in *E. coli*
MATL1/2/3*	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence -confirmed in *E. coli*
MATL5	pFGC5941	sequence-confirmed in *Agrobacterium*	N/A
CYP92A17/19	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence -confirmed in *E. coli*
CYP92A18	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence -confirmed in *E. coli*
CYP92A20 (v2)	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence -confirmed in *E. coli*
CYP92A24/25	pFGC5941 & pGSA1285	sequence-confirmed in *Agrobacterium*	sequence -confirmed in *E. coli*
LAC1a/1b/1c/1d	pFGC5941	sequence-confirmed in *Agrobacterium*	N/A
CYP92A20 (v1)	pFGC5941 & pGSA1285	First fragment sequence-confirmed	sequence -confirmed in *E. coli*
LAC2/3	pFGC5941	First fragment sequence-confirmed (2 independent lines)	N/A

Figures

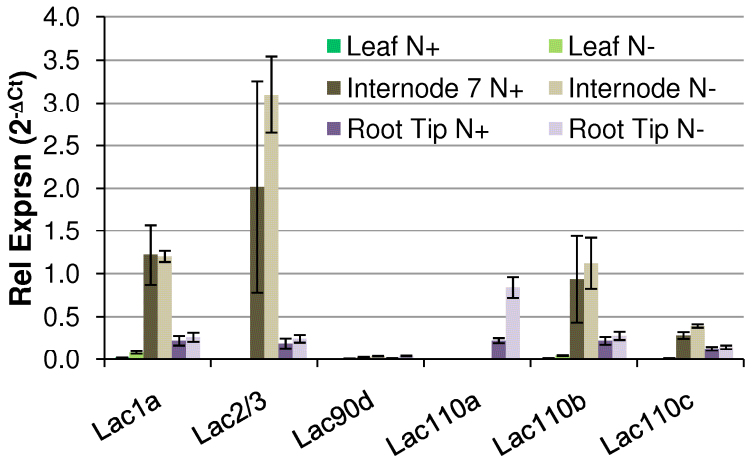

Figure A.1: Expression of Laccase Genes in *P. tremuloides* Genotype 271 Under Nitrogen-Replete (N+) and Nitrogen-Limited (N-) Conditions.

Data shown are means±SE relative to EF1β (n=3).

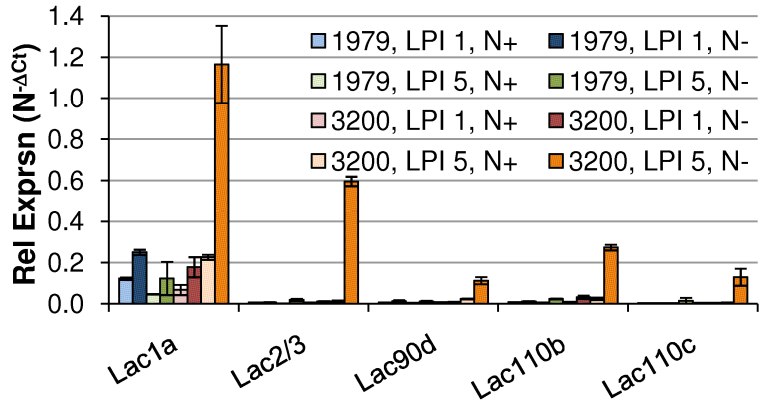

Figure A.2: Expression of Laccase Genes in Leaf Tissues of Two *P. fremontii* x *angustifolia* Hybrid Lines Under Nitrogen-Replete (N+) and Nitrogen-Limited (N-) Conditions.

Data shown are means+SD (n=2) relative to the geometric mean of EF1β and Cyp (genotype 1979), or EF1β, Cyp and UBCc (genotype 3200).

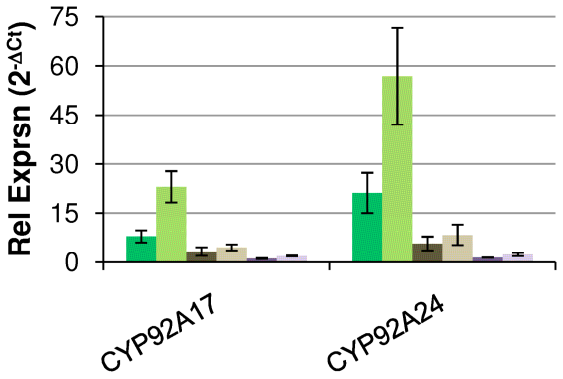

Figure A.3: Expression of CYP92A Genes in *P. tremuloides* Genotype 271 Under Nitrogen-Replete (N+) and Nitrogen-Limited (N-) Conditions.

Coloration of bars corresponds to samples as indicated in Figure A.1. Data shown are means±SE relative to EF1β (n=2 or 3).

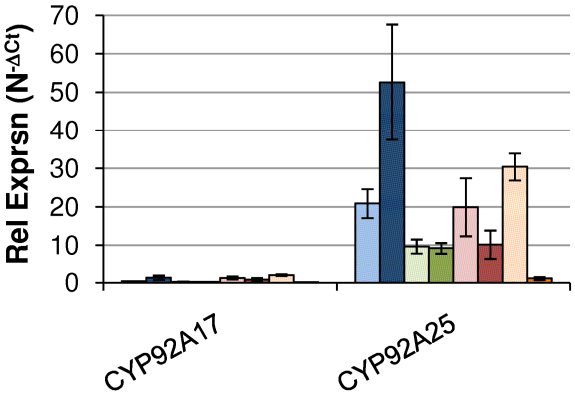

Figure A.4: Expression of CYP92A Genes in Leaf Tissues of Two *P. fremontii* x *angustifolia* Hybrid Lines Under Nitrogen-Replete (N+) and Nitrogen-Limited (N-) Conditions.

Coloration of bars corresponds to samples as indicated in Figure A.2. Data shown are means+SD (n=2) relative to the geometric mean of EF1β and Cyp (genotype 1979), or EF1β, Cyp and UBCc (genotype 3200).

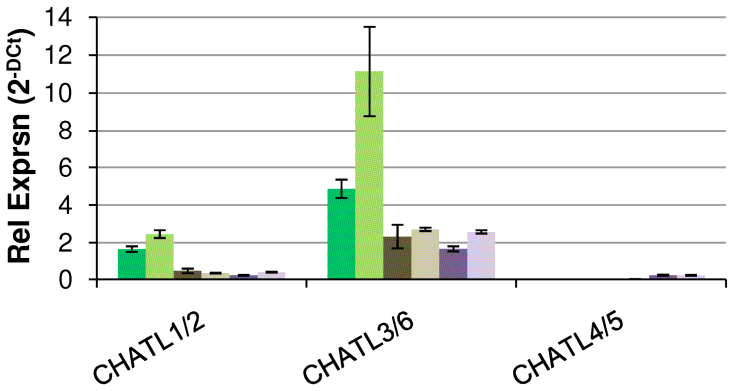

Figure A.5: Expression of CHATL Genes in *P. tremuloides* Genotype 271 Under Nitrogen-Replete (N+) and Nitrogen-Limited (N-) Conditions.

Coloration of bars corresponds to samples as indicated in Figure A.1. Data shown are means±SE relative to EF1β (n=3).

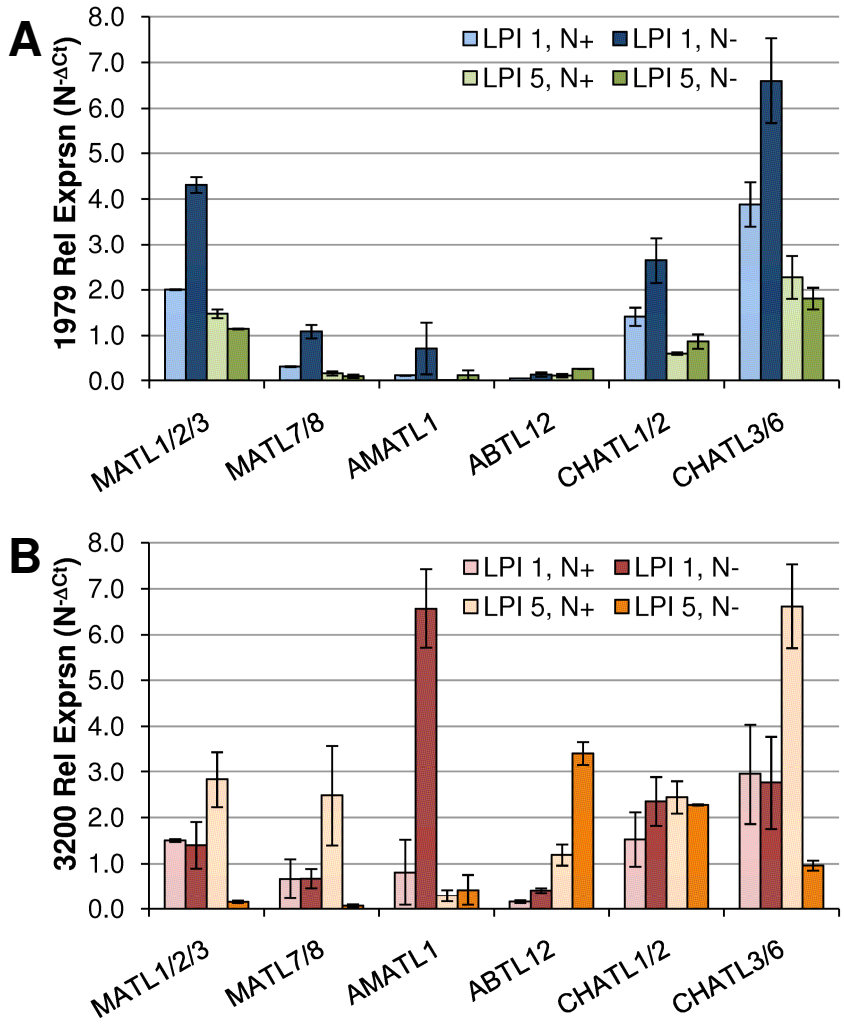

Figure A.6: Selected BAHD Acyltransferase Genes Well-Expressed in Leaves of Two *P. fremontii x angustifolia* Hybrid Lines and Response to Nitrogen Stress.

(A) Expression relative to the geometric mean of EF1β and Cyp in genotype 1979; (B) expression relative to the geometric mean of EF1β, Cyp and UBCc in genotype 3200. Data shown are means+SD (n=2). Any means below the detection limit of 0.01 are not shown.

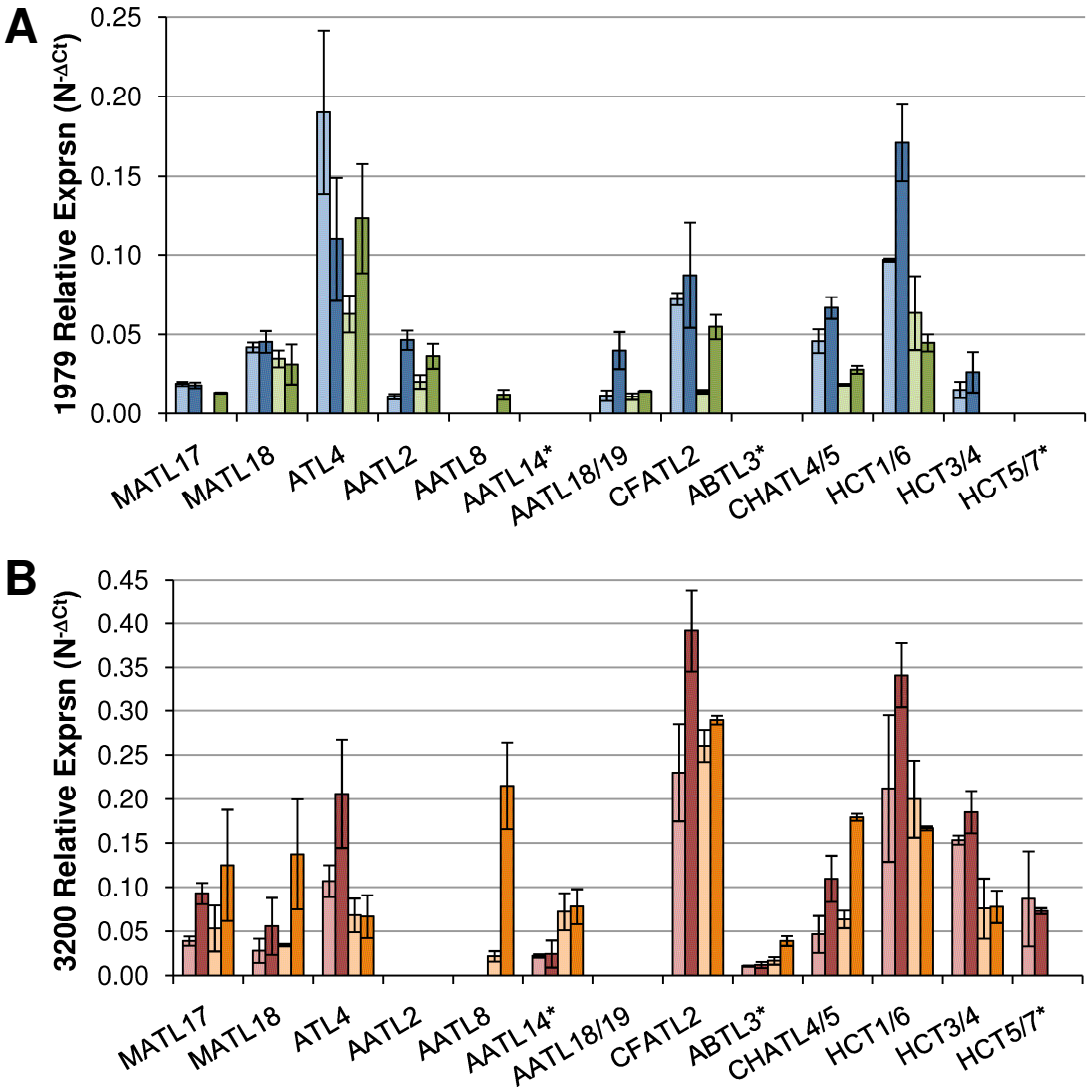

Figure A.7: Selected BAHD Acyltransferase Genes Weakly Expressed in Leaves of Two *P. fremontii x angustifolia* Hybrid Lines and Response to Nitrogen Stress.

(A) Expression relative to the geometric mean of EF1β and Cyp in genotype 1979; (B) expression relative to the geometric mean of EF1β, Cyp and UBCc in genotype 3200. Coloration of bars corresponds to samples as indicated in Figure A.6. Data shown are means+SD (n=2). Any means below the detection limit of 0.01 are not shown. *Below the detection limit in all tissues tested only for the genotype shown.

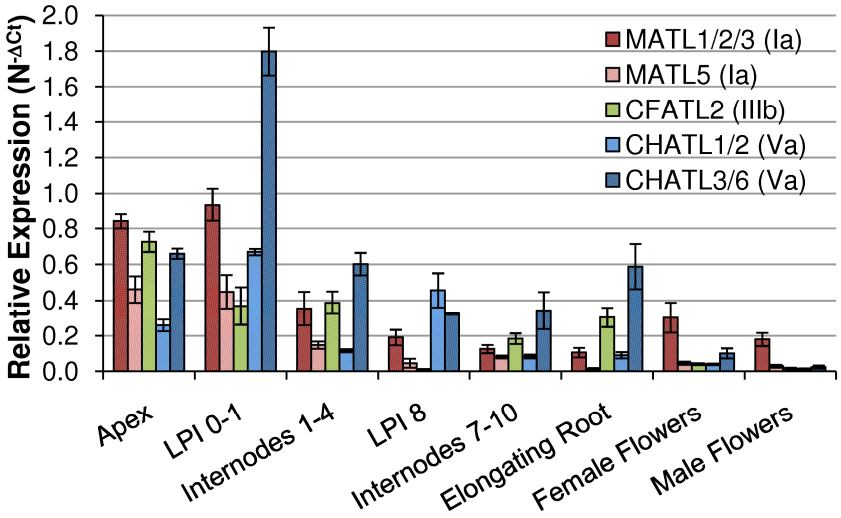

Figure A.8: Selected BAHD Acyltransferases Well-Expressed Across *P. tremuloides* Tissues.

Data shown are means±SE relative to the geometric mean of EF1β, Cyp, and UBCc (n=3). Data shown for CHATL genes are identical to those in **CHAPTER 2**.

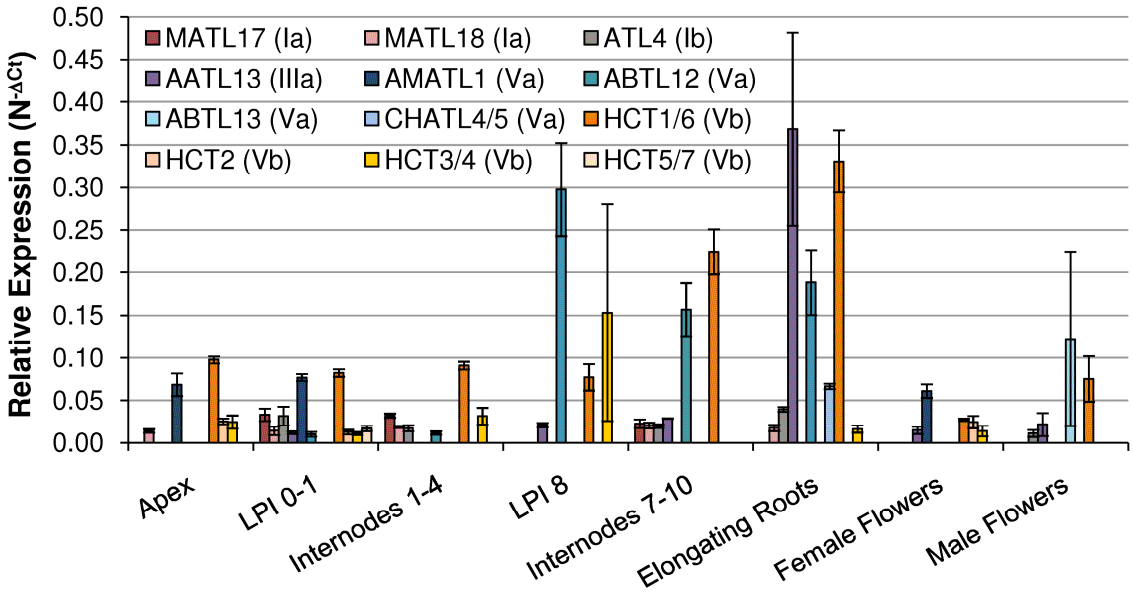

Figure A.9: Selected BAHD Acyltransferases Weakly Expressed Across *P. tremuloides* Tissues.

Data shown are means±SE relative to the geometric mean of EF1β, CYP, and UBCc (n=3); any means below the detection limit of 0.01 are not shown. Data for CHATL genes are identical to those in **CHAPTER 2**.

References

Ayabe S-i, Akashi T (2006) Cytochrome P450s in flavonoid metabolism. *Phytochemistry Reviews* 5: 271-282

Babst B, Harding S, Tsai C-J (2010) Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in *Populus*. *Journal of Chemical Ecology* 36: 286-297

Bao W, O'Malley DM, Whetten R, Sederoff RR (1993) A laccase associated with lignification in loblolly pine xylem. *Science* 260: 672-674

Berthet S, Demont-Caulet N, Pollet B, Bidzinski P, Cézard L, Le Bris P, Borrega N, Hervé J, Blondet E, Balzergue S, Lapierre C, Jouanin L (2011) Disruption of *LACCASE4* and *17* results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* 23: 1124-1137

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding *in vivo* benzenoid metabolism in Petunia petal tissue. *Plant Physiology* 135: 1993-2011

Cai X, Davis EJ, Ballif J, Liang M, Bushman E, Haroldsen V, Torabinejad J, Wu Y (2006) Mutant identification and characterization of the laccase gene family in *Arabidopsis*. *Journal of Experimental Botany* 57: 2563-2569

Czernic P, Chen Huang H, Marco Y (1996) Characterization of *hsr201* and *hsr515*, two tobacco genes preferentially expressed during the hypersensitive reaction provoked by phytopathogenic bacteria. *Plant Molecular Biology* 31: 255-265

Dai S-M, Zhou C-A, Xie B-Y, Xiao Q-M, Yang Y-H, Yin L-R (2007) Molecular cloning and sequence analysis of cytochrome P450 in pepper (*Capsicum annuum* L.) induced by β-aminobutyric acid (BABA). *Acta Horticulturae Sinica* 34: 637-643

Ehlting J, Hamberger B, Million-Rousseau R, Werck-Reichhart D (2006) Cytochromes P450 in phenolic metabolism. *Phytochemistry Reviews* 5: 239-270

Freudenberg K (1959) Biosynthesis and Constitution of Lignin. *Nature* 183: 1152-1155

Gavnholt B, Larsen K (2002) Molecular biology of plant laccases in relation to lignin formation. *Physiologia Plantarum* 116: 273-280

Gendler K, Paulsen T, Napoli C (2008) ChromDB: The Chromatin Database. [http://www.chromdb.org/index.html]

Harding SA, Jarvie MM, Lindroth RL, Tsai C-J (2009) A comparative analysis of phenylpropanoid metabolism, N utilization, and carbon partitioning in fast- and slow-growing *Populus* hybrid clones. *Journal of Experimental Botany* 60: 3443-3452

Holsters M, De Waele D, Depicker A, Messens E, van Montagu M, Schell J (1978) Transfection and transformation of *Agrobacterium tumefasciens*. *Molecular and General Genetics* 163: 181-187

Kang J-G, Yun J, Kim D-H, Chung K-S, Fujioka S, Kim J-I, Dae H-W, Yoshida S, Takatsuto S, Song P-S, Park C-M (2001) Light and brassinosteroid signals are integrated via a dark-induced small G protein in etiolated seedling growth. *Cell* 105: 625-636

Luo K, Harding SA, Tsai C-J (2008) A modified T-vector for simplified assembly of hairpin RNAi constructs. *Biotechnology Letters* 30: 1271-1274

Nelson D, Werck-Reichhart D (2011) A P450-centric view of plant evolution. *The Plant Journal* 66: 194-211

Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative genomics of Rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology* 135: 756-772

Nomura T, Bishop G (2006) Cytochrome P450s in plant steroid hormone synthesis and metabolism. *Phytochemistry Reviews* 5: 421-432

Orlova I, Marshall-Colon A, Schnepp J, Wood B, Varbanova M, Fridman E, Blakeslee JJ, Peer WA, Murphy AS, Rhodes D, Pichersky E, Dudareva N (2006) Reduction of benzenoid synthesis in Petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *Plant Cell* 18: 3458-3475

Persans MW, Wang J, Schuler MA (2001) Characterization of maize cytochrome P450 monooxygenases induced in response to safeners and bacterial pathogens. *Plant Physiology* 125: 1126-1138

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: 2966-2980

Ralston L, Kwon ST, Schoenbeck M, Ralston J, Schenk DJ, Coates RM, Chappell J (2001) Cloning, heterologous expression, and functional characterization of 5-*epi*-aristolochene-1,3-dihydroxylase from tobacco (*Nicotiana tabacum*). *Archives of Biochemistry and Biophysics* 393: 222-235

Ranocha P, Chabannes M, Chamayou S, Danoun Sd, Jauneau A, Boudet A-M, Goffner D (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiology* 129: 145-155

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. *Annual Review of Plant Biology* 54: 629-667

Tsai CJ, Cseke LJ, Harding SA (2003) Isolation and purification of RNA. In: Cseke LJ, Kaufman PB, Podila GK, Tsai CJ (eds) *Handbook of Molecular and Cellular Methods in Biology and Medicine*. CRC Press, Boca Raton, FL, pp 25-44

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan YN (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytologist* 172: 47-62

Zenk MH (1967) Pathways of salicyl alcohol and salicin formation in *Salix purpurea* L. *Phytochemistry* 6: 245-252

APPENDIX B.

EFFORT TOWARDS GENERATING RECOMBINANT PROTEINS OF FOUR PARALOGOUS *POPULUS* BAHD ACYLTRANSFERASES

Background

A logical step subsequent to the identification of paralogous *Populus* BAHD acyltransferase genes with differential expression patterns in different tissues, genotypes, or stress conditions is to investigate the functions of their encoded proteins. At least two approaches can be utilized to address this interest. The first approach, altering the expression of specific paralogous genes to determine their effects *in vivo*, was outlined in Appendix A. A second approach, outlined in this appendix, is to conduct *in vitro* biochemical assays of recombinant proteins. Although technical considerations and prioritization have led to this work being discontinued as part of the dissertation research, I outline here the progress I have made in cloning and generating recombinant proteins for the *Populus* BAHD acyltransferase genes *CHATL1*, *CHATL2*, *CHATL3*, and *CHATL6*.

It is worth noting that differences in expression of paralogous genes in no way predict differences in biochemical function in and of themselves. The sequence similarity of the four proteins of interest ranges from 81 to 92% (**CHAPTER 2**), suggesting some level of functional conservation is likely. Nevertheless, substrate specificity of metabolic pathway enzymes can be altered by changes in a single amino

acid (Kampranis et al. 2007), and BAHD enzyme substrate specificity has previously been shown to differ in homologues differing by only seven residues (Unno et al. 2007), so biochemical analysis is still a potentially informative step in understanding the processes by which secondary metabolic pathway genes undergo functional divergence. In this context, investigating the biochemical function of the four *CHATL* genes is of particular interest *because* (rather than in spite) of their sequence similarity. While differences in enzyme biochemical properties would be a clear indication of functional divergence, similar enzyme specificity among the four CHATLs would support a case of functional divergence via spatiotemporal regulation. In either case, transgenic manipulation by overexpression or silencing of *CHATL* genes would provide complementary information to help clarify the functional importance of the four paralogs *in vivo*.

The rationale for choosing these four genes as candidates for biochemical assays is twofold. First, as outlined above and in **CHAPTER 2**, *CHATL 1*, *2*, 3, and *6* form a small, *Populus*-specific phylogenetic cluster, with the pairs *CHATL 1* & *2* and *CHATL 3* & *6* more closely related to each other than other possible combinations of the genes. They appear to have originated from a single salicoid duplication (*CHATL3* and *6*), followed by tandem duplications (*CHATL1-3*) on linkage group (LG) XIII. *CHATL 1* & *2* appear to be specific to green tissues, while *CHATL 3* & *6* also exhibit expression in roots. The two pairs respond differentially to nitrogen depletion stress in the *P. fremontii* x *angustifoli* genotype 1979, with *CHATL1* and *CHATL2* further differing in their responses to nitrogen stress in hybrid genotype 3200. These differential expression patterns provide the first rationale for biochemical study. The second reason is their

inclusion within a poorly-resolved branch of the BAHD acyltransferase phylogeny, within Clade Va. The most closely-related member of this branch that has been biochemically characterized is the *Arabidopsis* enzyme CHAT, which is involved in the biosynthesis of the green leaf volatile compound (Z)-3-hexen-1-yl acetate (D'Auria et al. 2002; D'Auria et al. 2007). Other characterized proteins within the poorly-resolved branch include those responsible for the synthesis of benzoate compounds, including benzylbenzoate (Boatright et al. 2004; D'Auria et al. 2002), and several proteins within this branch exhibit multifunctionality (Boatright et al. 2004; D'Auria et al. 2002; El-Sharkawy et al. 2005; Okada et al. 2005; Souleyre et al. 2005; Wang and De Luca 2005). Thus, the *CHATL* genes are potential candidates for involvement in the *Populus* benzoate pathway, one that may be connected to the biosynthesis of salicylate glycosides, secondary metabolites specific to the Salicaceae (Babst et al. 2010).

Methods

Primer Design

Predicted cDNA sequences corresponding to the gene loci fgenesh4_pm.C_LG_XIII000317 (*CHATL1*), eugene3.00130737 (*CHATL2*), eugene3.00130740 (*CHATL3*), and grail3.0065010701 (*CHATL6*) were extracted from the JGI *Populus trichocarpa* v.1.1 genome database (Tuskan et al. 2006). Two primer pairs were designed using the PrimerQuest^{SM} tool (Integrated DNA Technologies Inc. 2011) for full-length cDNA cloning by two different strategies, TA Cloning and TOPO Cloning, as outlined below. For the TA Cloning primers, two restriction sites were

appended to the first primer set (underlined in Table B.1) to facilitate later subcloning. The *Bam*H I site is attached to the forward primer and the *Eco*R I restriction site to the reverse primer. For the TOPO Cloning primers, the sequence CACC was appended to the forward primer in compliance with cloning requirements for the TOPO vector (Invitrogen). When some genes proved recalcitrant to full-length cloning, internal primers were also designed for partial fragment amplification for use in overlapping PCR-based cloning. All primers are listed in Table B.1.

Template Preparation

Two ESTs available in our laboratory, MTUNUL1.P47.A10 and MTUNUL1.P35.B11, derived from *P. fremontii* x *angustifolia* genotype NUL, were used as templates for coding sequence amplification of *CHATL3* and *CHATL6*, respectively. Glycerol stocks were streaked on agar plates containing appropriate antibiotics, and incubated at 37^{o}C overnight. Single colonies were grown in liquid media with appropriate antibiotics overnight. Plasmid DNAs were extracted from the overnight liquid cultures by standard miniprep (Ausubel et al. 1992). Template for PCR amplification of *CHATL1* and *CHATL2* was cDNA generated by reverse transcription of total RNA extracted from mature leaf tissues (Leaf Plastochron Index [LPI] 3/4) of *P. fremontii* x *angustifolia* genotype 1979, as described in **APPENDIX A**.

TA Strategy: PCR Amplification, TA Cloning and Subcloning to the pET30(a)+ Vector

PCR was performed via a nested strategy for TA cloning using 20-41 ng of leaf cDNA (for *CHATL1* and *CHATL2*) or 5 ng of plasmid DNA (for *CHATL3* and *CHATL6*)

as the template (Table B.2) and 10 μM of gene-specific forward and oligo dT (XhoI-dT20VN) reverse primer, along with either 1 U of Phusion® high-fidelity DNA polymerase (Finnzymes) or 1 U of RedTaq® DNA Polymerase (Sigma-Aldrich Co.) supplemented with up to 0.8 U of Phusion®. Amplicons were gel purified using the UltraClean® GelSpin® DNA Extraction kit (MO BIO Laboratories, Inc.), then used as template for PCR of the coding region using TA Strategy primers. As an alternative approach, semispecific (i.e., matching two of the four paralogues) internal primers were used in conjunction with the TA Cloning forward and reverse primers for amplification of two separate but overlapping 5'- and 3'-partial fragments, respectively, using template as described above. Overlapping PCR was conducted by combining 2.0 μl template purified using the UltraClean® GelSpin® DNA Extraction kit and gene-specific primers. PCR products were analyzed by agarose gel electrophoresis. Full-length, cleaned products were incubated with RedTaq® and dNTPs for 20 minutes at 72°C to add 3'-A overhangs for TA cloning to the pCR®II vector according to manufacturers' instructions (Invitrogen).

Amplicons generated from TA Cloning Strategy procedures were ligated to the TA cloning vector pCR®II (Invitrogen). The constructs were transformed into *E. coli* TOP10F', and transformants were confirmed using colony PCR. Two positive colonies per clone were cultured in 5 mL LB with antibiotic at 37°C overnight, with 4.5 mL of each culture used for plasmid DNA extraction. A portion of the isolated plasmid DNA was used for clone confirmation by Sanger sequencing using vector-specific primers and BigDye XTerminator® kit (Applied Biosystems) according to manufacturer's instructions on an ABI 3730 automated sequencer (Georgia Genomics Facility, The University of

Georgia). Sequence-confirmed clones were digested using *Bam*H I and *Eco*R I restriction enzymes to release the full-length protein coding region. After purification using the UltraClean® GelSpin® DNA Extraction kit, these fragments were subcloned into the pET30(a)+ protein expression vector (Novagen, Inc.) predigested with *Bam*H I and *Eco*R I restriction enzymes. I also attempted to digest *CHATL* amplicon ends with restriction enzymes and directly ligate these amplicons into pET30a(+). The ligation mixtures were transformed into *E. coli* TOP10 competent cells and transformants confirmed using colony PCR as described above.

TOPO Strategy: PCR Amplification and Direct Cloning to the pET160/GW/D-TOPO® Expression Vector

CHATL cDNAs were PCR-amplified using 20-41 ng (*CHATL1 & 2*) or 15 ng (*CHATL3 & 6*) template DNA previously described, gene-specific primers designed for the TOPO Cloning Strategy, and 1 U Phusion® (supplemented with up to 0.5 U RedTaq®) enzyme in a single reaction (Table B.2). Initial attempts were made to conduct full-length cloning, with difficulties overcome by cloning 5' and 3' fragments separately and conducting a second, overlapping PCR reaction. Amplicons were successfully generated for all four genes using this method. Cleaned PCR products were cloned into the pET160/GW/D-TOPO® expression vector following the manufacturers' instructions (Invitrogen). The ligation mixtures were transformed into *E. coli* TOP10 and transformants confirmed using colony PCR and Sanger sequencing as described above.

Recombinant Protein Induction

Sequence-confirmed protein expression vectors containing full-length *CHATL* cDNAs were transformed into *E. coli* strain BL21 (DE3) pLysS. A biochemically characterized benzoate:CoA ligase (BZL) in the protein expression vector pPE204, kindly provided by Dr. Till Beuerle (Beuerle and Pichersky 2002), was separately transformed into the same *E. coli* strain for use as a positive control in protein induction experiments. For each construct, ten colonies were grown up individually overnight on a 250 rpm shaker at 37°C in 5 mL LB containing 100 µM ampicillin. The following morning, each sample was subcultured 1:50 in 5 mL LB with 100 µM ampicillin and grown for 2 h at 37°C. A replicate culture was prepared for one randomly selected sample per construct. Protein induction was initiated by the addition of 0.8 mM IPTG to all samples except the random replicates, which acted as non-induced controls. The incubation temperature was reduced to 23.5°C, and the cultures were induced for 6 to 24 h, after which 0.5 to 1.0 mL of sample was taken for total protein extraction and SDS-PAGE to check the presence and extent of induction of the desired proteins. This procedure is the same as that published for BZL induction, which used a 24h final induction period (Beuerle and Pichersky 2002).

An alternative induction procedure was also used for *CHATL* constructs only. In this case, overnight-grown cultures were subcultured 1:50 in 5 mL LB containing 100 µM ampicillin. The cultures were grown for 2.5 h at 37°C and 200 rpm, then moved to an 80 rpm shaker at 19°C. After 2 h, a 150 µL aliquot was taken from each sample as a non-induced control, then 0.1 mM IPTG was added to the remainder for protein induction. After 24 h, 1.0 mL of each culture was taken for total protein extraction and SDS-PAGE.

Protein Extraction, Quantification, and SDS-PAGE

Sample aliquots from induced and non-induced cultures were extracted using BugBuster® Master Mix (Novagen) according to the manufacturer's protocol for small-scale extraction. Upon cell lysis, soluble proteins were obtained by centrifugation. The remaining pellets were subjected to inclusion body preparation according to the manufacturer's protocol. No protease inhibitors were added during either procedure. Protein concentration was estimated using the BCA Protein Assay kit (Pierce) with a microplate reader, referenced against an eight-point standard curve of bovine serum albumin ranging from 25 to 2000 μg/mL. Protein extracts were analyzed by SDS-PAGE using discontinuous gels consisting of a 5% acrylamide stacking gel and a 10% acrylamide separating gel (Ausubel et al. 1992). Each gel contained at least one lane for a non-induced control to allow baseline comparison. Gels were stained with Coomassie Blue for protein visualization.

Results

cDNA Amplification

Predicted cDNA sequences corresponding to each of the four *Populus* v.1.1 *CHATL* gene models were subjected to BLASTN searches of *P. tremuloides* and *P. fremontii* x *angustifolia* EST databases on the AspenDB website (AspenDB 2011). Two putative EST matches were identified: MTUNUL1.P47.A10 for *CHATL3* (98% sequence identity over the first 826 nt listed in the database relative to a total length of 1380 nt), and MTUNUL1.P35.B11 for *CHATL6* (99% sequence identity over the 870 nt

listed in the database relative to a total length of 1383 nt). Both of these ESTs were derived from a full-length cDNA library made from leaf tissue of *P. fremontii* x *angustifolia* genotype NUL. The coding sequences were successfully PCR-amplified using both TA and TOPO Cloning strategies, although *CHATL6* required an overlapping PCR approach for successful amplification using TOPO strategy primers (Table B.2). *CHATL1* and *CHATL2* cDNAs were also successfully amplified using both TA and TOPO strategies, although overlapping PCR was required for amplification except in the case of the TOPO strategy primers for *CHATL1* (Table B.2). Images of full-length amplicons and, where appropriate, partial-length amplicons used for overlapping PCR are also shown by cloning strategy (Figures B.1 & B.2).

Cloning and Sequence Confirmation

CHATL6 cDNA was successfully cloned into pCRII TA vector using the TA Cloning strategy, and has been fully sequenced to confirm authenticity (Tables B.3 & B.4). The imputed full-length protein sequence has 98% similarity to CHATL6 from the *Populus trichocarpa* v1.1 genome; it was 81%, 81%, and 90% similar to CHATLs 1, 2, and 3 respectively. A full-length sequence of *CHATL3* was also obtained, but the flanking vector sequences were incorrect (data not shown), suggesting contamination of the EST plasmid rather than a successful TA cloning event. Due to difficulties with digestion of TA-cloned *CHATL6*, all four genes were also putatively cloned into the pET30(a)+ expression vector via direct digestion of amplicons, but these have not been sequence confirmed (Table B.3). *CHATL2* and *CHATL3* were also cloned into the pET160 expression vector and sequence confirmed (Tables B.3, B.5, B.6). The imputed

full-length protein sequence for CHATL2 has 98% similarity to the corresponding sequence from *P. trichocarpa*; it was 91%, 81%, and 80% similar to CHATLs 1, 3, and 6 respectively.The imputed full-length protein sequence for CHATL3 has 97% similarity to the corresponding *P. trichocarpa* sequence; it was 81%, 81%, and 90% similar to CHATLs 1, 2, and 6 respectively.

Recombinant Protein Generation

Successful BZL protein production was achieved by following previously published procedures (Beuerle and Pichersky 2002), with abundant expression of BZL evident in the soluble fraction of protein extract in nine of the ten colonies tested (Figure B.3A; samples I and J not shown). The recombinant protein bands were between the 58 and 46 kDa molecular markers, smaller than the previously estimated size of purified BZL at 61.5 kDa (Beuerle and Pichersky 2002). The discrepancy was probably due to different storage solvents in which the molecular mass markers and sample extracts were prepared. No clear difference was observed between soluble fractions of induced and non-induced samples of *CHATL2* and *CHATL3* (Figure B.3; samples F-I not shown for either gene). When the insoluble fractions were analyzed, a clear band was visible for most of the induced *CHATL3* samples (Figure B.4), indicative of successful recombinant protein induction in the form of inclusion body. As for BZL, the *CHATL3* bands were shifted downwards from the expected size of about 55 kDa. No expression of *CHATL2* was detected in the inclusion body extracts (Figure B.4A), although a band with a much larger molecular weight was observed in one of the *CHATL2* inclusion body fractions from a separate trial (Figure B.4C)

Discussion and Future Directions

E. coli expression constructs were successfully prepared for two of the four target genes, *CHATL2* and *CHATL3*. Although *CHATL1* and *CHATL6* pET30(a)+ constructs were also putatively generated, sequencing is required to determine if they match to the intended genes; partial sequencing of other constructs suggested that *CHATL2* or *CHATL3* were cloned instead. Thus, further cloning may be required to generate the *CHATL1* and *CHATL6* constructs. It should be noted that *CHATL2* and *3* are found adjacent to each other on a single LG, but phylogenetically they are less closely related to each other than to *CHATL1* and *CHATL6*, respectively (**CHAPTER 2**). Thus, these clones provide a good start for dissecting possible functional differences between the two pairs of paralogues and might be considered analogous to the RNAi constructs generated to silence both *CHATL1 & 2* and *CHATL3 & 6*, respectively, in terms of their abilities and limitations in determining functional similarities and differences among *CHATL* paralogues (**APPENDIX A**). Preliminary protein expression experiments confirmed the successful production of CHATL3 recombinant protein, though in the form of inclusion body. No recombinant protein induction was observed for CHATL2 under the same experimental conditions. Future optimization to improve the solubility of CHATL3 and to achieve induction for CHATL2 could include longer induction periods, altered induction temperature, or reduced rates of orbital shaking. Successful preparation of CHATL2 and CHATL3 recombinant proteins would allow for biochemical characterization of these two proteins, while successful preparation of BZL recombinant proteins would facilitate in-house production of benzoyl-CoA (Beuerle and Pichersky 2002), a common BAHD acyltransferase donor commercially available only at high cost.

Tables

Table B.1: Primers used for full-length cloning of four paralogous *CHATL* genes.

Code listed in parentheses below gene indicates the corresponding "BPBTL" coding from previous in house annotation. Underlined sequences indicate restriction sites for TA cloning or lead sequence for TOPO cloning; bold positions indicate wobbles added to amplify both genes in the pair.

Gene	Forward	Reverse	Purpose
CHATL1 **(B4)**	GGGATCCGCTTCATCACCCTCTTCT	GGAATTCTCAGCCCTTTAGCATGCC	TA Cloning strategy; full-length if used together, for overlapping PCR if used with internal primers
CHATL1	CACCGCTTCATCACCCTCTTCTCTA	TCAGCCCTTTAGCATGCCTTCAAGC	TOPO Cloning strategy; full-length if used together, for overlapping PCR if used with internal primers
CHATL1	AGTT**W**C**W**TGCAGACACCGAGCGTA	AGGGCAATGG**Y****M**CGACATATCCAA	Internal primers for cloning of 3' (forward primer) and 5' (reverse primer) fragments for overlapping PCR, used in conjunction with the above primers
CHATL2 **(B3)**	GGGATCCGTTTCATCACCCTCTTCT	GGAATTCTTATAAAGAAGATACGATTA	TA Cloning, full-length if used together
CHATL2	CACCGTTTCATCACCCTCTTCTCTA	TTATAAAGAAGATACGATTAACTTGGA	TOPO Cloning, full length if used together

Table B.1, Continued:

Gene	Forward	Reverse	Purpose
CHATL2	AGTT**W**C**W**TGCAGACACCGAGCGTA	AGGGCAATGG**Y**M**R**CGACATATCCAA	Internal primers for cloning of 3' and 5' fragments, respectively
CHATL3 **(B2)**	TGGATCCGCTTCATCACCCGCTTCT	GGAATTCTTATAAGGAAGACGCGAT	TA Cloning, full-length if used together
CHATL3	CACCGCTTCATCACCCGCTTCTCTG	TTATAAGGAAGACGCGATAAATTTGGA	TOPO Cloning, full length if used together
CHATL3	TGGCCCTTCAGA**R**AT**R**TCTGCTCT	AGTCACGTCAGCCTT**R**GCCTTTCT	Internal primers for cloning of 3' and 5' fragments, respectively
CHATL6 **(B1)**	TGGATCCGCATCATCACCCGCCTCT	GGAATTCTTATAGGGAAGATACAAT	TA Cloning, full-length if used together
CHATL6	CACCGCATCATCACCCGCCTCTCTG	TTATAGGGAAGATACAATAAATTTGGA	TOPO Cloning, full length if used together
CHATL6	TGGCCCTTCAGA**R**AT**R**TCTGCTCT	AGTCACGTCAGCCTT**R**GCCTTTCT	Internal primers for cloning of 3' and 5' fragments, respectively

Table B.2: Conditions for successful amplification of four *CHATL* genes.

Gene	Strategy	DNA Source/Quantity	Enzyme/Reaction Vol.	Thermal Cycling
CHATL1 **(B4)**	TA	20 ng RNA equivalent of *P. fremontii x angustifolia* cDNA 20 ng RNA equivalent of *P. fremontii x angustifolia* cDNA 2.0 µl each of 5' Frag and 3' Frag	RedTaq/50.4 µl (incl. 0.8 U Phusion) RedTaq/50.4 µl (incl. 0.8 U Phusion) RedTaq/39.3 µl (incl. 0.6 U Phusion)	**5' Frag**: 94°C-5 min; 37x[94°C-30s, 45°C-30s, 72°C-60s]; 72°C-7 min **3' Frag**: 94°C-5 min; 37x[94°C-30s, 45°C-30s, 72°C-60s]; 72°C-7 min **Overlap**: 94°C-5 min; 32x[94°C-30s, 45°C-30s, 72°C-90s]; 72°C-7 min
CHATL1	TOPO	23 ng RNA equivalent of *P. fremontii x angustifolia* cDNA	Phusion/50.1 µl	98°C-30s; 38x[98°C-10s, 55.4°C-20s, 72°C-45s]; 72°C-7 min
CHATL2 **(B3)**	TA	20 ng RNA equivalent of *P. fremontii x angustifolia* cDNA 41 ng RNA equivalent of *P. fremontii x angustifolia* cDNA 2.0 µl each of 5' Frag and 3' Frag	RedTaq/50.4 µl (incl. 0.8 U Phusion) RedTaq/50.4 µl (incl. 0.8 U Phusion) RedTaq/39.3 µl (incl. 0.6 U Phusion)	**5' Frag**: 94°C-5 min; 37x[94°C-30s, 45°C-30s, 72°C-60s]; 72°C-7 min **3' Frag**: 94°C-5 min; 35x[94°C-30s, 45°C-30s, 72°C-90s]; 72°C-7 min **Overlap**: 94°C-5 min; 32x[94°C-30s, 45°C-30s, 72°C-90s]; 72°C-7 min
CHATL2	TOPO	2.0 µl each of 5' Frag and 3' Frag from TA Strategy	Phusion/50 µl (incl. 0.5 U RedTaq)	**Overlap**: 98°C-30s; 38x[98°C-10s, 52°C-20s, 72°C-45s]; 72°C-7 min
CHATL3 **(B2)**	TA	5 ng of EST plasmid MTUNUL1.P47.A10	Phusion/50 µl	98°C-30s; 33x[98°C-10s, 55°C-30s, 72°C-45s]; 72°C-7 min [Plus 72°C-20 min in RedTaq after cleanup to add sticky ends]

Table B.2, Continued:

Gene	Strategy	DNA Source/Quantity	Enzyme/Reaction Vol.	Thermal Cycling
CHATL3	TOPO	15 ng of EST plasmid MTUNUL1.P47.A10	Phusion/50.1 µl	98°C-30s; 38x[98°C-10s, 55.4°C-20s, 72°C-45s]; 72°C-7 min
CHATL6 **(B1)**	TA	5 ng of EST plasmid MTUNUL1.P35.B11	Phusion/50 µl	98°C-30s; 33x[98°C-10s, 55°C-30s, 72°C-45s]; 72°C-7 min [Plus 72°C-20 min in RedTaq after cleanup to add sticky ends]
CHATL6	TOPO	15 ng of EST plasmid MTUNUL1.P35.B11	Phusion/50.1 µl	**5' Frag**: 98°C-30s; 38x[98°C-10s, 60.4°C-20s, 72°C-30s]; 72°C-7 min
		15 ng of EST plasmid MTUNUL1.P35.B11	Phusion/50.1 µl	**3' Frag**: 98°C-30s; 38x[98°C-10s, 53.6°C-20s, 72°C-30s]; 72°C-7 min
		2.0 µl each of 5' Frag and 3' Frag	Phusion/50 µl (incl. 0.5 U RedTaq)	**Overlap**: 98°C-30s; 38x[98°C-10s, 53.6°C-20s, 72°C-45s]; 72°C-7 min

Table B.3: *CHATL* clones obtained in this study.

Clone name shows name listed on lab stocks and correspond to an older annotation system than that used in **CHAPTER 2**. [a]Full-length protein coding region has been confirmed by Sanger sequencing. [b]Partial confirmation by Sanger sequencing. [c]Multiple glycerol stocks from different colonies, not sequence-confirmed. [d]*BZL* in an expression-ready strain.

Name	Gene	Vector	*E. coli* strain
pCRII:BPBTL1-D	*CHATL6*[a]	TA Cloning	TOP10F'
pET30a(+):BPBTL1-E, F[c]	*CHATL6*	Protein Expression	TOP10
pET30a(+):BPBTL2-A, B[c]	*CHATL3*	Protein Expression	TOP10
pET30a(+):BPBTL3-A, B[c]	*CHATL2*	Protein Expression	TOP10
pET30a(+):BPBTL4-A, B[c]	*CHATL1*	Protein Expression	TOP10
pET160:BPBTL1-H	*CHATL3*[a]	Protein Expression	TOP10
pET160:BPBTL2-A	*CHATL3*[b]	Protein Expression	TOP10
pET160:BPBTL3-A	*CHATL2*[a]	Protein Expression	TOP10
pET160:BPBTL4-K	*CHATL2*[b]	Protein Expression	TOP10
pET160:B1H-A, D, E[c]	*CHATL3*	Protein Expression	BL21 (DE3) pLysS
pET160:B3A-A, D, E[c]	*CHATL2*	Protein Expression	BL21 (DE3) pLysS
pCRT7/CT-TOPO:BZL-E, G, J[c,d]	*BZL*	Protein Expression	BL21 (DE3) pLysS

Table B.4: Full-length gene and imputed protein sequences of PfaCHATL6 from Sanger sequencing of clone pCRII:BPBTL1-D.

Initial ATG nucleotides were excluded in order to ensure proper reading frame in the pET30a(+) vector. Imputed protein sequence is referenced against CHATL6 as found in the JGI *Populus trichocarpa* v.1.1 genome database. Positions listed in blue match CHATL 1, 2, or 3 reference sequences; positions listed in red are different from any of the four CHATL reference sequences.

```
>BPBTL1-D-consensus
GCATCATCACCCGCCTCTCTGTTGTTCAAAGTTCACAGACGTGAACCAGAACTGATCAAGCCTGCCAAGCCCACCCCACATGAGTTCAAGCTATTATCTGA
CATCGATGACCAAGAAGGTCTTCGATTCCACATTCCAGTCATGCAATTTTATCGCAACAATCCCTCTATGCAAGGGAAAGACCCCGTCAAAATCATTAGGG
AGGCACTCGCCAAAACATTAGTGTTTTACTATCCATTTGCCGGTAGACTCAGGGAAGGGCCTAACCGCAAGCTCATGGTGGAATGTACCGGTGAGGGTATC
TTGTTCATAGAAGCTGATGCCGATGTTACACTTGAGCAATTCGGTGATGCGCTTCAACCACCTTTCCCCTGCTTGGAGGAGCTCCTCTTTGATGTCCCTGG
CTCTAGTGGGGTGTTGAACTGCCCTTTGTTGCTTATTCAGGTGACGCGCCTCAAGTGTGGTGGTTTTCTGTTTGCCCTCCGCCTGAACCATACCATGAGTG
ATGCCGTAGGCCTAGTCCAATTCATGGCAGCAGTGGGTGAGATGGCACGGGGAGCCAATGCGCCCTCCGTCCCAGCTGTGTGGGAAAGACAAGTTCTCAAT
GCTAGTAACCCTCCACGAGTTACATGCACACACCGTGAGTACGAGGAGGTAGCTGACACCAAGGGTACCATTATTCCACTTGATGATATGGCTCATCGTTC
CTTCTTCTTTGGCCCTTCAGAAATGTCTGCTCTTCGAAAATTTGTCCCGCCTCACCTTAGCCATTGTTCTACTTTCGAAATTCTAACAGCATGTCTTTGGA
AATGTCGTACCATTGCCCTCCAACCAGATCCTACCGAGGAGATGCGCATACTATGCATTGTCAATGCTCGTGAGAAATTTAACCTTCCATTGCCAAGAGGA
TACTATGGTAATGGCTTTGCTTTTCCGGTTGCAGTGGCAACTGCGGGAGAACTCTCAAGAAATCCATTTGGATACGCCTTAGAATTGGTGAGAAAGGCTAA
GGCTGACGTGACTGAGGAATACATGCGATCAGTATCATCTTTGATGGTGATTAAGGGGAGGCCTCACTTTACAGTGGTAAGGGCATACCTAGTATCGGACC
TGAGACGTGCAGGATTCGAAGAGGTAGATTTTGGATGGGGTAATGCTATATATGGTGGCGCTGCCAAAGGTGGGGTTGGCGCCATCCCTGGAGTTGCAAGC

TTTTATATCCCATTTACAAACAAGAAAGGAGAAAATGGGGTTGTGGTCCCATTTTGCTTGCCGGCTCCTGCCATGGAAAGATTTGTCAAGGAGCTTGACGG
CATGTTGAAGGACGACCAGCCAGTTAGCGCGCAAACTAAGTCCAAATTTATTGTATCTTCCCTATAA
```

```
BPBTL1-D-c      -ASSPASLLFKVHRREPELIKPAKPTPHEFKLLSDIDDQEGLRFHIPVMQFYRNNPSMQG 59
CHATL6          MASSPASLLFKVHRREPELIKPAKPTPHEFKLLSDIDDQEGLRFHIPVMQFYRNNPSMQG 60
                 ************************************************************

BPBTL1-D-c      KDPVKIIREALAKTLVFYYPFAGRLREGPNRKLMVECTGEGILFIEADADVTLEQFGDAL 119
CHATL6          KDPVKIIREALAKTLVFYYPFAGRLREGPNRKLMVECTGEGILFIEADADVTLEQFGDAL 120
                ************************************************************
```

Table B.4, Continued:

```
BPBTL1-D-c      QPPFPCLEELLFDVPGSSGVLNCPLLLIQVTRLKCGGFLFALRLNHTMSDAVGLVQFMAA 179
CHATL6          QPPFPCLEELLFDVPGSSGVLNCPLLLIQVTRLKCGGFLFALRLNHTMSDAVGLVQFMAA 180
                ************************************************************

BPBTL1-D-c      VGEMARGANAPSVPAVWERQVLNASNPPRVTCTHREYEEVADTKGTIIPLDDMAHRSFFF 239
CHATL6          VGEMARGANAPSVPAVWERQVLNASDPPRVTCTHREYEEVADTKGTIIPLDDMAHRSFFF 240
                ************************:***********************************

BPBTL1-D-c      GPSEMSALRKFVPPHLSHCSTFEILTACLWKCRTIALQPDPTEEMRILCIVNAREKFNLP 299
CHATL6          GPSEMSALRKFVPPHLSHCSTFEILTACLWKCRTIALQPDPTEEMRILCIVNAREKFNPP 300
                ************************************************************ *

BPBTL1-D-c      LPRGYYGNGFAFPVAVATAGELSRNPFGYALELVRKAKADVTEEYMRSVSSLMVIKGRPH 359
CHATL6          LPRGYYGNGFAFPVAVATAEELSKNPFGYALELVRKAKADVTEEYMRSVSSLMVIKGRPH 360
                ******************* ***:************************************

BPBTL1-D-c      FTVVRAYLVSDLRRAGFEEVDFGWGNAIYGGAAKGGVGAIPGVASFYIPFTNKKGENGVV 419
CHATL6          FTVVRAYLVSDLRRAGFEEVDFGWGNAIYGGAAKGGVGAIPGVASFYIPFTNKKGENGVV 420
                ************************************************************

BPBTL1-D-c      VPFCLPAPAMERFVKELDGMLKDDQPVSAQTKSKFIVSSL- 459
CHATL6          VPFCLPAPAMERFVKELDGMLKDDQTVSAQTKSKFIVSSL- 460
                ************************.**************
```

Table B.5. Full-length gene and imputed protein sequences of PfaCHATL2 from Sanger sequencing of clone pET160:BPBTL3-A.

Initial ATG nucleotides were excluded in order to ensure proper reading frame in the pET160 expression vector. Imputed protein sequence is referenced against CHATL2 as found in the JGI *Populus trichocarpa* v.1.1 genome database. Positions highlighted in blue match CHATL1, 3, or 6 reference sequences; positions in red are different from any of the four reference sequences.

```
>B3A_imputed_060110
GTTTCATCACCCTCTTCTCTAGTTTTCAAAGTTCACAGACGTGAACCCGAGCTGATCAAACCAGCGAAGCCCACCCCACGTGAGTTCAAACTGTTATCTGA
CATTGATGACCAAGAAGGGCTTCGATTCCACGTTCCAATCATACAATTTTATCGCCACAATCCCTCAATGCACGGGAAAGACCCCGTCAAGGTCATCAGAG
AGGCAATTGCTAAAACATTAGTGTTTTACTATCCATTTGCCGGTAGGCTGAGGGAAGGGCATAACCGCAAGCTCATGGTGGAATGCACTGGCGAGGGTATC
TTGTTTATAGAGGCTGACGCTGATGTTACACTTGAGCAGTTTGGTGATCCACTTCAACCTCCATTTCCTTGCTTGGAGGAGCTCCTCTTTGATGTCCCTGG
CTCTAGCGGGGTGCTAAACTGCCCTCTGTTACTTATTCAGGTGTCACGGCTCAAGTGTGGTGGTTTTCTCTTTGCCCTCCGCCTCAACCATACCATGAGTG
ATGGCCCAGGATTAGAGCAATTCATGGCAGCGGTGGGTGAGATGGCCCGCGGAGCCAACGCCCCCTCTGTCCCTCCAGTGTGGGAAAGACATGTCCTTAAT
GCAACTGACCCACCTCGAGTTACATGCAGACACCGAGCGTACGAGGAGGTAGCTGGTTCGAAGAGCTCAATTCTTACACATGATCATCTGGTTCATCGTTC
ATTTTTCTTTAGCCCTTCAGATATAACTGCTCTTCGAAGATTGGTCCCACCTCACCTCGGCCATTGTTCTACTTTCGAAATATTAACGGCATGTCTTTGGA
TATGTCGGACCATTGCCCTCCAACCAGATCCTAATGAAGAAATGCGCGTAATTTGCCTCGTCAATGCACGTGAAAAATTTAACCCTCCATTATTACCAAGA
GGTTACTATGGTAATGGTTTTTTTCTTCTAGCAGCAGTAGCAACTGCAGGGGAACTTTCGAAAAAGCCAATTGGATATGCTTTGGAGCTGGTAAGGAAGGA
TAAGGCGGACATGACTGAGGAATACATGCGATCTACAGCATCTTTGATGGTGAGCAAGGGAAGGCCTCTTTTTACTGTGCCAGGGATCTACATAGTTTCGG
ACTTGAGACGTGCGGGACTTGAAAAGGCAGATTTCGGATGGGGAAATGCTATATATGCTGGTACTGCAAAAGCCATCCCTGAACTTGCAAGCTTTTATATT
CCGTTTACAAATAAGAAAGGAGAAGATGGGATCGTAGTACCATTTTGCTTGCCATCTCCTGCTGTGGAAAGATTTTACAAGGAGCTTGAAGGCATGCTAAA
GGGACAGCTAGTTAGTGGTGGAGCAAATTCCAAGTTAATAGTATTTTCTTTATAA

B3A_imputed_p_060110      -VSSPSSLVFKVHRREPELIKPAKPTPREFKLLSDIDDQEGLRFHVPIIQFYRHNPSMHG 59
CHATL2                    MVSSPSSLVFKVHRREPELIKPAKPTPHEFKLLSDIDDQEGLRFHVPIIQFYRHNPSMHG 60
                           **************************:******************************

B3A_imputed_p_060110      KDPVKVIREAIAKTLVFYYPFAGRLREGHNRKLMVECTGEGILFIEADADVTLEQFGDPL 119
CHATL2                    KDPVKVIREAIAKTLVFYYPFAGRLREGHNRKLMVECTGEGILFIEADADVTLEQFGDPL 120
                          ************************************************************
```

Table B.5, Continued:

```
B3A_imputed_p_060110      QPPFPCLEELLFDVPGSSGVLNCPLLLIQVSRLKCGGFLFALRLNHTMSDGPGLEQFMAA 179
CHATL2                    QPPFPCLEELLFDVPGSSGVLNCPLLLIQVSRLKCGGFLFALRINHTMSDGPGLVQFMAA 180
                          ******************************************:********** *****

B3A_imputed_p_060110      VGEMARGANAPSVPPVWERHVLNATDPPRVTCRHRAYEEVAGSKSSILTHDHLVHRSFFF 239
CHATL2                    VGEMARGANAPSVPPVWERHVLNATDPPRVTCRHRAYEEVAGSKSSILTHDHLVHRSFFF 240
                          ************************************************************

B3A_imputed_p_060110      SPSDITALRRLVPPHLGHCSTFEILTACLWICRTIALQPDPNEEMRVICLVNAREKFNPP 299
CHATL2                    SPSDITALRRLVPPHLSHCSTFEILTACLWICRTIALQPDPNEEMRVICLVNAREKFNPP 300
                          ****************.*******************************************

B3A_imputed_p_060110      LLPRGYYGNGFFLLAAVATAGELSKKPIGYALELVRKDKADMTEEYMRSTASLMVSKGRP 359
CHATL2                    LLPRGYYGNGFFLLAAVATAGELSKKPIGYALELVRKVKADMTEEHMRSTASLMVSKGRP 360
                          ********************************** *******:**************

B3A_imputed_p_060110      LFTVPGIYIVSDLRRAGLEKADFGWGNAIYAGTAKAIPELASFYIPFTNKKGEDGIVVPF 419
CHATL3                    LFTVPGTYIVSDLRRAGLEKADFGWGNAIYGGTAKAIPELASFYIPFTNKKGEDGIVVPF 420
                          ****** ***********************.****************************

B3A_imputed_p_060110      CLPSPAVERFYKELEGMLKGQLVSGGANSKLIVFSL 455
CHATL2                    CLPSPAVERFYKELEGMLKGQLVSGGANSKLIVSSL 456
                          ********************************* **
```

Table B.6: Full-length gene and imputed protein sequences of PfaCHATL 3 from Sanger sequencing of clone pET160:BPBTL1-H.

Initial ATG nucleotides were excluded in order to ensure proper reading frame in the pET160 vector. Imputed protein sequence is referenced against CHATL3 as found in the JGI *Populus trichocarpa* v.1.1 genome database. Positions listed in blue match CHATL 1, 2, or 6 reference sequences; positions listed in red are different from any of the four CHATL reference sequences.

```
>B1H_imputed_070210
GCATCATCACCCGCCTCTCTGGTTTTCAAAGTTCACAGACGTGAGCCCGAGCTGATCAAACCAGCGAAGCCCACCCCACATGAGTTCAAACTGTTATCTGA
CATTGATGACCAAGAAGGGCTTCGGTTCCACATTCCAGTCATACAATTCTATCGCCACAATCCATCAGTGCAAGGGAAAGACCCCGTCGAGGTCATCAGAG
AGGCAATTGCTAAAACATTGGTGTTTTACTATCCATTTGCCGGTAGGCTGAGGGAAGGGCAAAACCGCAAGCTCATGGTGGAATGCACTGGCGAGGGTATC
TTGTTTATAGAGGCTGACGCTGATGTTAAACTTGAGGAGTTTGGTGATGCACTTCAACCTCCATTTCCTTGCTTGGAAGAGCTCATCTTTGATGTCCCTGG
CTCTAGCGGGGTACTAAACTGCCCTCTGTTACTTATTCAGGTGACACGCCTCAAGTGTGGTGGTTTTATCTTTGGCCTTCGCCTCAATCATACCATGAGTG
ATGCCTCCGGCATAGTCCAATTCATGGCAGCGGTTGGTGAGATGGCACGCGGAGCCACTGCCCCCTCTGTCCCAGCTGTGTGGGAAAGGCATGTTCTGAAT
GCAAGAAACCCACCACGGGTTACATGCATACACCGTGAGTACGAGGAGGTAGCTGACACCAAGGGTACAATTATTCCACTTGATGATATGGCTCATCGTTC
CTTTTTCTTTGGCCCTTCAGAGATATCTGCTCTTCGAAAATTGATCCCGCCTCACCTTAGGCGTTGTTCCACTTTCGAAATATTAACAGCATGTCTTTGGA
CATGTGTTACCATTGCCCTCCAACCAGATCCTACTGAAGAGATGCGCATAATATGCATTGTCAATGCTCGTGAGAAATTTAACCCTCCATTACCAACTGGA
TACTACGGTAATGGCTTTGCTTTTCCGGTAGCAGTGGCAACTGCCGGGGAACTCTCGGAGAAGCCATTTGGATATGCCTTGGAATTGGTAAGAAAGGCCAA
GGCTGACGTGACTGAGGAATATATGCGATCGGTAGCATCTTTGATGGTAACTAAGGGGAGGCCTCATTTTACAGTGGTAAGGGCATACCTTGTATCGGACT
TAAGAAGTGCAGGATTCGAAGTGGTAGATTTCGGTTGGGGTAATGCTAAATACGGCGGGGCTGCCAAAGGTGGAGTTGGGGCAATCCCTGGAGTTGCAAGC
TTCTATATTCCATTTAAAAACAAAAAAGGAGAAAATGGGATTGTGGTGCCATTTTGCTTGCCAGCTCCTGCCATGGAAAGATTTGTCGAGGAGCTTGACGG
TATGTTGAAGGGCCAGCTACAAAGTGGGCAAACTCATTCCAAATTTATTGTATCTTCCCTATAA

B1H_imputed_p_070210      -ASSPASLVFKVHRREPELIKPAKPTPHEFKLLSDIDDQEGLRFHIPVIQFYRHNPSVQG 59
CHATL3                    MASSPASLVFKVHRREPELIKPAKPTPHEFKLLSDIDDQEGLRFHIPVIQFYRHNPSVQG 60
                           ************************************************************

B1H_imputed_p_070210      KDPVEVIREAIAKTLVFYYPFAGRLREGQNRKLMVECTGEGILFIEADADVKLEEFGDAL 119
CHATL3                    KDPVKVIREAIAKTLVFYYPFAGRLREGQNRKLMVECTGEGILFIEADADVTLEQFGDAL 120
                          ****:*********************************************.**:*****

B1H_imputed_p_070210      QPPFPCLEELIFDVPGSSGVLNCPLLLIQVTRLKCGGFIFGLRLNHTMSDASGIVQFMAA 179
CHATL3                    QPPFPCLEELIFDVPGSSGVLNCPLLLIQVTRLKCGGFIFGLRLNHTMSDASGIVQFMAA 180
                          ************************************************************
```

Table B.6, Continued:

```
B1H_imputed_p_070210      VGEMARGATAPSVPAVWERHVLNARNPPRVTCIHREYEEVADTKGTIIPLDDMAHRSFFF 239
CHATL3                    VGEMARGATTPSVPAVWERHVLNARNPPRVTCIHREYEEVADTNGTIIPLDDMAHRSFFF 240
                          *********:**********************************:****************

B1H_imputed_p_070210      GPSEISALRKLIPPHLRRCSTFEILTACLWTCVTIALQPDPTEEMRIICIVNAREKFNPP 299
CHATL3                    GPSEISALRKLIPPHLSRCSTFEILTACLWKCRTIALQPDPTEEMRIICIVNAREKFNPP 300
                          **************** *************.* ***************************

B1H_imputed_p_070210      LPTGYYGNGFAFPVAVATAGELSEKPFGYALELVRKAKADVTEEYMRSVASLMVTKGRPH 359
CHATL3                    LPTGYYGNGFAFPVAVATAGELSEKPFGYALELVRKAKADVTEEYMRSVASLMVTKGRPH 360
                          ************************************************************

B1H_imputed_p_070210      FTVVRAYLVSDLRSAGFEVVDFGWGNAKYGGAAKGGVGAIPGVASFYIPFKNKKGENGIV 419
CHATL3                    FTVVRAYLVSDLRSAGFEVVDFGWGNAIYGGAAKGGVGAIPGVASFLIPFKNKKGENGIV 420
                          *************************** ****************** *************

B1H_imputed_p_070210      VPFCLPAPAMERFVEELDGMLKGQLQSGQTHSKFIVSSL 458
CHATL3                    VPFCLPAPAMERFVEELDGMLKGQLQSGQTHSKFIASSL 459
                          ***********************************.***
```

Figures

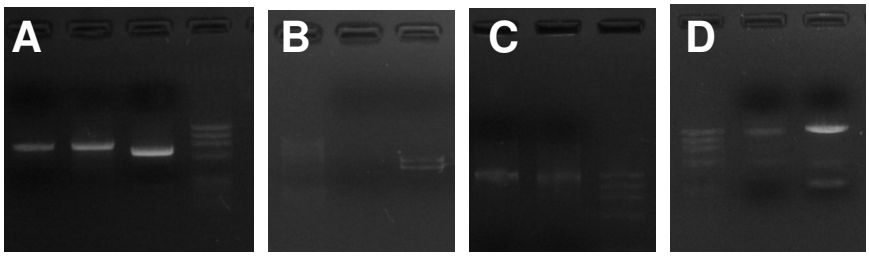

Figure B.1: Partial-length and full-length amplicons of *CHATL* cDNAs generated using TA Cloning strategy.

Lanes are listed in order from left to right for each panel: (A) *CHATL2*-5' fragment, *CHATL1*-5' fragment, *CHATL1*-3' fragment, and ΦX174 *Hae* III (molecular weight markers); (B) ΦX174 *Hae* III, failed reaction, and *CHATL2*-3' fragment; (C) full-length *CHATL2*, full-length *CHATL1*, and ΦX174 *Hae* III; (D) ΦX174 *Hae* III, full-length *CHATL6*, and full-length *CHATL3*.

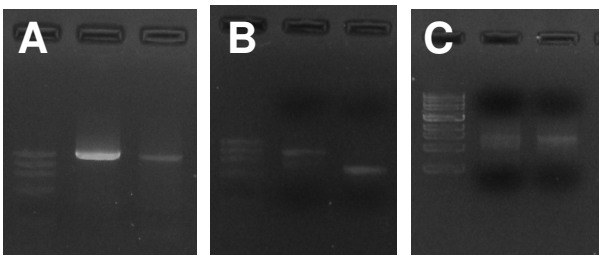

Figure B.2: Partial-length and full-length amplicons of *CHATL* cDNAs generated using TOPO Cloning strategy.

Lanes are listed in order from left to right for each panel: (A) ΦX174 *Hae* III, full-length *CHATL3*, and full-length *CHATL1*; (B) ΦX174 *Hae* III, *CHATL6*-5' fragment, and *CHATL6*-3' fragment; (C) λ *Hind* III (reference for band size), full-length *CHATL6*, and full-length *CHATL2*.

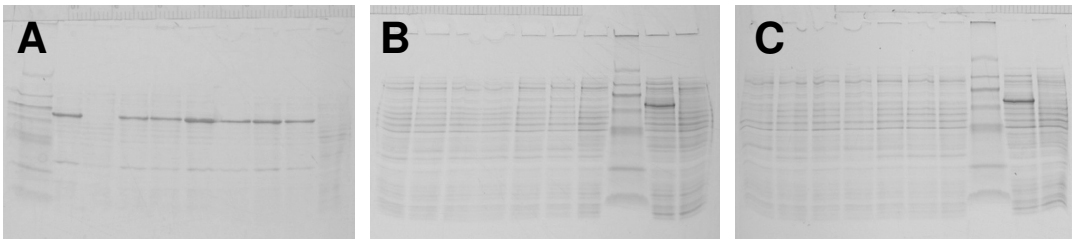

Figure B.3: Soluble fractions of protein extracts on SDS-PAGE gels.

(A) BZL, left to right: molecular mass markers, protein extracts from induced BZL A-H, and protein extract from non-induced BZL J. (B) CHATL 2, L to R: induced samples A-E & J, non-induced J, markers, induced BZL J, non-induced BZL J. (C) CHATL 3, L to R: induced samples A-E & J, non-induced J, markers, induced BZL J, non-induced BZL J.

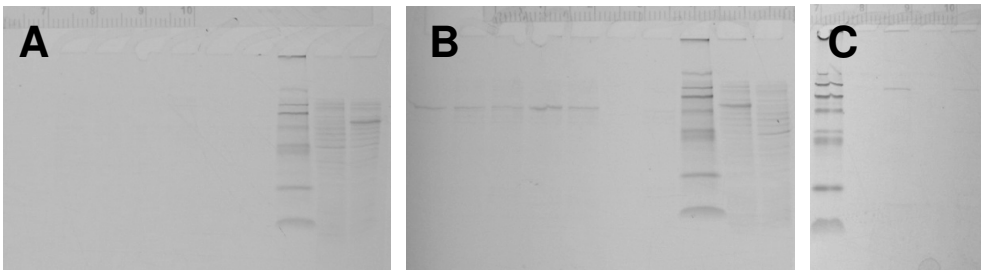

Figure B.4: Insoluble fractions of protein extracts on SDS-PAGE gels.

(A) CHATL 2, L to R: induced samples A-E & J, non-induced J, mass markers, induced BZL J, non-induced BZL J. (B) CHATL 3, L to R: induced samples A-E & J, non-induced J, markers, induced BZL J, non-induced BZL J. (C) CHATL 2, L to R: markers, non-induced A, induced samples A, D, E.

References

AspenDB (2011) [http://aspendb.uga.edu/]

Babst B, Harding S, Tsai C-J (2010) Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in *Populus*. *Journal of Chemical Ecology* 36: 286-297

Beuerle T, Pichersky E (2002) Purification and characterization of benzoate:coenzyme A ligase from *Clarkia breweri*. *Archives of Biochemistry and Biophysics* 400: 258-264

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding *in vivo* benzenoid metabolism in Petunia petal tissue. *Plant Physiology* 135: 1993-2011

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds. (1992) *Short protocols in molecular biology.* John Wiley & Sons, New York, NY

D'Auria JC, Chen F, Pichersky E (2002) Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. *Plant Physiology* 130: 466-476

D'Auria JC, Pichersky E, Schaub A, Hansel A, Gershenzon J (2007) Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (*Z*)-3-hexen-1-yl acetate in *Arabidopsis thaliana*. *The Plant Journal* 49: 194-207

El-Sharkawy I, Manríquez D, Flores F, Regad F, Bouzayen M, Latché A, Pech J-C (2005) Functional characterization of a melon alcohol acyl-transferase gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity. *Plant Molecular Biology* 59: 345-362

Integrated DNA Technologies Inc. (2011) PrimerQuest^{SM}. [http://www.idtdna.com/Scitools/Applications/Primerquest/]

Kampranis SC, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Anssour S, Dunwell JM, Degenhardt J, Makris AM, Goodenough PW, Johnson CB (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: Structural insights into the evolution of terpene synthase function. *Plant Cell* 19: 1994-2005

Okada T, Hirai MY, Suzuki H, Yamazaki M, Saito K (2005) Molecular characterization of a novel quinolizidine alkaloid *O*-tigloyltransferase: cDNA cloning, catalytic activity of recombinant protein and expression analysis in *Lupinus* plants. *Plant and Cell Physiology* 46: 233-244

Souleyre EJF, Greenwood DR, Friel EN, Karunairetnam S, Newcomb RD (2005) An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. *FEBS Journal* 272: 3132-3144

Tuskan GA, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604

Unno H, Ichimaida F, Suzuki H, Takahashi S, Tanaka Y, Saito A, Nishino T, Kusunoki M, Nakayama T (2007) Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *Journal of Biological Chemistry* 282: 15812-15822

Wang J, De Luca V (2005) The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. *The Plant Journal* 44: 606-619