Hydroxycinnamoyl transferases in *Populus* and their roles in vascular development

by

Cuong Hieu Le

Bachelor of Science, University of British Columbia, 2009

A Dissertation Submitted in Partial Fulfilment

of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

University of Victoria

Supervisory Committee

Hydroxycinnamoyl transferases in *Populus* and their roles in vascular development

by

Cuong Hieu Le
Bachelor of Science, University of British Columbia, 2009

Supervisory Committee

Dr. Christoph H. Borchers, Department of Biochemistry and Microbiology
Co-Supervisor

Dr. Jürgen Ehlting, Department of Biology
Co-Supervisor

Dr. Caren Helbing, Department of Biochemistry and Microbiology
Departmental Member

Dr. Peter Constabel, Department of Biology
Outside Member

Abstract

Supervisory Committee
Dr. Christoph H. Borchers, Department of Biochemistry and Microbiology
Supervisor
Dr. Jürgen Ehlting, Department of Biology
Co-Supervisor
Dr. Caren Helbing, Department of Biochemistry and Microbiology
Departmental Member
Dr. Peter Constabel, Department of Biology

Hydroxycinnamoyl conjugates (HCC)s are an extremely diverse class of natural products that serve a wide variety of key functions in plant physiology, for example during wood formation, and in defence. They have diverse biological properties and act as antioxidants, antimicrobials, and antivirals. The biochemical basis of HCC diversity, however, has not yet been fully elucidated. Plants in the *Populus* genus are known to produce a particularly diverse range of HCCs and they constitute up to 5% of the leaf dry mass in some *Populus* species. HCCs can be formed by hydroxycinnamoyl transferases (HCTs) and distinct HCT isoforms in Populus may have distinct biological functions related to the synthesis of specific classes of HCCs. These can be identified on the basis of their evolutionary history and I show that many of the biochemically characterised HCTs belong to the BAHD superfamily of acyltransferases. My phylogenetic reconstruction of the BAHD superfamily has also defined a subclass containing most of the already-characterised HCTs, including nine potential HCT candidates in *Populus*.

Caffeoyl-shikimate is a central precursor in the formation of lignin, a biopolymer (along with cellulose) that imparts mechanical stability to wood. Based on the transcript abundance of two candidate genes PtHCTA1 (Potri.001G042900)) and PtHCTA2 (Potri.003G183900) were

hypothesised to be responsible for caffeoyl-shikimate formation in secondary xylem (i.e., wood). As part of this project, RNAi whole-plant knock-downs were generated for the xylem-associated PtHCTA1/2. The PtHCTA1/2 RNAi knock-downs have stunted growth, reminiscent of other mutants with impaired lignin biosynthesis. Based on thioacidolysis GC-MS, I found that the mutants produced a lignin with enriched hydroxyphenyl (H) subunits, which were derived from precursors upstream of the HCT-catalysed reaction and normally do not occur in *Populus* lignin. Interestingly, in one of the RNAi lines, the lignin phenotype was uncoupled from the developmental dwarfing phenotype. This is of high interest from a bioethanol perspective, since wood rich in H-lignin is more easily fermented than wood that is rich in guaiacyl (G) and syringyl (S) lignin. Another candidate gene (Potri.018G109900, HCT-E2) was linked to the formation of caffeoyl-spermidine in male catkins (which function in pollen coat formation), and one candidate gene (Potri.018G104700, HCT-C2) was associated with the formation of bioactive, soluble HCCs in leaves and roots. Since RNAi-mediated down-regulation proved ineffective, CRISPR-based gene knock-out methodology was developed and utilised for the *Populus* hairy root system. Targeted knock-out mutants for the leaf-associated HCT-C2 were generated. HCC identity was determined by metabolite purification and subsequent MS/MS/MS from leaf extracts, and the metabolite concentrations were determined by LC-MS. A decrease in chlorogenic acid concentration was apparent in CRISPR hairy-root knockouts of HCT-C2 indicating that HCTC2 is involved in HCC biosynthesis and can directly produce chlorogenic acid. Candidates for the HCTs involved in lignin biosynthesis, soluble ester biosynthesis, and pollen coat formation were identified and plant genetics confirmed the role of the lignin and soluble ester HCT candidates.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4CL - 4-coumaroyl CoA ligase

AC - activated charcoal

AHCT - anthocyanin O-hydroxycinnamoyl transferase

BAHD - BEAT, AHCT, HCBT, DAT (superfamily)

BAP - benzylaminopurine

BEAT - benzyl alcohol O-acetyltransferase

BLAST - Basic local alignment search tool

C4H - cinnamate 4-hydroxylase

CCR - cinnamoyl-CoA reductase

CGA - chlorogenic acid

CHS - chalcone synthase

CIM - callus induction media

CSE - caffeoyl-shikimate esterase

CYP - cytochrome P

CoA - coenzyme A

COMT - caffeic acid/5-hydroxyferulic acid O-methyltransferase

CRISPR - Clustered regularly interspaced short palindromic repeat

DAT - deacetylvindoline 4-O-acetyltransferase

DFR - dihydroflavonol-4-reductase

EF1β - elongation factor 1β

ESI - electrospray ionization

FPKM - fragments per kilobase of transcript per million mapped reads

GC - gas chromatography

gDNA - Genomic DNA

HCA - hydroxycinnamoyl amide

HCBT - anthranilate N-hydroxycinnamoyl/benzoyltransferase

HCC - hydroxycinnamoyl conjugate

HCE - hydroxycinnamoyl ester

HCT - hydroxycinnamoyl transferase

HPLC - high-performance liquid chromatography

HQT - hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase

HST - hydroxycinnamoyl-CoA shikimate hydroxycinnamoyl transferase

HRM - high-resolution melt

HTT - hydroxycinnamoyl-CoA tyramine hydroxycinnamoyl transferase

IBA - indole butyric acid

INRA - Institut National de la Recherche Agronomique

IUPAC - International Union of Pure and Applied Chemistry

LC - liquid chromatography

MGL - mannitol glutamic acid: lysogeny (medium)

MS - mass spectrometry

M&S - Murashige and Skoog

NAA - naphthalene acetic acid

NHEJ - non-homologous end joining

NNI - nearest neighbour interchange

PAL - phenylalanine ammonia lyase

PCR - polymerase chain reaction

PDA - photo diode array

PET - paired-end tag

qPCR - quantitative polymerase chain reaction

q-ToF - quadrupole time of flight

RIN - RNA integrity number

RNAi - ribonucleic acid inhibition

RP - ribosomal protein

RT - reverse transcriptase

SHT - spermidine hydroxycinnamoyl-CoA hydroxycinnamoyl transferase

SA - salicylic acid

SCP - serine carboxypeptidases

SEM - shoot elongation medium

SM - secondary metabolites

SNP - single nucleotide polymorphism

SPC - soluble phenolic compounds

SPR - subtree pruning and regrafting

TEM - transmission electron microscopy

THT - tyramine N-hydroxycinnamoyl transferase

UPLC - ultra performance liquid chromatography

UV - ultraviolet (light)

WPM - woody plant medium

XIC - extracted ion chromatogram

1. Introduction

1.1. Plant Metabolism

Plants are an important class of eukaryotic organisms that have colonised almost all of the terrestrial and aquatic ecosystems on the planet. They are omnipresent and comprise a large proportion of the planet's biomass. Archaeplastida (Plantae) encompasses a group of organisms that include the glaucophytes, rhodophytes (red algae), chlorophytes (green algae), and the embryophytes (land plants), which all contain a primary plastid surrounded by a double membrane derived from an ancient endosymbiotic incorporation of a free cyanobacterium (Rodríguez-Ezpeleta *et al.*, 2005). The incorporated cyanobacterium evolved to a specialised organelle, which frequently functions in photosynthesis. The Chloroplastida (Viridiplantae) clade contains only the chlorophytes and embryophytes. The clade is defined by having a plastid with chlorophylls a and b, and cell walls containing cellulose (Adl *et al.*, 2005). In the Ordovician period, descendants of streptophyte green algae began to colonize land leading to the development of the most familiar form of vegetation on the planet, the embryophytes or land plants (Sanderson *et al.*, 2004). A terrestrial environment has many benefits, but it leads to many different challenges including moisture retention, temperature fluctuations, oxygen stress, UV stress, and gravity (Waters, 2003). Having a solid substrate on which to root allows for upward growth and greater photosynthetic potential; this upward growth, however, increases the mechanical stress of gravity, to which plants have adapted by producing the polymer lignin, which is deposited where there is the highest mechanical load (Volkmann and Baluska, 2006). The challenges associated with colonization of land during

land-plant evolution have been addressed by plants, to a large extent, by the expansion of their small molecule repertoire, leading to the production of diverse classes of specialised compounds in plants, also called plant natural products or plant secondary metabolites.

Secondary metabolites produced by plants are *not essential* for survival, but they serve key biological purposes and can function in the interaction of plants with their environment. Secondary metabolites can protect plants from abiotic stresses (such as UV irradiation, nutrient deficiency) and biotic stresses (such as herbivores and pathogens) (Wink, 2010). In addition to protective roles, secondary metabolites can also be used offensively by plants; in allelopathic interactions, plants produce secondary metabolites to inhibit their competitors. In addition to roles as offensive and defensive weapons, secondary metabolites also serve roles in biotic communication. Plants and their cells are generally immotile and chemical volatiles allows communication between different organs within a plant or potentially from plant to plant, and can also function as chemical and visual signals that mediate interactions with animals, e.g., by attracting pollinators, seed dispersers, or parasites of herbivores (Dicke and Baldwin, 2010). Because many these biological chemicals are produced by plants in order to have physiological effects on their targets, they are important to humans, as they can often be used as a source of pharmaceuticals or nutraceuticals, and are extensively used in traditional medicine. Understanding the enzymes involved in secondary metabolism, and the genes encoding them is crucial to decipher the biological functions of these compounds and how they act in the chemical ecology of plants.

Plant secondary metabolites are categorised based on functional groups or by common biosynthetic origin. Major classes include the terpenoids, alkaloids, and phenolics, the final group is the focus of this dissertation. The shikimic acid pathway links primary carbohydrate metabolism to the production of aromatic compounds, including the phenolics. The primary metabolic end-products of the shikimate acid pathway are the aromatic amino acids: tryptophan, tyrosine, and phenylalanine. The intermediates and these end-products serve as precursors for secondary metabolism, leading to the diverse array of plant phenolics that are seen in nature (Vogt, 2010). Phenolics comprise a biochemically diverse group of compounds (Figure 1-1). Branching from many of the early steps in the pathway leads to simple phenolics with a single ring, at least one hydroxyl group, and (frequently) one carboxyl group. Ellagitannins and gallotannins are common forms of hydrolysable tannins derived from 3-dehydroshikimate and are an example of the expansion of the chemical defence repertoire that plants develop as they evolve. The complexity and the oxidation status of hydrolysable tannins have been correlated with plant evolution, indicating potential optimization over evolutionary time, as the potency of these compounds depends on their oxidation stage (Okuda *et al.*, 2000). The salicylates and benzoates are also derived from early intermediates in the shikimate pathway. These are both (large) groups of phenolic compounds with a long history of human use. Willow trees (Salix) have been used for their medicinal properties for centuries (Boatwright and Pajerowska-Mukhtar, 2013). They are high in salicylic acid (SA), which is the precursor the pharmaceutical acetylsalicylic acid which was patented by Bayer in 1898 (Hoffmann, 1900).

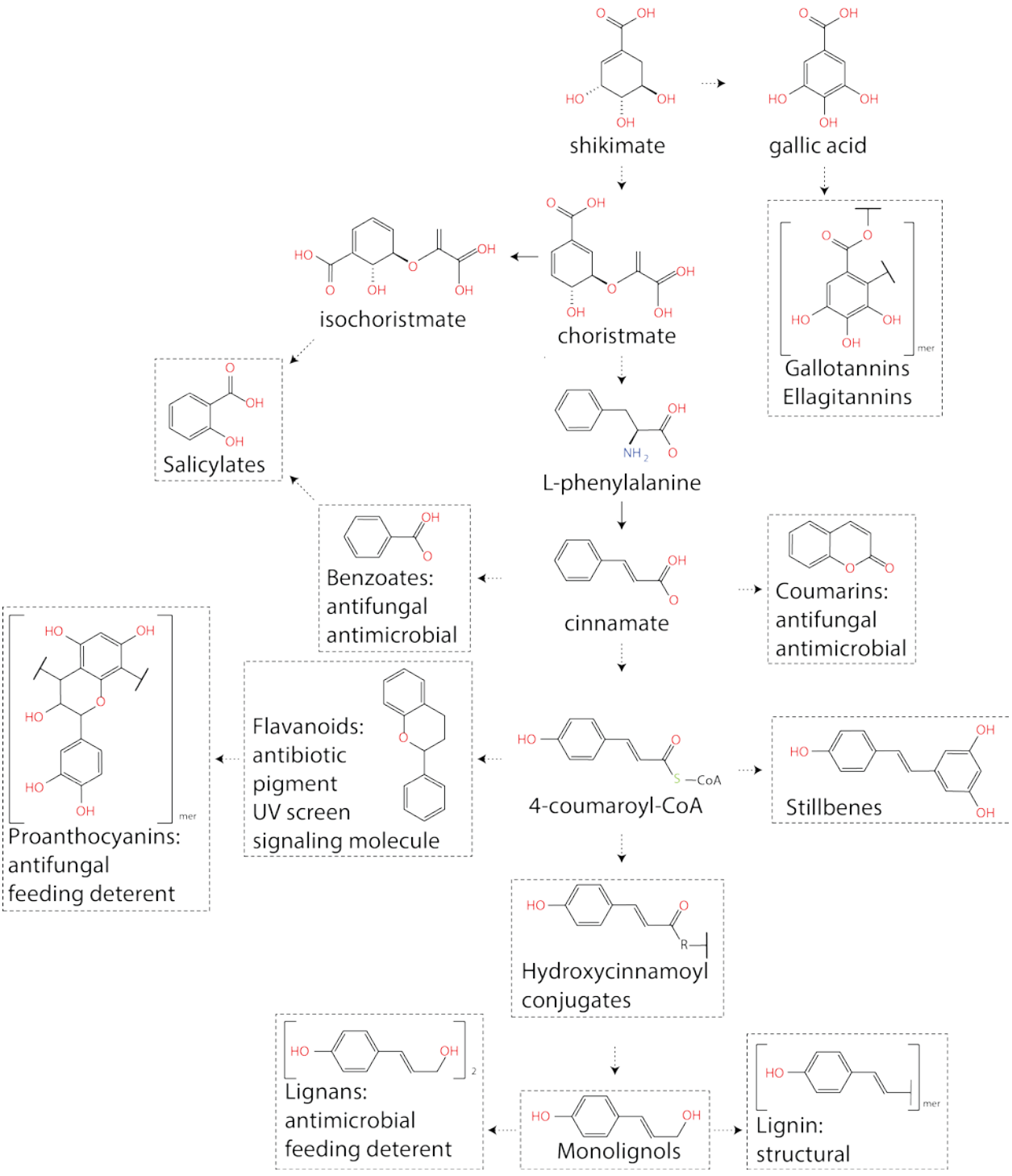


Figure 1-1. Schematic Overview of Phenolic Compound Biosynthesis
Phenolic secondary metabolites are primarily derived from the shikimate pathway. Boxes indicate compound classes, with a skeleton structure of each class being shown. Dashed arrows represent multiple steps. 4-coumaryl-CoA is key branch point that leads to many different classes of compounds.

SA -- although primarily known for its central role in pathogen-induced defence responses -- is also involved in a wide range of other functions, including regulation of seed germination, seedling development, cell growth, stomatal aperture, respiration, temperature tolerance, fruit yield, legume nodulation, and senescence (Vlot *et al.*, 2009). Benzoates have been known to be an antifungal compound for well over 100 years (Krebs *et al.*, 1983). Benzoic acid and its potassium and sodium salts are common food additives that have antibacterial and antifungal properties. Benzyl benzoate is of note as it is a pharmaceutical used to treat ectoparasite infections, and is considered by the World Health Organization to be an essential medicine required for a basic health-care system (World Health Organization, 2015). It is synthesised in leaves, along with other volatile benzoate esters, by *Clarkia breweri* in response to wounding by a class of enzymes known as BAHD acyltransferases (D'Auria *et al.*, 2002; Dudareva *et al.*, 1998; Nam *et al.*, 1999).

1.1.1. **Phenylpropanoids**

Phenylpropanoids constitute a major class of phenolic compounds derived from the shikimate pathway. They are named for their basic backbone structure which consists of an aromatic six carbon phenyl group with an attached three-carbon propyl group. The highly conserved biochemical reactions that convert phenylalanine to 4-coumaroyl-CoA mark the entry point into the secondary metabolism of phenylpropanoids, which is therefore termed the “general phenylpropanoid pathway”. These enzymatic reactions control how much of the carbon fixed through photosynthesis is allocated by the plant to the production of phenylpropanoids (Ro and Douglas, 2004). Phenylalanine ammonia-lyase (PAL), situated at a branch point between primary and secondary metabolism, converts L-phenylalanine into

cinnamate and ammonia (Camm and Towers, 1973). Cinnamate can then be modified by the action of several hydroxylases and/or O-methyl transferase. One such enzyme, cytochrome P450 (CYP)450 monooxygenase cinnamate 4-hydroxylase (C4H), hydroxylates cinnamate to form *p*-coumaric acid ((E)-3-(4-hydroxyphenyl)-2-propenoic acid) (Russell, 1971). *p*-Coumaric acid is then ligated to coenzyme-A by 4-coumaryl:CoA ligase to form 4-coumaroyl-CoA (Knobloch and Hahlbrock, 1977) from which the various diverse phenylpropanoids are derived.

A large proportion of the biomass of our planet is comprised of phenylpropanoids. In addition to existing as soluble phenolic compounds, they are also polymerised to form major structural polymers in plants. In trees, for example, a major portion of their mass is made up of phenylpropanoid polymer lignin and up to 20% of the dry mass of wood in *Populus* is comprised of lignin (Andersson-Gunneras *et al.*, 2006). Lignin is the second most abundant terrestrial biopolymer after cellulose, comprising up to 30% of the organic carbon in the biosphere (Boerjan *et al.*, 2003).

The phenolic compounds are also abundant in diverse plant tissues and they also constitute a major carbon sink as they make up a large portion of the leaf dry mass, for example they comprise between 10 and 35% of leaf dry mass in *P. tremuloides* (Hwang and Lindroth, 1997). In contrast to lignin, much of the fixed carbon that is stored in soluble phenolics will not be stored long-term, and will quickly re-enter the carbon cycle although some will remain locked in the soil. Once in the soil, these compounds have an impact on soil microbial ecology and nitrogen availability, and thereby can influence the overall C-uptake of forest soils (Shay,

2017). Many of the immediate ecological functions for these compounds remain unknown. The soluble phenylpropanoids are extremely diverse and members of this group have a wide variety of functions in the plant kingdom (Vogt, 2010) and humans use these compounds for many different purposes, including as chemical precursors. For example, the phenylpropene eugenol is a floral attractant that is used in perfumes and flavouring. Many of the phenylpropanoids are also used by humans as precursors in the synthesis of useful bioactive compounds. For example safrole derivatives can be used in the synthesis of pesticides (piperonyl butoxide) (Herman, 1949), perfumes (piperonal) (Blair, 1959), and narcotics (methylenedioxymethamphetamine) (Noggle *et al.*, 1991).

Phenylpropanoids coming from the general phenylpropanoid pathway are also used by plants as precursors in the biosynthesis of the different classes of bioactive compounds derived from phenylpropanoids. Coumarins, chalcones, flavonoids, and stilbenoids are other classes of phenylpropanoids generated by the initial condensation of additional units to the phenylpropanoid backbone followed by cyclization to form additional aromatic moieties (Figure 1-1). Coumarins are a class of phenylpropanoid in which the carboxy group of cinnamate is transesterified internally to an ortho-hydroxyl group on the phenyl ring, leading to the formation of the heterocyclic six-membered ring (Shimizu, 2014). The prototypical compound coumarin gives hay its characteristic odour, and many other compounds in this class are used in the perfume industry for their pleasant smell. The induction of coumarins has been observed in response to herbivory (Olson and Roseland, 1991) and directly from cell debris (Davis and Hahlbrock, 1987). Warfarin, a common anticoagulant listed on the WHO

Model List of Essential Medicines (World Health Organization, 2015), is a coumarin-derived drug that interferes with Vitamin K metabolism (Whitlon *et al.*, 1978).

Polyketide synthase can extend the structure of a phenylpropanoid by condensing acetate extender units, derived from malonyl-CoA. These units are cyclised to form additional aromatic moieties leading to chalcones, flavonoids, and stilbenoids (Figure 1-1). Stilbenoids (including resveratrol, a stress-induced phytoalexin commonly found in grapes and red wine (Wang *et al.*, 2010), and pinosylvin, (an antifungal compounds which accumulates in *Pinaceae* heartwood), are formed from stilbene synthase (Schanz *et al.*, 1992). Chalcones are also produced by a polyketide synthase, chalcone synthase (CHS), and contain two aromatic rings linked by a three-carbon enone moiety. Soluble chalcones, which are abundant in plants, show a wide array of biochemical activities and function as potential antimitotic, anti-infective, anti-inflammatory, antiviral, antibacterial, anticancer, antioxidant, and antimitotic drugs (Anto *et al.*, 1995; Ducki *et al.*, 1998; Mahapatra *et al.*, 2015; Nowakowska, 2007). Chalcone isomerase is responsible for the formation of the basic flavonoid structure form, chalcone. It facilitates the closure of chalcone to form an additional heterocyclic ring forming the flavanone, naringenin. From here, modification by various enzymes can lead to the many different classes of flavonoids including flavanonols, flavanones, flavones, flavonols, and flavanols (Winkel-Shirley, 2001). Some flavonoids are known phytoalexins, for example, the isoflavonoid medicarpin has been shown to be a defensive response induced by two distinct pathways in *Medicago truncatula* (Naoumkina *et al.*, 2007). Anthocyanin and their glycosylated derivative, anthocyanidins, are plant pigments belonging to the flavonoid class. Anthocyanins are the primary pigments that colour flowers, and directly influence the pollinators which are

attracted to the plant (Bradshaw and Schemske, 2003). The accumulation of anthocyanin in other vegetative tissues has also been proposed to provide protective antioxidant properties (Gould *et al.*, 2002). They are formed by the dihydroflavonol-4-reductase (DFR) which reduces the ketone on the flavonoid at the 4 position of ring B to a hydroxyl group. Dehydration of the hydroxy group at this position leads to an extension of aromaticity to ring B; this conjugated aromatic system is responsible for the colouration that these compounds provide.

The conjugation of hydroxycinnamic acids to polar molecules forming hydroxycinnamoyl conjugates (HCC)s is widespread and naturally occurring in plants (Salvador *et al.*, 2013). Simple phenylpropanoids, such as hydroxycinnamic acids, are used as antioxidants in foods, cosmetics, and drugs -- including use as antivirals due to their inhibition of interleukin-8 (Hirabayashi *et al.*, 1995), as well as sunscreens due to their UV absorption. Cinnamate itself has been shown to alter soybean growth, increasing lignin production and inhibiting root growth, and potentially functioning as an allelopathic chemical (Salvador *et al.*, 2013). The free acids are not readily soluble in aqueous solutions and the conjugation of hydroxycinnamic acids to polar molecules forming HCCs has been proposed to solve this problem in human products (Compton *et al.*, 2000; Tsuchiyama *et al.*, 2007). HCCs in plants may be formed by the action of hydroxycinnamoyl transferases (HCTs) which catalyses the transfer a hydroxycinnamoyl moiety from a hydroxycinnamoyl-CoA thioester (e.g., 4-coumaroyl-CoA) to a hydroxyl or amino group of an acceptor (e.g., shikimate, quinate, glutaric acid, malic acid, glycerol, an anthocyanin, anthranilate, or spermidine). HCCs are an abundant class of phenylpropanoids, and the enzymes that form HCC esters and amides will be the primary focus of this dissertation. Several dozen enzymes have been found to have HCT activity, in

that they perform a transesterification reaction, which forms an HCC; however, most of the enzymes that produce HCCs are currently uncharacterised and we are only starting to understand the biological significance of these compounds.

1.1.2. **Hydroxycinnamoyl Conjugate Structure and Function**

An HCC is a conjugate of an hydroxycinnamoyl derivative and an alcohol, amide, or sulfide. The carbocyclic acid moiety is based on the hydroxycinnamic acid, *p*-coumaric acid (4-hydroxylation only). CYP450s have also been shown to hydroxylate *p*-coumaric acid conjugates leading to the corresponding caffeic acid conjugate (3, 4-hydroxylation); further methoxylations and hydroxylations give rise to ferulic acid (4-hydroxy, 3-methoxy), and sinapinic acid (4 hydroxy, 3,5 methoxy) (Ehlting *et al.*, 2006). Variations in the carbocyclic acid moiety of this class of compounds are thus limited to differential hydroxylation and methoxylation patterns at the 3- and 5-positions of a phenyl ring. In contrast, the conjugated moiety can be chemically diverse and includes a wide array of alcohols and amines, as well as coenzyme-A which serves as an activating moiety in subsequent transesterification reactions during the biosynthesis of HCCs (Figure 1-2). In addition to being a precursor of lignin, makes up a large proportion of woody tissues, many different HCCs accumulate as freely soluble compounds. Soluble HCCs can contribute to the biomass of plant tissues. They comprise up to 5% of the leaf dry mass and are abundant in *Populus* bud exudates; although the quantity and composition of HCCs *in Populus* can vary greatly from species to species (English et al., 1991; Greenaway and Whatley, 1990, 1991; Greenaway et al., 1987; Isidorov and Vinogorova, 2003). HCCs are widespread in the plant kingdom and have diverse functions in plant biochemistry; some have been connected to plant defences against pests and pathogens.

Many HCCs have been found in nature including hydroxycinnamoyl esters of tartaric acid (Scarpati and Oriente, 1958), glycerol (Kim *et al.*, 2012), simple sugars (Harborne and Corner, 1961), and choline (Tzagoloff, 1963). Two of the most well-studied and abundant HCCs are quinate esters (chlorogenic acids; CGA) (Sondheimer, 1958) and 3,4-dihydroxyphenyllactic acid esters (rosmarinic acid; RA). *In vitro*, HCCs have been shown to have antioxidant (Kikuzaki *et al.*, 2002), antibacterial (Huang *et al.*, 2009), antifungal (Huang *et al.*, 2009), and antiviral (Kishimoto *et al.*, 2005) properties. They have been shown to accumulate upon various stresses, including fungal infection (Lyons *et al.*, 1990). Although direct evidence of ecological function has not been shown for many of these compounds, some can be readily taken up by herbivores and their metabolites can be widely distributed in the body, e.g., caftaric acid has been shown to penetrate the blood-brain barrier in mice (Vanzo *et al.*, 2007).

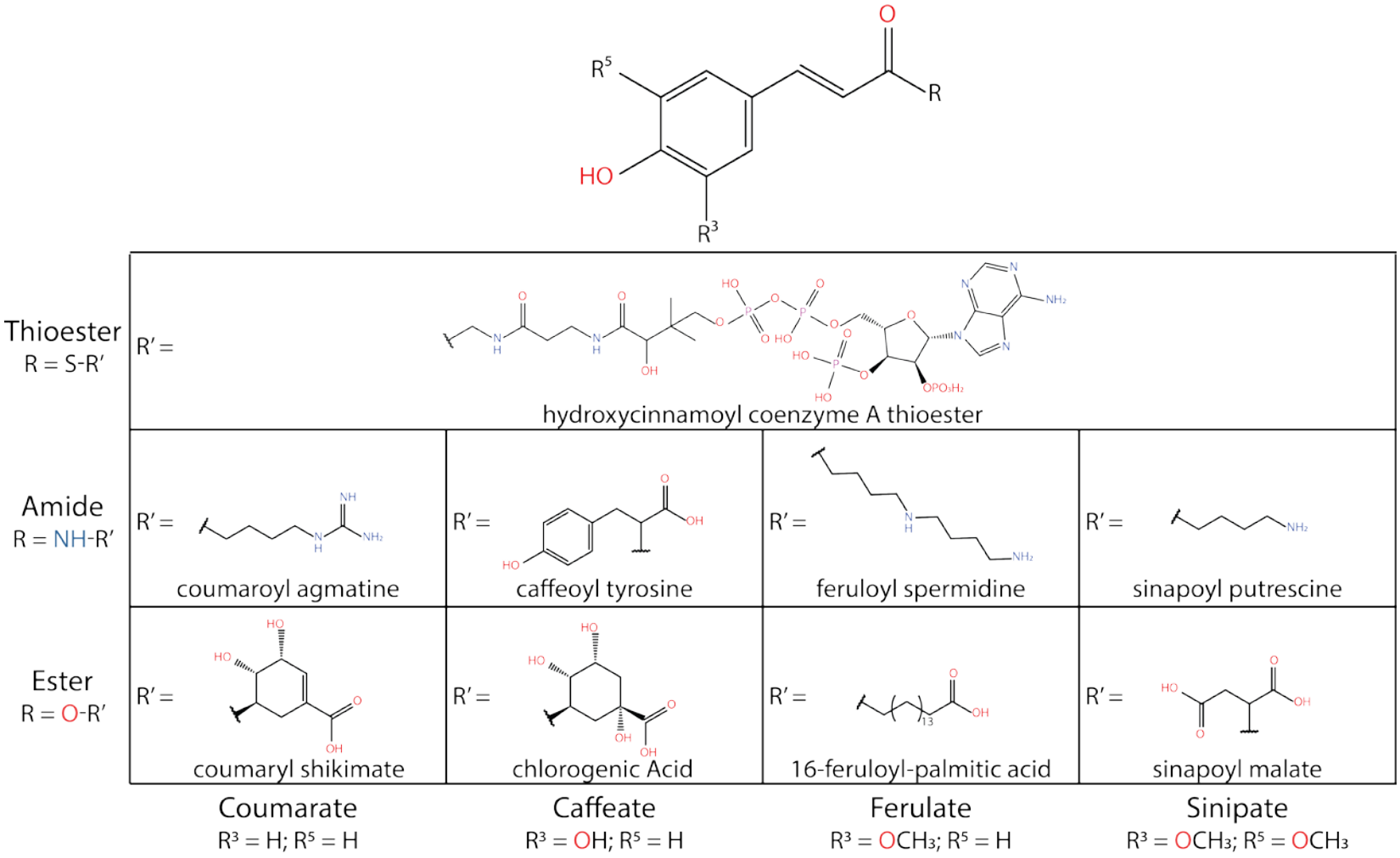


Figure 1-2. The General Structure of HCCs with a Few Notable Examples Shown
From HCC-CoA thioesters (S-bond), many other compounds are formed. HCTs transesterify the CoA moiety with another molecule: transesterification with an alcohol forms an ester (O-bond) and transesterification with an amine forms an amide (N-bond). Both esters and amides are commonly found in diverse plants. The hydroxycinnamoyl moiety varies based on 3′ and 5′ methoxy/hydroxylation with the most frequent being coumaryl, caffeoyl, feruloyl, or sinapoyl.

CGAs are esters between hydroxycinnamic acids and quinic acid, a direct derivative of the shikimate pathway, a hydrated derivative of shikimate (Guo *et al.*, 2014). CGAs are some of the most common HCCs found in plants and are distributed in a variety of plants, including coffee, potato, tobacco, apple, and poplar (Clifford, 2000). The International Union of Pure and Applied Chemistry (IUPAC) established an IUPAC commission on the nomenclature of cyclitols in 1976, which recommended the name of the most abundant chlorogenic acid to be 5-O-caffeoylquinic acid (5-CQA) as opposed to the previous biological numbering, 3-O-caffeoylquinic acid (3-CQA). In this document, we will follow the IUPAC nomenclature and numbering -- we will refer to 5-O-caffeoylquinic acid (5-CQA) as simply chlorogenic acid (CGA), and we will use the term chlorogenic acids (CGAs) to represent the class (IUPAC, 1976). CGA has been demonstrated to be a strong antioxidant that may be produced in response to stress such as UV exposure and herbivory (Clé *et al.*, 2008; Grace *et al.*, 1998; Izaguirre *et al.*, 2007). It is commonly known as an astringent, providing the bitter taste found in coffee and the unpleasant taste of many plant leaves, and has been shown to increase the resistance of various plants to pests. Its anti-herbivory effect has been demonstrated in willow against leaf beetle (Grace *et al.*, 1998; Ikonen *et al.*, 2001) and thrips in chrysanthemum (Leiss et al., 2009). In addition to its well-characterised function as an antifeeding compound, CGA has a proposed function in the defence against pathogens, iron chelation, and protection against UV radiation or environmental stress (Mondolot *et al.*, 2006). This highlights the multi-functionality frequently seen for plant secondary metabolites. Although quinic acid derivatives are some of the most common conjugates found soluble in plants, there is a wide array of alternate (and sometimes species-specific) HCCs.

Another well-studied HCC is rosmarinic acid (RA), an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid, which was originally discovered in *Rosmarinus officinalis* (Scarpati and Oriente, 1958). It has been shown to accumulate in Boraginaceae and the Lamiaceae sub-family Nepetoideae; as well as in species of several orders of mono- and eudicotyledonous angiosperms (Petersen *et al.*, 2009). RA has been proposed to have a wide range of biological functions. Similarly to CGA, RA has strong anti-oxidative properties (Lin *et al.*, 2002; Osakabe *et al.*, 2004b) and it has been explored for its use as an anti-inflammatory (Osakabe *et al.*, 2004a; Osakabe *et al.*, 2004b; Sanbongi *et al.*, 2004; Swarup *et al.*, 2007), anti-carcinogenic (Osakabe *et al.*, 2004a), and anti-viral (Hooker *et al.*, 2001; Swarup *et al.*, 2007) drugs. RA has also been shown to inhibit snake venom-induced haemorrhage in mice (Aung *et al.*, 2010a; Aung *et al.*, 2010b).

HCCs, particularly caftaric acid, the ester of caffeic acid and tartaric acid, are the major phenols (besides flavonoids) in *V. vinifera*. Caftaric acid is highly abundant in grape juice and HCCs are considered to be important phenolic compounds in wine. The concentrations of caftaric acid and its glutathione conjugate 'grape reaction product' (2-S-glutathionyl caftaric acid) contribute the sensory impact of wine (Gawel *et al.*, 2014; Hufnagel and Hofmann, 2008). It is present in *Cichorium intybus* (chicory) and *Echinacea purpurea,* which also produces the dicaffeoyl ester of tartaric acid, chicoric acid. Chicoric acid was initially isolated from *C. intybus* roots and it is also found abundantly in *Taraxacum* spp., *Melissa officinalis,* and *Ocimum basilicum*. Similar to other HCCs, chicoric acids have a wide range of biological activities, including anti-viral activity -- specifically inhibition of HIV-1 integrase (Robinson and Mansfield, 2009).

The majority of known hydroxycinnamoyl conjugates fall into the ester class, but amines are also prevalent and have important biochemical functions. Hydroxycinnamoyl spermidines are found in the pollen coats of many plants. Genetic knockout of the BAHD acyltransferase that is responsible for the formation of these compounds in *A. thaliana* (Grienenberger *et al.*, 2009), leads to a crushed pollen phenotype with visible pollen wall irregularities. Introduction of a *Malus pumila* paralogue was shown to recover the phenotype (Elejalde-Palmett *et al.*, 2015). Likewise, genetic interruption of later stages of hydroxycinnamoyl-spermidine biosynthesis also impacts wall integrity and pollen viability (Elejalde-Palmett *et al.*, 2015; Matsuno *et al.*, 2009). It, therefore, appears that HCCs -- particularly hydroxycinnamoyl-spermidines -- are essential for pollen structure and function, making these compounds primary metabolites.

1.1.3. **Hydroxycinnamoyl Conjugates in Lignin Biosynthesis.**

Although HCCs and their derivatives are diverse in structure and function in a wide range of plants, by far the largest sink of phenylpropanoids, in terms of biomass, are polymerised structures -- lignin, in particular. The HCC caffeoyl-shikimate is one of the chemical control points in the biosynthetic route leading to lignin biosynthesis. Lignin is a major structural component of plant secondary cell walls, which is formed after the cell has stopped elongating(Vanholme *et al.*, 2010). Secondary cell walls are found in many plant tissues, particularly in wood-forming tissues, and provides additional rigidity and mechanical stability. Lignin content and composition has a considerable impact on the chemical and physical properties of many plant-derived materials and affects properties that are associated with wood quality. Understanding how lignin is synthesised, and what factors affect its

composition, have direct implications for the practical use of wood products. Because caffeoyl shikimate is an intermediate in lignin synthesis, understanding the HCTs that form it has obvious implications in wood formation, pulp/paper making, and the use of lignocellulose materials in second-generation biofuel production.

Lignin is a racemic aromatic heteropolymer that is formed by the oxidative combinatorial coupling of 4-hydroxyphenylpropanoids (Boerjan *et al.*, 2003; Vanholme *et al.*, 2010). It is mainly composed of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) subunits, which are derived from phenylpropanoids. H-lignin is formed from *p*-coumaryl alcohol; G-linin and S-lignin are derived from coniferyl alcohol, and sinapyl alcohol respectively. Only the pathway to G- and S-lignin requires an HCT-catalysed reaction (Boerjan *et al.*, 2003). Lignin is responsible for the structural integrity of the cell wall and provides mechanical stability to woody structures by stiffening and strengthening the stem. This enables high negative-pressure water transport, and allows for the extension of the plant's vascular system (Boerjan *et al.*, 2003). The structure of lignin also serves to protects plants against pathogens. Cell wall apposition formation serves to halt the initial progression of pathogenic fungi and inhibition of lignin biosynthesis attenuates this function (Bhuiyan *et al.*, 2009) . Lignin composition varies in response to development, biotic and abiotic stresses, wounding (including by herbivores), infection, and ions in the cell-wall structure (Vanholme *et al.*, 2010).

1.2. Populus

Poplars (all *Populus* species, including aspens and cottonwoods, will be referred to as poplars here) are deciduous trees spread across the Northern hemisphere. They are abundant in the

boreal forest and are the keystone species in riparian ecosystems. Owing to their fast growth rate and an established biotechnological toolkit, poplars have become a tree model species (Tuskan *et al.*, 2006). Poplars are also recognised for their quantity and quality of phenolic secondary metabolites, including HCCs.

Poplars produce wood products that are commercially important, including as feedstock for pulp and paper. The composition of the wood and the ester products that poplars produce are important because they affect how poplar can be employed in wood, biofuel, and paper products. Poplars are popular in forestry production due to their fast growth rate, which increases harvest rates, and for their ability to regenerate easily, which also allows large-scale asexual reproduction, which decreases the financial burden of reforestation. Aspens, which belong to the *Populus* genus, are a large carbon sink, particularly in the boreal forest, and their decline due to climate change may have a significant impact (Natural Resources Canada, 2017). Poplars thrive on land which is not suitable for food growth making their production non-competitive with food crops. In Europe, closely related willows are used as a biofuel in short-rotation coppice systems; in Canada, due to their rapid growth rates in Canadian climates, poplars are an excellent candidate for a long-term carbon sink to mitigate the climate-change effects of elevated atmospheric CO_2 levels, and as a feedstock sustainable for biofuel generation for transportation.

P. trichocarpa is commonly known as black cottonwood or western balsam poplar. Its native range is the west coast of North America, from Mexico to Alaska, and it is a common native plant in western Canadian riparian habitats. An individual tree from the Nisqually river basin

in Washington State was the first tree to have its genome sequenced (Tuskan *et al.*, 2006). The genome sequence revealed a genome duplication 60 million years ago, shortly after the Chicxulub impact event -- the bolide impact that may have led to the Cretaceous–Paleogene extinction -- which has been proposed to have led to an increase in whole genome duplications (Vanneste *et al.*, 2014). This genome duplication makes *Populus* an interesting genus to study from an evolutionary perspective, as the recent genome duplication makes the species a good model system for the evaluation of the evolutionary fates of gene duplicates. In terms of practical consequences, it makes genetic manipulation slightly more difficult as most genes occur in pairs of paralogues that are nearly identical (Tuskan *et al.*, 2006).

Due to the practical limitations of genetic transformation and culturing techniques with *P. trichocarpa,* other related species are often used for reverse genetic analysis. Within the French Institut National de la Recherche Agronomique (INRA) species collection there are several *Populus* hybrids, i.e. crosses between two different species of *Populus*, that are commonly used for reverse-genetic interrogation. Protocols have been developed for two hybrid lines (Meilan and Ma, 2007): clonal propagates of a male *Populus* INRA 717-1B4 (*Populus tremula* x *Populus alba*) and of a female *Populus* INRA 353-38 (*Populus tremuloides* x *Populus tremula*). *Populus* INRA 717-1B4 is commonly used for its rapid growth rate and ease of transformation, and work in in this hybrid is facilitated by a variant-substituted custom genome scaffold (Xue *et al.*, 2015). This assembly is based on sequence data for *Populus* INRA 717-1B4 mapped to the *P. trichocarpa* Nisqually-1 genome. This leads to a genome sequence that is primarily based on *P. trichocarpa* Nisqually-1 with single nucleotide polymorphisms

(SNPs) within transcribed regions of the genome having been replaced with the INRA 717-1B4 variant. A draft genome of *Populus* INRA 717-1B4 has also been recently published (Mader *et al.*, 2017).

Although a plethora of genomic information is available for many different plant species, it is still difficult to link genomic information to a specific biological phenotype. The genes that are specifically involved in the biosynthesis of many secondary metabolites are still unknown. In this dissertation, I will attempt to use bioinformatic analyses and reverse-genetic approaches to decipher which specific genes in *Populus* are responsible for the formation of hydroxycinnamoyl conjugates.

1.3. Hypothesis and Objectives

I hypothesize that the chemically and functionally diverse hydroxycinnamoyl esters found in *Populus* are produced by specific hydroxycinnamoyl transferases (HCT) that can be identified by bioinformatic analyses and characterised through reverse genetics and metabolic phenotyping. Distinct HCT isoforms in poplar have distinct biological functions related to the synthesis of either protective, soluble HCEs or the synthesis of lignin, and that they can be differentiated on the basis of their evolutionary divergence and the transcript abundance in tissues where they are synthesized.

Objective One: To identify candidate genes responsible for the production of functionally different hydroxycinnamoyl conjugates present in *Populus* through phylogenetic reconstruction and expression profiling.

Objective Two: To functionally characterize specific hydroxycinnamoyl transferases responsible for the production hydroxycinnamoyl esters leading to lignin formation in *Populus*. Knock down of the primary candidate should lead to a clear and observable phenotype in lignin.

Objective Three: To validate that specific hydroxycinnamoyl transferases are responsible for the production of functionally and chemically distinct soluble hydroxycinnamoyl esters, particularly chlorogenic acids, in *Populus*. Knockout plants should have a clear metabolic phenotype.

2. Hydroxycinnamoyl Transferases in *Populus*: Evolutionary Classification and Gene Expression Profiling

2.1. Introduction

As outlined in detail in Chapter 1 of this thesis, HCCs are chemically diverse compounds that have wide-ranging functions in plant development and chemical ecology. Identification of the enzymes that are responsible for their production is the first step in understanding their biosynthesis. A large majority of the enzymes known to form HCCs belong to the BAHD superfamily of plant acyl-CoA dependent acyltransferase, which will be the focus of this chapter. The few exceptions of enzymes with HCT activity belonging to other families include glucosyltransferases, which use UDP-glucose as an acyl-donor in an HCT reaction, indolamine-specific tyramine N-hydroxycinnamoyl transferase (THT)s, and serine carboxypeptidases (SCP)s, which are a class of enzymes that generally function in the cleavage of peptide bonds (Supplementary Table 2). The vast majority of enzymes with HCT activity belong to the BAHD superfamily. The BAHD superfamily of enzymes is a very large family encompassing many distinct acyltransferases in addition to HCTs, for example, acetyltransferases, malonyl transferase, and benzoyl transferases.

2.1.1. **BAHDs**

BAHDs are a superfamily of acyltransferases that are encoded by large gene families in plants typically containing 50 -100 members. They are named after the first four characterised enzymes in this family: benzyl alcohol O-acetyltransferase (BEAT), (Dudareva *et al.*, 1998), anthocyanin O-hydroxycinnamoyltransferase (AHCT) (Fujiwara *et al.*, 1998), N-

hydroxycinnamoyl/benzoyltransferase (HCBT) (Yang *et al.*, 1997), deacetylvindoline 4-O-acetyltransferase (DAT) (Power *et al.*, 1990).

BAHDs are involved in the production of a wide array of secondary metabolites in plants and produce compounds such as phenolics glycosides, waxes, volatile esters, and HCCs (D'Auria, 2006). Many medicinally important compounds including the pain medication, morphine, the chemotherapy drug paclitaxel, and the antiarrhythmic agent ajmaline, involve BAHD enzymes in their biosynthesis (Lallemand *et al.*, 2012b; Ma *et al.*, 2005). Although they are primarily associated with plants and are absent from animals, similar enzymes producing mycotoxins have been found in *Fusarium graminearum* (*Gibberella zeae*) the causative agent of Fusarium head blight (Garvey *et al.*, 2008; McCormick *et al.*, 1999; Tokai *et al.*, 2005).

The first crystal structure of a BAHD acyltransferase, vinorine synthase, was elucidated by Ma et. al. (Ma *et al.*, 2004). Thus far, all known BAHD enzymes are soluble monomeric enzymes located in the cytosol; there have not been any identified enzymes with an organelle-localization signal (D'Auria, 2006). These enzymes consist of two distinct domains, each containing pockets connected by a small channel. Each of the substrates appears to enter through its respective pocket and the transesterification reaction occurs in the channel that links the two domains, where the active site is located (Ma *et al.*, 2005). BAHDs have two characteristic motifs: an HXXXDG motif involved in catalysis, which is present in the active site, and a DFGWG motif which is located far from the active site and which does not function in catalysis. Instead, the DFGWG motif appears to be related to the stabilization of the two domains (Ma *et al.*, 2005).

BAHDs transfer an acyl group from a CoA-thioester to an alcohol or amine acceptor, which yields an acyl-ester or an -acyl-amide respectively. Substrates can vary widely within the superfamily, but a group is always transferred from a CoA-activated thioester. BAHD-mediated catalysis begins with the histidine residue of the HXXXDG motif, which deprotonates the hydroxyl or amino group on the acceptor molecule making a nucleophilic attack on the carbonyl carbon of the CoA-thioester possible. Subsequent nucleophilic attack leads to a short-lived tetrahedral intermediate between the two substrates. Protonation releases the free CoA and the newly conjugated compound (D'Auria, 2006).

Many HCC-forming HCTs in the BAHD superfamily have been characterised. Some of the most well-studied BAHDs are the shikimate specific "Hydroxycinnamoyl-CoA shikimate hydroxycinnamoyl transferases" (HST)s. The crystal structure of HST from *Coffea canephora* is very similar to other BAHD transferases and shows structural similarity to chloramphenicol acetyltransferase-like domains and contains large mixed β sheets flanked by α helices (Lallemand *et al.*, 2012a; Lallemand *et al.*, 2012b). In the HST structure, a flexible α1-β3 loop was found to be important in docking both the hydroxycinnamoyl moiety and the acyl acceptor; histidine, valine, and proline were involved in catalysis and another histidine moiety was engaged in π-stacking to constrain the rotation of the imidazole ring in the active site, similar to other BAHDs (Lallemand *et al.*, 2012b). The functional significance of HST and its products has been shown by analysing *in vivo* transgenics. Knock-downs and knock-outs of the HCCs that form polymerised structures lead to clear physical defects in the structure of plant organs. HSTs transfer a coumaroyl moiety from coumaroyl-CoA to shikimate. This leads to coumaryl shikimate, which is subsequently modified in order to produce G- and S-

monolignols which are polymerised to form G-lignin and S-lignin. Repression of HST has been shown to change the amount and composition of lignin in a several species, including *Arabidopsis thaliana* (Besseau *et al.*, 2007; Hoffmann *et al.*, 2005) *Nicotiana benthamiana* (Hoffmann *et al.*, 2005), *Populus nigra* (Vanholme *et al.*, 2013a), *Pinus radiata* (Wagner *et al.*, 2007), *Panicum virgatum* (Eudes *et al.*, 2016) and *Medicago sativa* (Shadle *et al.*, 2007); however, the involvement of HCT in lignin formation in *Populus* has been disputed (Vanholme *et al.*, 2013a; Vanholme *et al.*, 2013b). Lignin-associated HCTs are the focus of Chapter 3 of this thesis and will be thoroughly described there.

The HSTs are the best characterised HCTs and they appear in most plant lineages, but the broad diversity of HCCs indicates that other HCT activity likely exist and that these may be responsible for the synthesis of these soluble compounds. Certain HCTs, which have greater specificity towards quinate than shikimate, have also been described and are termed "Hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferases" (HQT)s (Niggeweg *et al.*, 2004). *In vivo* gene-silencing experiments have shown that HQTs are likely to be responsible for secondary metabolite biosynthesis, including chlorogenic acid. Silencing these genes does not affect the formation of wood and does not lead to any obvious physiological phenotype in normal unstressed plants, but it does cause changes in secondary metabolite production (Lepelley *et al.*, 2007). In the *Solanaceae* family, which includes many agriculturally important crops such as tomato, potato, and tobacco, HQTs are specifically responsible for the accumulation of chlorogenic acid (Lepelley *et al.*, 2007).

The difference between quinic and shikimic acids is the presence of a hydroxyl group at C-1 in quinic acid and a double bond between C-1 and C-2 in shikimic acid, resulting in a different geometry of the polyol ring. The structures of HST and HQT from *C. canephora* show that the residues lining the substrate-binding pocket appear to adjust the volume to selectively accommodate different substrates -- leucine and phenylalanine residues in the pocket were required for the specificity of quinate. The different CGAs that are found in coffee are created by HQT, with 3,4-di-O-caffeoylquinic acid (3,5-diCQA) being produced first with subsequent isomerization into 3,4- and 4,5-diCQA (Lallemand *et al.*, 2012b). Surprisingly, although the HXXXDG motif is universally conserved in all known BAHD acyltransferases, conversion of the catalytic histidine to aspartate resulted in an *increase* in chlorogenic acid production, so the highly conserved histidine in the characteristic motif appears to not be essential for catalysis in the case of chlorogenic acid.

The characterised HCTs appear to have varying substrates specificities, and other BAHD HCTs are responsible for the biosynthesis of other HCCs, including phaselic acid (Sullivan, 2008; Sullivan, 2009; Sullivan and Zarnowski, 2011) and rosmarinic acid (Berger *et al.*, 2006). Individual HCTs may produce a wide range of 4-coumaroyl-esters; these could then be 3-hydroxylated (mediated by broad-range or specific CYP98As) to form the caffeoyl esters observed (Alber, 2016). Alternatively, CYP98As could preferentially act on 4-coumaroyl-shikimate and, in concert with HST, could produce caffeoyl-CoA. Caffeoyl-CoA could then be the substrate of a range of other HCTs to produce the caffeoyl-ester diversity observed. In either case, HCT-like genes would be key players in producing the variety of esters present.

2.1.2. **Phylogenetic Classification of BAHD Enzymes**

A previous phylogenetic reconstruction of the BAHD family of acetyltransferases focused on the monocot, *O. sativa*, and the eudicots, *P. trichocarpa*, *A. thaliana*, *M. truncatula*, and *V. vinifera* (Tuominen *et al.*, 2011). Their reconstruction supported eight distinct clades which were numbered based on the expansion of specific clades in a previous phylogenetic reconstruction (D'Auria, 2006). The initial assignment of functions to the enzymes in these clades was based on the functions of members that had been characterised. Currently, the genomes of *P. trichocarpa, A. thaliana,* and *P. patens* have been sequenced – and, therefore, the number of members in the BAHD families are known. It is still unclear, however, whether phylogenetic classification coincides with BAHD enzymatic function and, in particular, if within-clade sub-division is caused by the evolution of distinct enzyme functions (for example within the clade containing the characterised HCTs). The expansion of a phylogenetic reconstruction to include more divergent lineages, including gymnosperm, lycopods, and bryophyte sequences and the additional functional characterisation of enzymes within the eudicots is expected to clarify the evolutionary interpretation and functional divergence via a phylogenetic reconstruction of this family.

In my analysis, the BAHD superfamily contains 127 (*P. trichocarpa*) to 15 (*P. patens*) members in chloroplastids with completed genomes. In any given species, only a small fraction of the family members has been functionally characterised. For example, 16 out of the 55 BAHD family members in *A. thaliana* have been biochemically characterised, and a clear biological function been assigned for only 14 of these (Supplementary Table 3). For most other species, the numbers are much smaller and, for most species, only sequence data, but no functional

data are available. Overall, more than 119 BAHDs (distributed among 62 species) have been characterised. Bringing complete gene families into a phylogenetic context allows the reconstruction of common evolutionary origins and may allow the assignment of putative functions based on evolutionary relationships to characterised members from other species. Using an expanded phylogenetic inference, I have identified HCT candidate genes for each of the HCC groups in *Populus* and have further substantiated the candidate genes' proposed function through expression profiling based on microarrays, RNAseq, and qPCR.

2.2. Materials and Methods

2.2.1. **Phylogenetics**

The sequences of all characterised BAHD enzymes were manually compiled by retrieving the sequence from GenBank based on the identifier noted in the respective publication (see Supplementary Table 3 for a complete reference list for the 119 sequences included). Enzymes were selected on the basis of an extensive literature review. All papers published before 2014 that included the term 'BAHD' as well as papers that they cited were manually reviewed for descriptions of enzymatic activity and the associated enzyme sequences were retrieved from the cited databases. All sequences were aligned using Clustal Omega v1.2.4 (Sievers *et al.*, 2011). Sequences were manually trimmed and any positions in the alignment that contained a gap in more than 99% of the sequences were removed. A maximum likelihood phylogeny of the biochemically characterised BAHD acyltransferases, with 1008 bootstrap replicates, was generated using PhyML 3.3, which generates an initial distance-based tree and iteratively refines the tree to improve likelihood (Guindon *et al.*, 2010),. The VT

model (Müller and Vingron, 2000) was selected using ProtTest3 (Darriba *et al.*, 2011). Nearest neighbour interchange (NNI), which exchanges the connectivity four adjusted subtrees (Guindon and Gascuel, 2003), and subtree pruning and regrafting (SPR), which prunes and regrafts a subtree to a new position (Guindon *et al.*, 2010), were used to for tree topology search. Basic local alignment search tool (BLAST) v2.2.60+ (Altschul *et al.*, 1990) was used to retrieve other enzymes with sequences similar to those of characterised BAHDs from the UniProt Database v2016_11 (The UniProt Consortium, 2017), the RefSeq non-redundant protein sequence database v09/12/2016 (O'Leary *et al.*, 2016) and from all sequenced plant genomes available through Phytozome v9 (Goodstein *et al.*, 2012) (Supplementary Data). These sequences were subsequently reviewed, in their respective database, in order to determine if any of these sequences have associated publications which were missed in the initial literature review. Sequences were compiled and duplicate sequences retrieved from different databases were removed. Sequences that were identical except for deletions were determined using CD-HIT v4.6.6 (Fu *et al.*, 2012) and were considered as fragments and assigned the same ID, with the longest length sequence being used. All other alleles and isoforms were given a unique ID.

Alignment of the full phylogeny of the BAHD superfamily was generated as described above using the alignment of the characterised enzymes as a profile. Phylogenetic reconstruction was carried out using FastTree v2.1.9 SSE3, OpenMP (Price *et al.*, 2010) with 1000 bootstrap replicates (bootstrap alignments were generated using SeqBoot 3.696 from the PHYLIP package (Felsenstein, 1989) and were then used for phylogenetic reconstruction using FastTree as above). Bootstrap results were mapped onto the original tree using a python

script as described previously (Price *et al.*, 2010). Sequences were grouped into bootstrap supported major clades and sub-phylogenies were generated separately for each clade. Full-length sequences were used for the subclades and alignment, trimming and phylogenetic reconstructions were generated with PhyML as described above. All bioinformatic software was compiled from source code, and all analyses were performed on the WestGrid/ComputeCanada computational platform.

2.2.2. **Microarray Based Analysis of Transcriptional Abundance**

Public microarray expression data were compiled, normalized, and filtered from 43 experimental series, covering 17 *Populus* spp. and 697 array hybridizations as described previously (Guo *et al.*, 2014). Each sample microarray contained 61413 probes, and these probes represent most of the transcripts from *Populus* species. Only data from the target probes and target tissue are shown.

2.2.3. **Plant Growth and Sample Collection**

Xylem, mature leaf, bark, dormant buds, phloem, flushing bud, and expanding leaf tissue were harvested from four *P. trichocarpa* genotype 'Nisqually-1' clones grown in a field without irrigation at the University of Victoria. For xylem samples, the bark was stripped from a three- to five-year-old branch from the mid-crown and the layer of developing cells was scraped from the wood using a single-blade industrial razor blade. Likewise, the innermost layer of the developing bark, i.e. phloem, was harvested by scraping the peeled bark. The whole root was harvested from twelve one-month old *P. trichocarpa* 'Nisqually-1' plants grown hydroponically in 1/10 Hoagland's solution (Hoagland and Arnon, 1950) in growth chambers

under long-day conditions (16 h of light/8 h of dark, 25°C). Male catkins, female catkins, and seed tissue were harvested from east-facing, exposed side branches from the lower crown of 12 mature *P. trichocarpa* trees from the Bowker Creek basin close to the University of Victoria using a pole pruner. Samples were harvested and immediately frozen in liquid nitrogen. Tissues were ground into a fine powder under liquid nitrogen and stored at -80°C until used.

Ground plant tissues were extracted as previously described (Kolosova *et al.*, 2004). Dried RNA pellets were resuspended in TURBO DNase digestion buffer and digested with TURBO DNase according to the manufacturer instructions (Life Technologies). Samples were washed with phenol (pH 4.5):chloroform (1:1 v/v) to remove residual DNase. Samples were analysed on a NanoDrop 2000C in order to determine yield. RNA was visualised on a TAE/formamide gel (Masek *et al.*, 2005). cDNA synthesis was performed with 5 µg of total RNA in a 20 µL reaction with Oligo(dT)20 using SuperScript III reverse transcriptase as per manufacturer instructions (Life Technologies).

2.2.4. **qPCR**

The "Mix for CFX" (Life Technologies) was used as per manufacturer's instructions on a CFX96 (BioRad) thermocycler. Gradient PCR between 55 and 65°C was used to optimize the annealing temperatures for each primer set, and 59°C was chosen as an acceptable annealing temperature for all primers (Supplementary Table 1). Samples were manually pipetted into 96-well low-profile PCR plates (Thermo Scientific). Primers for elongation factor 1β (EF1β; Potri.009G018600), ubiquitin (Potri.014G115100.1) and ribosomal protein (RP; Potri.001G342500) which had previously been validated (Alber, 2016), were used as reference

transcripts. All samples were tested in triplicate; four biological replicates were used for xylem, mature leaf, bark, dormant buds, phloem, flushing bud, and expanding leaf; and three pooled replicates -- each consisting of at least three plants -- were used for root, male catkin, female catkin, and seed tissue. A single automatic cycle threshold for quantitation (Cq) was selected for each target using BioRad CFX manager 3.1. The change in Cq (ΔCq) values for each sample were determined by normalization to the geometric mean of the reference transcripts for that sample.

2.2.5. **RNAseq**

RNAseq expression data were provided by the POPCAN project (Corea *et al.*, 2017). A set of 435 *P. trichocarpa* ecotypes were previously collected by the BC Ministry of Forests, Lands and Natural Resource Operations. Each ecotype was assigned a unique accession ID when added into the BC Ministry of Forests, Lands and Natural Resource Operations collection, and subsequent clones from the original ecotype are referred to as being the same accession and having the same accession ID. A subset of these *P. trichocarpa* accessions was grown in a replicated common garden trial at the University of British Columbia (McKown *et al.*, 2014; Xie *et al.*, 2009). Leaf and xylem samples were harvested from nearly 200 accessions, which were selected to be representative of genotypes across the range. Samples for each accession were harvested between 11:00 and 13:00 over 2 days (July 3-4, 2012) during an extended period of stable, calm, and clear weather. Current-year developing xylem (excluding the cambium) was collected after being exposed in windows cut from bark at breast height from the north side of replicate five-year-old trees grown in the common garden, as described (Porth *et al.*, 2013). Duplicate developing leaves at a uniform growth stage (first fully unfurled leaf below the

shoot apex) were harvested from two independent current-year shoots derived from an individual clone of each accession grown in a clonal bank at the same location. For developing xylem, RNA was acquired from a total of 385 samples representing 195 accessions, and for leaves, RNA from 389 samples from 193 unique accessions was acquired. Most samples (182 and 181 accessions, respectively, for xylem and leaf) were harvested in duplicate or more highly replicated. Leaf tissue for RNA purification was ground using Precellys 24 tissue homogenizer (5 s, 5500 RPM x 2; tissue frozen in liquid N_2 between grindings). Xylem tissue was ground in liquid N_2 using a mortar and pestle. RNA was purified in a two-step procedure: the first step was carried out using PureLink® Plant RNA Reagent (Invitrogen, Life Technologies) following the manufacturer's protocol. The second step was carried out using RNeasy Plant Mini Kit (Qiagen) following the RNA clean-up and On-Column DNase digestion steps from the manufacturer's protocol. The quantity and quality of the purified RNAs was evaluated using a 2100 BioAnalyzer instrument (Agilent) with the Agilent RNA 6000 Nano kit (Agilent). RNA samples with RNA integrity number -- an (RIN) (Mueller *et al.*, 2004) -- greater than or equal to 7 were submitted to the Michael Smith Genome Sciences Centre (Vancouver, Canada) for non-strand specific library preparation and transcriptome sequencing. Sequencing was carried out on an Illumina HiSeq instrument with 75-base paired-end tags (PET). Samples were indexed at 6 per HiSeq lane with an average of 30 Gb sequence per lane obtained.

A local Galaxy (http://usegalaxy.org) pipeline was implemented for analysis of the raw RNA-Seq data obtained from the Genome Science Centre as described by (Hefer *et al.*, 2015). Paired-end reads were trimmed to remove low-quality reads and adapters using

Trimmomatic (a 4bp sliding window, an average quality of 20, reads shorter than 50bp were discarded) (Bolger *et al.*, 2014). Tophat 2.0.8 (Trapnell *et al.*, 2012) was used to map the trimmed reads (mean inner distance of 300bp, read anchor length of 8bp, and allowing for 2 mismatches) to V3.1 of the *P. trichocarpa* genome, containing 41,335 putative gene transcripts. Cufflinks v2.1.1 (Trapnell *et al.*, 2012) was used to calculate fragments per kilobase of transcript per million mapped reads (FPKM) values for each transcript (performing bias correction and multi-read corrections). FPKM values were provided by POPCAN (Hefer *et al.*, 2015) for co-expression analysis performed as part of this thesis. A Pearsons correlation coefficient was generated between each of the annotated gene transcripts in the *P. trichocarpa* genome V3.1 (Tuskan *et al.*, 2006) using R v3.4.0. A custom R script was made which generated pairwise Pearson correlations coefficient for every gene vs every other gene in the *Populus* genome using median-centred gene expression data across accession for the xylem and leaf datasets separately. The genes with Pearson correlation coefficient (r^2)> 0.6 for the HCTs were included in this analysis and identifiers were retrieved using Phytomine (Goodstein *et al.*, 2012) and MapMan (Thimm *et al.*, 2004).

2.3. Results and Discussion

2.3.1. **Phylogeny**

In order to determine which genes to consider as candidates for a particular function, phylogenetic sub-grouping and determining evolutionary relationship to characterized genes is a powerful tool. Many of the sequences that are included in phylogenies are from large-

scale sequencing projects and are based on computational gene modelling only. Therefore, the phylogenetic reconstruction performed here was anchored with high-quality sequence data available for characterised enzymes. Extensive literature reviews identified 119 BAHD enzymes with experimentally characterized biochemical functions (Supplementary Table 3), which were aligned and used for an initial phylogenetic reconstruction (Supplementary Figure 1). Expanding on this sequence collection, available sequences from Genbank, Phytozome and Uniprot were identified through BLAST searches using enzymes with known functions from each functionally distinct clade as baits. This resulted in a total of 5266 unique sequences from 313 species in 186 genera. The profile HMM for the initial alignment of functionally characterized enzymes was used as a template for expanding the alignment towards the entire publicly-available BAHD family. Adding these to the phylogeny allows subdivision of the family based on evolutionary relationship and the identification of putative orthologs, i.e. close relatives across species with the same function. Alignment and phylogenetic reconstructions grew increasingly computationally intense as the number of sequences increased; therefore, FastTree was used for the complete phylogeny as it requires less computational resources and completes in a shorter time than PhyML (Price *et al.*, 2010).

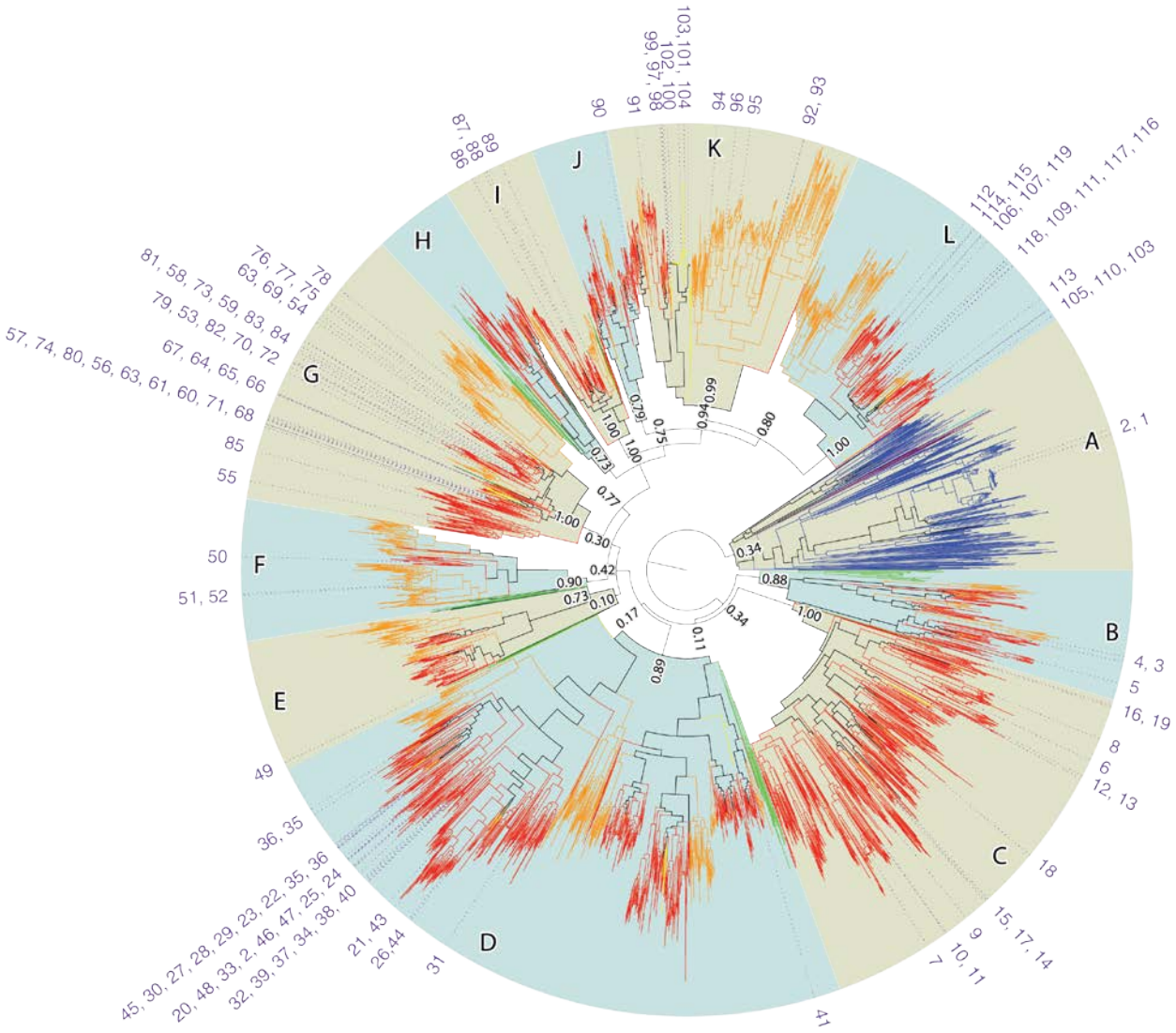


Figure 2-1. Phylogenetic Reconstruction of Available Putative BAHD Acyltransferases
A phylogenetic reconstruction which includes all ~5300 BAHD acyltransferases with available sequence data in the GenBank, UniProt, and Phytozome databases. Clade groupings are based on bootstrap support which is indicated on the figure. The identities of the characterized enzymes labelled in purple and the numbers are represented in Supplementary Table 3 and are used to infer functional relationships. Every line represents a unique protein sequence and branch colour coding is indicated in (Table 2-1).

Table 2-1. Colour Coding System Used for Branches in Phylogenetic Reconstructions.
Colour groupings are based on systematic groups to aid in visual interpretation of the phylogenetic reconstruction.

Taxa	Colour	# of sequences
Eudicots	Red	2937
Angiosperms excluding eudicots (primarily monocots)	Orange	1511
Gymnosperms	Yellow	157
Lycopods	light green	126
Bryophytes	dark green	22
Chlorophytes	light blue	5
Fungi	dark blue	472
Eukaryotes excluding fungi and viridiplantae	light purple	12
Bacteria and archaea	dark purple	13
Other	Black	1

The final phylogeny included ~5,300 BAHD protein sequences and identified several distinct phylogenetic clades, the grouping of these clades were statistically supported based on the bootstrapping method (Felsenstein, 1985), with groupings in which 0.70 of clades generated by random sampling with replacement remained the same (Figure 2-1). These clades were arbitrarily assigned letters A thru K. Previously the BAHD superfamily had been broken into 5 clades labelled I-V (D'Auria, 2006) and clades I, III, and V were subsequently subdivided (Tuominen *et al.*, 2011). In my phylogenetic reconstruction I have further subdivided clade Va into four major groupings (clade I, J, K, and L). The addition of sequences beyond the angiosperms allowed for the clarification of evolutionary context. Several of the clades are rooted by bryophytes and lycopods, indicating the ancient origin and preservation of the group. Independent phylogenetic reconstructions of each of these clades were generated to substantiate the sub-divisions (Table 2-2; Supplementary Figure 2 to Supplementary Figure 13).

Table 2-2. Summary of Clade Definitions Based on Phylogenetic Reconstructions
The clades defined below are based on the phylogenetic reconstruction of the BAHD acyltransferases. The previously published phylogenetic reconstructions are provided for reference. The substantial additional sequence information used in this work allowed for breakdown of the major clades into several additional subclades.

Clade	D'Auria Clade[1]	Tuominen Clade[2]	Bootstrap Supported Subclades
A			2
B	II	II	4
C	III	IIIa	4
D	I	Ia	3
E	III	IIIb	3
F	IV	IV	2
G	V	Vb	4
I	V	Va	2
J	V	Va	2
K	V	Va	3
L	V	Va	2

The enzymes within the BAHD superfamily are known to be quite promiscuous with respect to substrate specificity, so caution is warranted when drawing functional conclusions based on evolutionary relationship within these clades. Subclade phylogenies were generated using PhyML, with SPR and NNI for tree refinement, which provides a more accurate result than FastTree, but takes up to 100 times longer for computational analysis (Price *et al.*, 2010). Several of the clades contain few to no characterised members. Clade H does not contain any characterised enzymes, clades E and J each only contain one characterised member and clade F contains three characterised members. Group E-I consists of bryophyte and lycopods sequences, forming the base for sister groups E-II and E-III. This suggest separation of clade E

[1] D'Auria, J.C. 2006. Acyltransferases in plants: a good time to be BAHD. Current Opinion in Plant Biology **9**(3): 331-340.
[2] Tuominen, L.K., Johnson, V.E., and Tsai, C.-J. 2011. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. BMC Genomics **12**(1): 236-236.

from other BAHD clades prior to the bryophyte/vascular plant split some 400 million years ago. The one characterised member in group E is in group E-II and is an acetyl-CoA coniferyl alcohol transferase which is involved in the biosynthesis of the volatile ester isoeugenol, a prominent floral scent component of a *Petunia* hybrid (Supplementary Table 3 #49).

Clade J is subdivided into three distinct sub-clades. Group J-I contains primarily monocot sequences. Groups J-II and J-III each are rooted by an *Amborella* sequence and each contain both monocots and eudicots. This suggests divergence of the three clades early in or prior to angiosperm diversification. Only J-III contains a characterised member and it is involved in volatile compounds biosynthesis; this enzyme from *A. eriantha* primarily forms small esters particularly butyl acetate and butyl propionate (Supplementary Table 3 #90).

Group F-II is rooted by bryophyte, lycopods, amborella sequences, again suggesting early divergence within the land plants. All three characterised members are hydroxycinnamoyl amine transferases including an enzyme from *N. attenuata* which is responsible for coumaryl-putrescine (Supplementary Table 3 #50) and two isoforms of an enzyme from *H. vulgare,* which is responsible for the synthesis for the antifungal hydroxycinnamoyl-agmatine and its derivatives (Supplementary Table 3 #51-52). Given that a protein from both a eudicot and a monocot in this group can produce hydroxycinnamoyl-amides, this may be a common function of clade F proteins.

Group A-I consists of fungal sequences rooted by bacterial and archaea sequences. Group A-II consists solely of fungal sequences and contains the trichothecene 3-O-acetyltransferases which protect the fungi from mycotoxins that they produce. These mycotoxins are

responsible for many of the symptoms associated with Fusarium head blight which is caused by the pathogenic fungi *F. graminearum* and F. sporotrichioides (Supplementary Table 3 #51-52). BAHD proteins are predominantly found in plants, but this clade appears to be exclusively from non-plant species and may represent an early horizontal gene or frequent horizontal gene transfers from a plant source. It is noteworthy that plant-pathogenic fungi are frequently found among the fungi which contain BAHDs; however, fungi which affect humans and their crops may be over represented as sequence data is more likely to be available.

Clade B is rooted by the lycopod group B-I and contains three other groups, each containing gymnosperm, monocot, and eudicot sequences. The only characterised members of Group B-IV are acyltransferases involved in wax formation. A group of *Arabidopsis* epicuticular wax layer mutants, called *eceriferum* (*cer*) mutants are defective in long chain fatty acid biosynthesis. The *Arabidopsis* cer26 (Supplementary Table 3 #5) mutant cannot produce aliphatic chains longer than 30 carbons(Pascal *et al.*, 2013); the CER2 (Supplementary Table 3 #3) protein has been shown to interact with condensing enzymes and modify the chain length of the substrate produced from 28 carbons to 30 carbons in *Arabidopsis* (Negruk *et al.*, 1996). Similarly, the glossy 2 BAHD acyltransferase mutant from maize is also deficient in long chain wax formation (Supplementary Table 3 #4). Thus, the group as a whole appears to be primarily responsible for long chain fatty acid biosynthesis involved in wax formation.

Group C is primarily composed of eudicot sequences subdivided into four groups and containing members representing a wide substrate range, both related to the acyl groups transferred and to the acyl acceptors. Acetyl-CoA and flavonoids, however, are the most

common substrates of characterised members. Group C-I contains acetyl and malonyl transferases including an enzyme from *S. lycopersicum* shown to be involved in triacyl sugar biosynthesis (Supplementary Table 3 #6), the well-characterised acetyl-CoA: benzylalcohol acetyltransferase from *C. breweri* which forms benzyl acetate, a major constituent of floral scent (Supplementary Table 3 #8), and the anthocyanin malonyltransferase involved in salvianin colouration of *S. splendens* (Supplementary Table 3 #19). A C-I subgroup contains two characterised *N. attenuata* members that are responsible for spermidine transferases activity (Supplementary Table 3 #12, 13). Clade C-II contains acetyl transferases responsible for morphine biosynthesis (Supplementary Table 3 #16) and vinorine synthesis (Supplementary Table 3 #18). Group C-III contains acyl transferases that function on alkaloids and are involved in the biosynthesis of deacetylvindoline and minovincinine in *C. roseus* (Supplementary Table 3 #10-11) and capsaicin in *C. annuum* (Supplementary Table 3 #7). Group C-IV contains several acetyl transferases involved in biosynthesis of volatiles (Supplementary Table 3 #9, 14, 15, 17).

Group D breaks down into three subclades; one bryophyte/lycopod clade rooting two sister clades, again suggesting separation of this clade early during vascular plant evolution. D-III contains several well characterised malonyl-CoA flavonoid glycoside transferases (Supplementary Table 3 #21-40, 42-48). With the exception of a single amide acyltransferase from a *Torenia* hybrid (Supplementary Table 3 #20), it is exclusively composed of malonyl-CoA flavonoid glycoside transferases. Although enzymes in this group are generally characterised as malonyl-CoA flavonoid glycoside transferases, many enzymes in this group appear to be promiscuous in their choice of acyl-CoA substrates: in addition to malonyl-CoA, the group can

include donors as large as hydroxycinnamoyl-CoA's. In contrast to group D-III, group D-II is less well categorised, and the only characterised member of this group is the *Defective in Cuticular Ridges* (*DCR*) gene from *Arabidopsis*; this gene functions in the creation of triacylglyceride by acetylating diacylglycerides (Supplementary Table 3 #41).

Clade K is split into two groups. One of these, group K-I, contains only gymnosperm sequences from *T. baccata*. These enzymes are mainly related to the biosynthetic pathway of paclitaxel, a compound used in cancer therapy and sold under the brand name Taxol, and no other characterised enzymes are contained within this group. This grouping is likely biased in the phylogenetic reconstruction since there is a lack of sufficient gymnosperm sequence information and the pharmaceutical importance of Paclitaxel has resulted in the addition of a wealth of sequence information for genes related to its biosynthesis. Group K-II contains several hydroxycinnamoyl transferases. This group is enriched in monocot sequences and has been proposed to be a clade of BAHD acyl transferases involved in integration of hydroxycinnamates into grass cell walls, termed the Mitchell clade (Supplementary Table 3 #95-96). Two other enzymes from the monocot *I. hollandica*, which form hydroxycinnamoyl flavonoid conjugates, also fall into this clade (Supplementary Table 3 #92-93), as well as an alcohol acetyl transferase involved in volatile ester biosynthesis from *M. sapientum* (Supplementary Table 3 #94).

Group L contains almost exclusively plant volatile ester synthesis enzymes with the exception of a 13-hydroxylupanine O-tigloyltransferase from *Lupis albinus* which functions in alkaloid biosynthesis. 3-hexen-1-yl acetate in *A. thaliana* (Supplementary Table 3 #108) and *N.*

tabacum (Supplementary Table 3 #116) is formed from enzymes in the clade. Enzymes in *Petunia* x *hybrida* cv. Mitchell 'Diploid' (Supplementary Table 3 #117) and *C. breweri* (Supplementary Table 3 #109) form benzyl benzoate in flowers. A wide range of enzymes from *V. pubescens* (Supplementary Table 3 #118), *V. labrusca* (Supplementary Table 3 #119), *C. melo* (Supplementary Table 3 #110, 111 105) *M. pumila* (Supplementary Table 3 #114, 115) *F. ananassa* (Supplementary Table 3 #112) and Actinidia ssp. (Supplementary Table 3 #106, 107) form a diverse array of fruit ripening volatile esters including benzyl and butyl acetate. The majority of enzymes with known HCT activity which act on small alcohols and amines, were found in clade G. Additionally, no characterised enzyme in this clade has been shown to have any activity other than HCT. A few characterised HCTs exist outside this clade, but they have unusual substrates such as C30 alcohols, large flavonoid conjugates, or hormones such as melatonin. Thus, based on these phylogenetic analyses, the HCTs responsible for HCC formation are likely to be found in this clade. Within the clade, multiple subdivisions are apparent, defining 3 well-supported (>0.80) sub-clades and another cluster with moderate (0.63) bootstrap support (Figure 2-2). The clade we have labelled G-I contains only benzoyl-CoA:anthranilate N-benzoyltransferase sequences from *Dianthus caryophyllus*. In our phylogenetic reconstructions, they often formed the base of G-II; however, their grouping with G-II had poor support (0.33) so they were kept as a separate group. These sequences are annotated as benzoyl-CoA transferases as *Dianthus caryophyllus* produces the N-benzoylanthranilate as the precursor of several sets of dianthramides; however, the original biochemical characterization indicates that the activity of these enzymes with 4-coumaroyl-CoA far exceeds their activity with benzoyl-CoA (Yang *et al.*, 1997), suggesting that these are

also primarily HCTs similar to the other members of clade G. Two out of the four groups contain only two characterised members, making predictions regarding their functions problematic. Clade G-II exclusively contains sequences from eudicots; including the *Arabidopsis* spermidine hydroxycinnamoyl transferase (SHT) which is involved in pollen-coat formation (Supplementary Table 3 #55) and the HCT from *Trifolium pretense* which acts on malate, potentially being involved in the synthesis of the terpenoid, phaselic acid. While clade G-III contains exclusively monocot sequences, one enzyme from *O. sativa* has been characterised as producing glyceryl ferulate, acting on glycerol and ferulate as its preferred substrates. The other enzyme appears to be an HQT that forms chlorogenic acid. The largest clade G-IV contains sequences from all land-plant taxonomic groups, and also contains the majority of characterised HCTs which catalyze a wide diversity of HCT reactions. Notably, within G-IV, a number of hydroxycinnamoyl shikimate transferases (HST) form subclade G-IV-α, rooted by a *P. patens* gene (Figure 2-3). The hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HST) clade within the HCT clade typically has very short branch lengths, indicating that there is very little evolutionary divergence between these enzymes. The HSTs that have been characterised are involved in lignin biosynthesis and are essential to plant growth and survival. As such, it would be plausible to assume that the short branch length is indicative of purifying selection to maintain the shikimate specificity of this enzyme; additional selection tests would be warranted for confirmation. Although in general, the G-IV-α. clade has very short branch lengths, there are a few nodes with long branch lengths that appear to shift position, which may explain why the G-IV boot strap support is low (0.63).

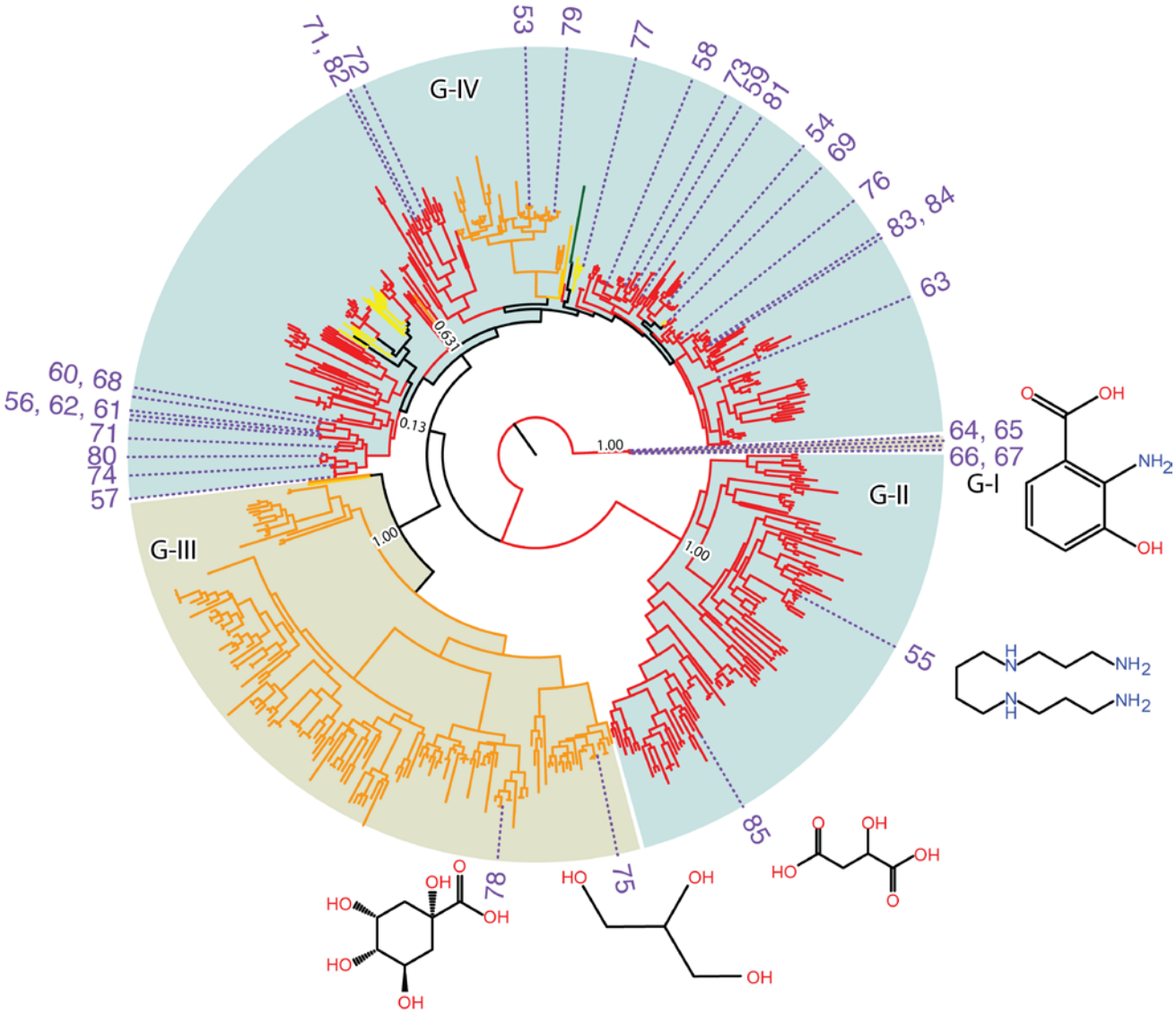


Figure 2-2. Phylogenetic Reconstruction of the Putative HCT Clade (Clade G)
The phylogenetic reconstruction shows that the large majority of the enzymes that have been characterized belong to group G-IV. The primary substrate of each of selected characterised enzymes is displayed for Groups G-I, G-II, and G-III. Branch colour coding as indicated in Table 2-1. Bootstrap values indicate the fraction of bootstrap replicates with the same topology.

As all of the sequences available in NCBI were included in this analysis, the phylogenies generated are strongly biased towards species and genes that have previously been considered important by humans. As such, the lignin HCTs in eudicots and chlorogenic acid forming HCTs are highly overrepresented, and this may have caused sub clades to shift

position. Bryophytes, lycopods, and gymnosperms are underrepresented, and expansion of the sequence coverage in the bryophytes, lycopods, and gymnosperms may help to define the clades. One node that frequently shifts position in G-IV-α, contains monocots, including the chlorogenic acid -specific *Pancium virgatum* HQT, and the *Avena sativa* 5-hydroxyanthranilate-specific HCT. The other node that shift positions contains exclusively eudicots, including the rosmarinic acid synthases from *Melissa officinalis* and *Solenostemon scutellarioides* and the HQT from *Lonicera japonica.*

The *P. patens* (Pp3c2_29140V1.1) isoform in G-IV-α has not yet had a biological function associated with it, and *P. patens* does not produce lignin; however, the cloned enzyme can form many substrates including the S-lignin and G-lignin precursor caffeoyl shikimate (Eudes *et al.*, 2016). Although *Physcomitrella* does not produce lignin it produces a waxy cuticle enriched in hydroxycinnamoyl conjugates which has been proposed to be an essential evolutionary adaptation in the transition of plants to land (Renault *et al.*, 2017). Being phylogenetically located at the base of the lignin specific HCT subclade, it is tempting to speculate that this gene may be involved in the formation of this phenolic enriched cuticle in *Physcomitrella*. This gene may have undergone positive selection early in vascular plant evolution and evolved to a shikimate specific lignin specific biosynthetic gene. The ability to synthesize lignin is essential for vascular plants, and this may explain why a copy of this gene has been maintained in all plants and the members in this subgroup are more similar to each other than in other groups. As such, it would be plausible to assume that the short branch length is indicative of purifying selection to maintain the shikimate specificity of this enzyme; however, additional selection tests would be required for confirmation.

Group G-IV-β appears to contain exclusively enzymes involved in chlorogenic-acid biosynthesis; therefore, this clade is the most likely candidate when looking for an enzyme involved in chlorogenic acid biosynthesis. In group G-IV-β there is an uncharacterised *A. thaliana* gene (At5g57840) which has been previously annotated as a potential anthranilate benzoyl transferase, due to high similarity to a protein in *Dianthus caryophyllus*. The expected gene product of this enzyme has not been characterised in *A. thaliana*, but the phylogenetic evidence indicates that this enzyme could be a chlorogenic acid synthase; however, chlorogenic acid cannot be detected in *A. thaliana* extracts indicating that the functional products of this clade may be more diverse than the phylogeny would indicate. HCTs can be very promiscuous with their substrate specificity, for example the *alcohol acyltransferase* from *F. vesca* can act on substrates as varied as methanol, cinnamoyl alcohol and eugenol. The substrate promiscuity demonstrated by these enzymes indicates that genomic analysis may be insufficient for the determination of substrate specificity.

Because there are a large number of sequences and only a few characterised enzymes, this phylogenetic reconstruction does not provide a definite picture of enzyme function; however, it provides a solid basis for selection of candidate genes within the large BAHD superfamily. It identified nine HCT candidates and suggested distinct functions related to lignin and soluble HCC biosynthesis in *Populus*. Overall, Clade G contains hydroxycinnamoyl alcohol transferases (HCTs) along with a few amine transferases. The vast majority of enzymes with known HCT activity are contained within clade G. A few characterised HCTs exist outside of this clade, but they have unusual substrates such as C30 alcohols, large flavonoid conjugates, or hormones such as melatonin. Most of the characterised HCTs which use small alcohols and amines as

substrates belong to this clade. Additionally, all of the characterised enzymes in this clade (33) have HCT activity. Thus, based on these phylogenetic analyses, the HCTs that are responsible for HCC formation are likely to be found in this clade (Supplementary Table 3; Figure 2-3).

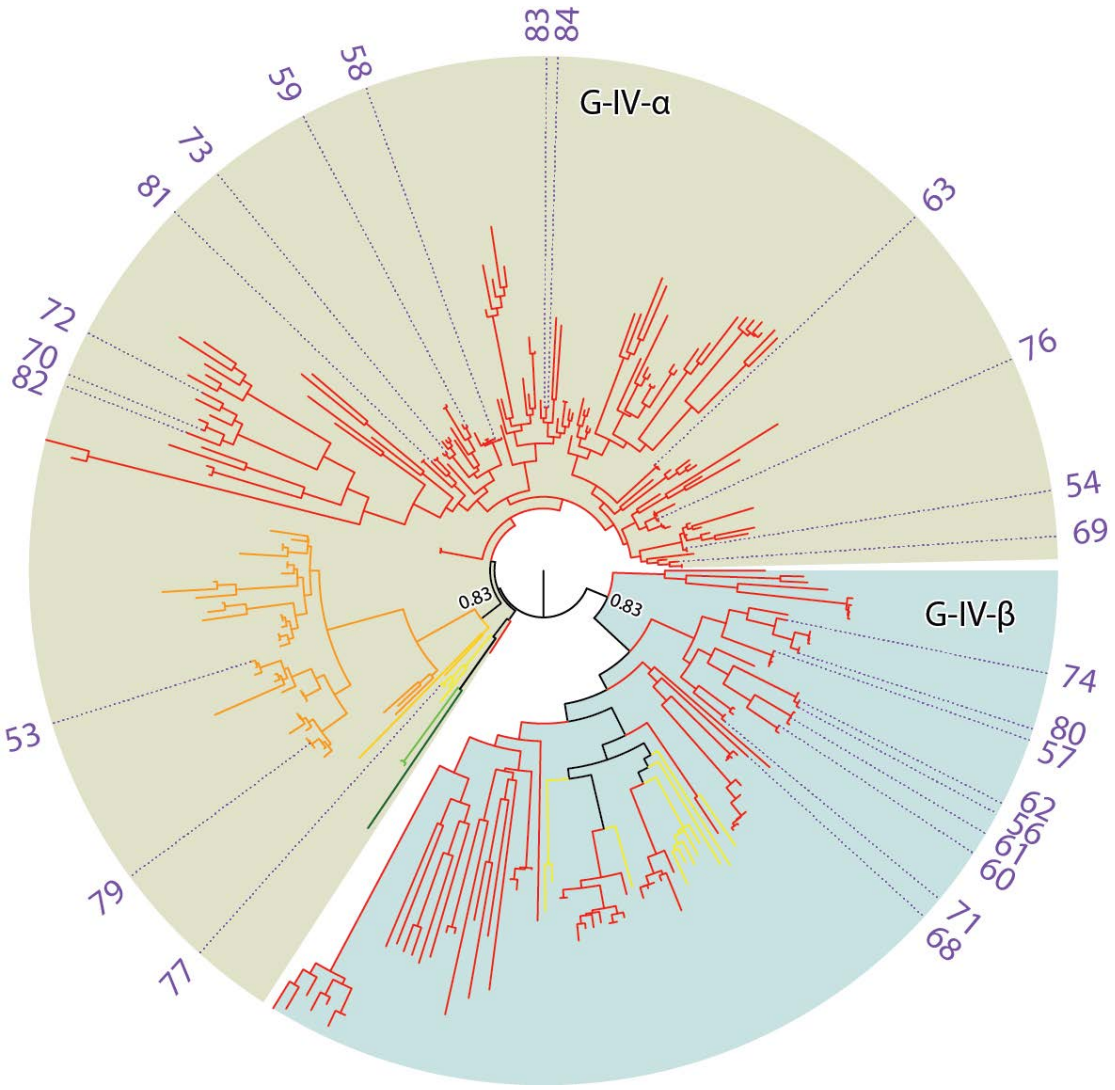


Figure 2-3. Phylogenetic Reconstruction of BAHD Group G-IV
Clade G-IV represents the primary hydroxycinnamoyl transferases clade with G-IV-α containing primarily lignin associated enzymes and G-IV-β containing primarily chlorogenic acid associated enzymes. The identities of the characterized enzymes labelled in purple are included in Table 2-1 and are used to infer functional relationships. Every line represents a unique protein sequence and branch colour coding is as indicated in Table 2-1.

2.3.2. **The HCT Family in *Populus***

Nine genes in *P. trichocarpa* were identified as being potential HCT homologs out of a total of 106 *P. trichocarpa* BAHD family members identified here (Table 2-3). Due to the recent whole-genome duplication, most of the poplar isoforms exist in pairs; the isoforms are named such that the pairs share the same letter identifier and were arbitrarily assigned either 1 or 2. Although there are instances in which a pair do not have identical function (Chedgy *et al.*, 2015), they frequently have a similar function. When designing transgenics, there is a strong risk that gene pairs are redundant and if knock-outs are generated for only one of the paralogs that there would be no detectable phenotype. Two of the HCT isoforms in *P. trichocarpa* (*PtHCTE1*, Potri.006G165200; *PtHCTE2*, Potri.018G109900) show significant similarity to an HCT involved in hydroxycinnamoyl-spermidine production during pollen-coat formation and have been placed into clade G-II. All other *P. trichocarpa* HCT candidates are located in clade G-IV. Among them, two of the *P. trichocarpa* HCT isoforms fall into group G-IV-α (*PtHCTA1,* Potri.001G042900; *PtHCTA2*, Potri.003G183900) and are closely related to a characterised HST involved in lignin biosynthesis (Supplementary Table 3 #54-84) . They thus constitute candidates for shikimate ester biosynthesis leading to lignin biosynthesis, predominantly during wood formation. The remaining five genes in group G-IV-β (*PtHCTB*, Potri.018G105500; *PtHCTC1*, Potri.018G104800; *PtHCTC2*, Potri.018G104700; *PtHCTD1*, Potri.005G028100; *PtHCTD2*, Potri.005G028000) are candidates that are likely to be involved in other transferase activities, and may be responsible for the formation of the diverse HCCs found in Populus. These genes are closely related to each other and have a high sequences similarity, placing them in the same distinct section of clade G-IV-β. This clade thus far has

been exclusively characterised as having chlorogenic-acid forming HCTs; however, the duplication events leading to 5 different isoforms present the possibility for functional divergence. These are the most likely candidates for enzymes responsible for the diverse soluble HCCs that are found in *Populus*.

Table 2-3. Genes Identified in *P. trichocarpa* Encoding Hydroxycinnamoyl Transferase Candidates
Genes were selected based on proximity to characterized hydroxycinnamoyl transferases in the phylogenetic reconstruction (Figure 2-2). *P. trichocarpa* genome ID numbers are included with all isoforms, variants, and transcripts being represented by the primary ID.

Gene	Group	Potri ID
PtHCTA1	G-IV-α	Potri.001G042900
PtHCTA2	G-IV-α	Potri.003G183900
PtHCTB	G-IV-β	Potri.018G105500
PtHCTC1	G-IV-β	Potri.018G104800
PtHCTC2	G-IV-β	Potri.018G104700
PtHCTD1	G-IV-β	Potri.005G028100
PtHCTD2	G-IV-β	Potri.005G028000
PtHCTE1	G-II	Potri.006G165200
PtHCTE2	G-II	Potri.018G109900

Phylogenetic reconstructions can decipher evolutionary relationships which frequently indicates orthology, i.e., functional equivalence, but proper biochemical characterisation is required to assign function. Phylogenetic analyses are a valid starting point for providing functional hypotheses; but caution must be taken, as sometimes an enzyme responsible for a function can have no obvious sequence similarity to other enzymes with a similar function (i.e. they acquired the same function through convergent evolution), and vice versa -- enzymes with high sequence similarity may have distinct catalytic activities because sometimes only a few amino acid changes at the active site may be sufficient to change substrate specificity. Complementing phylogenetic analyses with alternative assessment,

such as transcript expression profiling, is therefore necessary to further substantiate functional hypotheses.

2.3.3. **Tissue and Organ Expression profiling**

The abundance of the candidate HCT transcripts were extracted from publicly available microarray data (Guo *et al.*, 2014). Due to the high similarity between *P. trichocarpa* isoforms, probes were not available to distinguish between the PtHCTA1/PtHCTA2 and the PtHCTE1/PtHCTE2 pairs; however, based on microarray expression data, at least one of the transcripts in the PtHCTA1/PtHCTA2 pair is more highly expressed in xylem tissue and at least one of the transcripts in the PtHCTE1/PtHCTE2 pair is more highly expressed in catkins, indicating that their respective roles may be lignin formation in xylem and pollen coat formation in catkins. PtHCTB could not be readily detected in any of the target tissues. PtHCTD1 and PtHCTD2 appeared to be highly expressed in outer bark and phloem, indicating that these transcripts may have a role in the production of the secondary metabolites produces there. The primary tissues in which the HCC chlorogenic acid is produced is young leaves (Tsai *et al.*, 2006) and PtHCTC1 and PtHCTC2 were highly expressed in these tissues (Figure 2-4).

For further confirmation of the expression of the target gene transcripts, qPCR was carried out (Figure 2-5). Primers designed for PtHCTA2 did not successfully amplify the target; however, PtHCTA1 showed high expression in xylem, leaves, and male catkins. In the microarray data PtHCTA was much higher in the xylem than it was in the leaves or male catkins. Lignin is produced in many tissues including vascular tissue in the leaves and the male flowers.

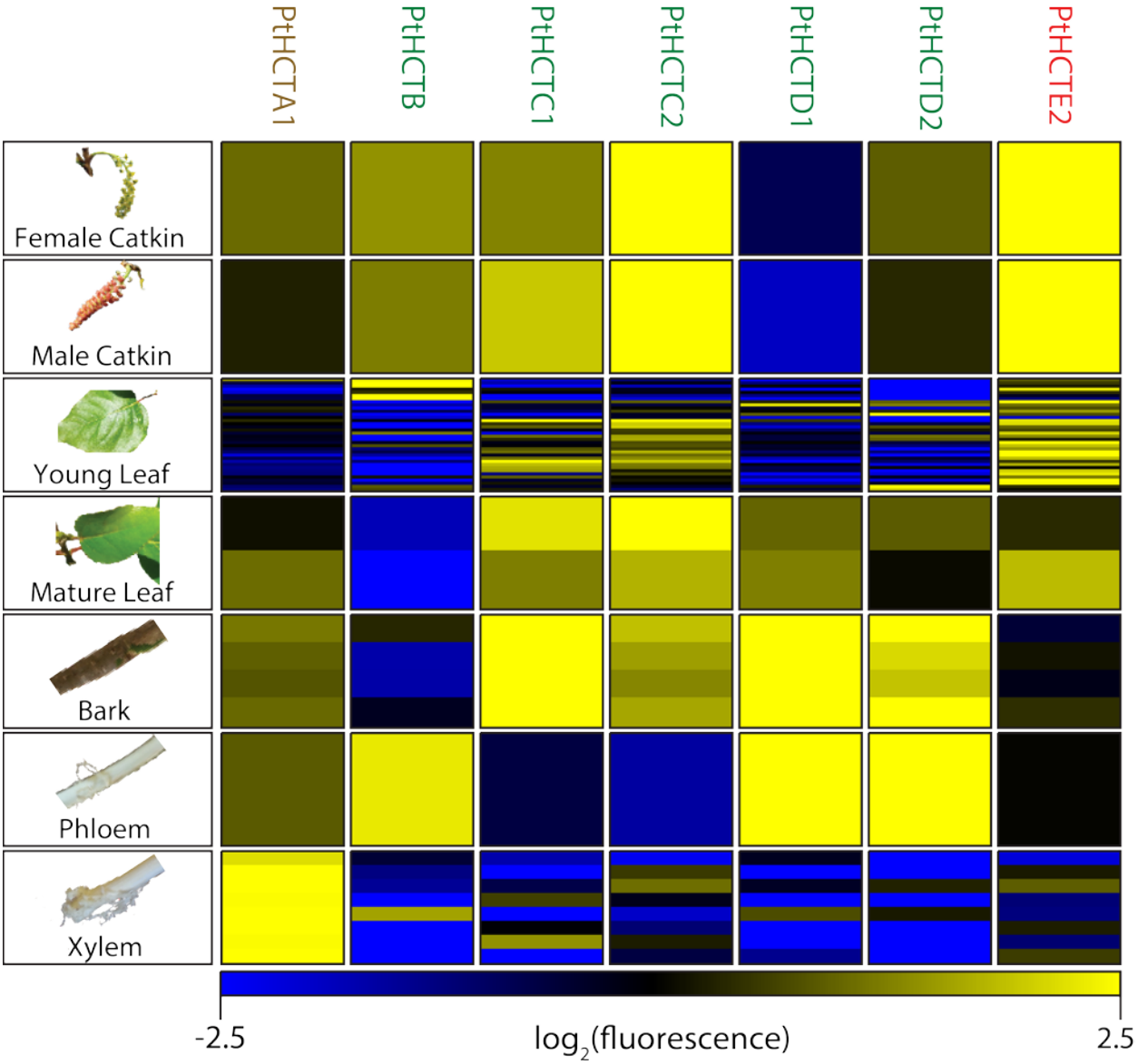


Figure 2-4. Microarray Expression of PtHCT Genes in Selected Tissues
Microarray data collated from publically available data (Guo *et al.*, 2014) show PtHCTA, PTHCTC, and PTHCTCE expression in xylem, mature leaf, and male catkin, respectively. Each line represents a biological replicate available for that tissue type in the compiled data. PtHCTA and PtHCTE isoforms could not be distinguished.

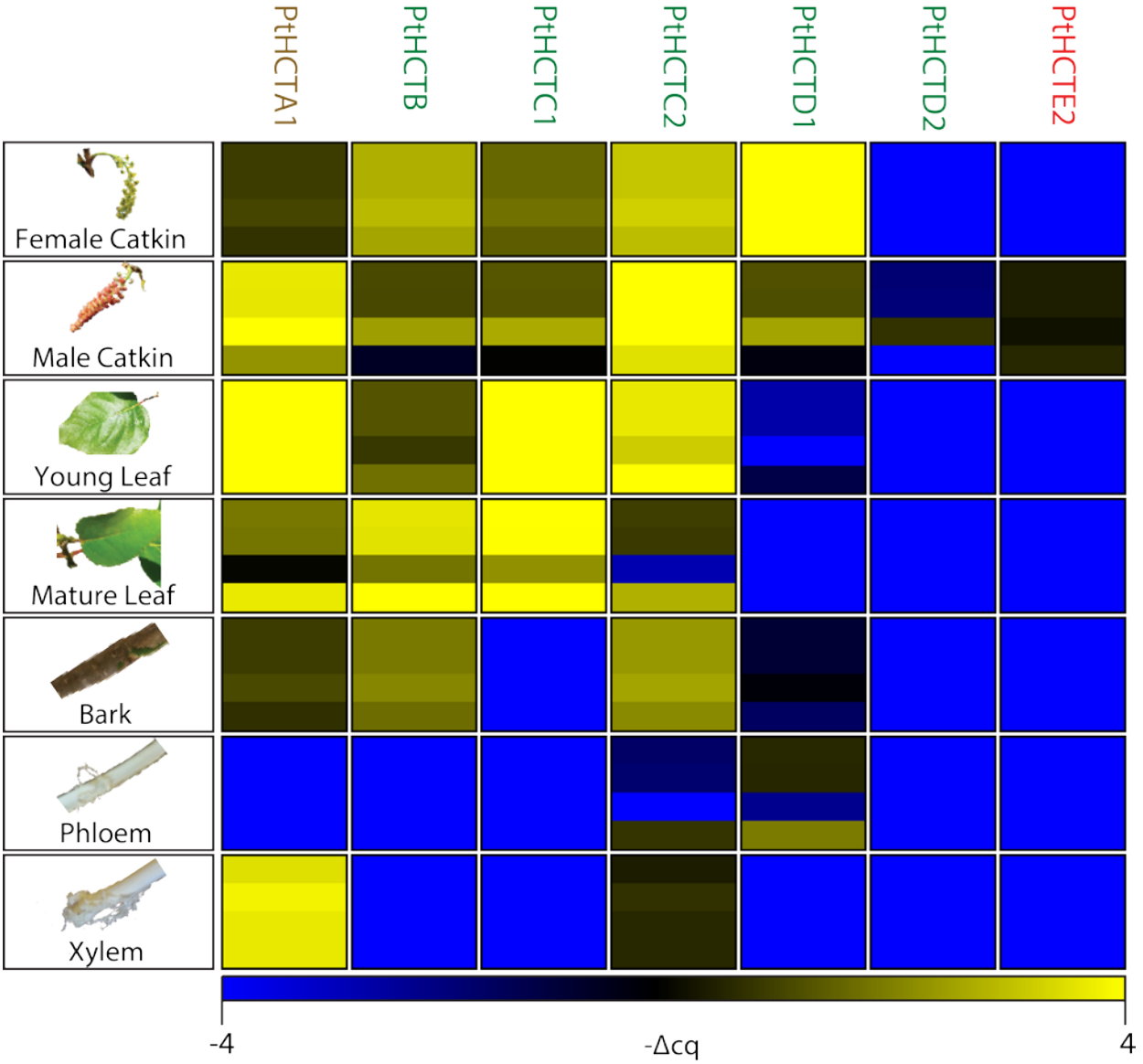


Figure 2-5. qPCR Expression Data for PtHCT Targets in Selected Tissues
PtHCTA1 and PtHCTE2 show strong expression in xylem (wood) and male catkin (pollen), respectively. Samples with no detectable expression assigned the lowest value for colouring. Each line represents a biological replicate available for that tissue type in the compiled data. (n=4 biological; 3 technical).

The discrepancy between the qPCR and microarray data for this tissue may be due to different sampling practices. Whole leaf and male flowers, including the lignified petiole were extracted for qPCR. Primers designed for PtHCTE1 also did not successfully amplify the target. Although other enzymes showed higher expression in male catkins in the qPCR dataset than PtHCTE2, it was most highly expressed in male catkins and was expected to be in male catkin based on the phylogenetic reconstruction. These qPCR data support the hypothesis that these genes are the primarily candidates for lignin formation in xylem and pollen coat formation in male catkins. PtHCTC1 and PtHCTC2 both showed the strong expression in young leaves supporting their assignment as the leaf chlorogenic acid HCT.

2.3.1. **Targeted Expression Profiling Using RNASeq**

Using RNASeq data acquired from xylem and leaf tissue from almost 200 natural *P. trichocarpa* accessions, the relative expression of all HCT candidates were examined (Figure 2-6). PtHCTA1 and PtHCTA2 appear to be highly expressed in xylem tissues and moderately expressed in leaves indicative of lignin biosynthesis. PtHCTC1 and PtHCTC2 have strong expression in young leaves making them possible candidate's for chlorogenic acid production. Overall combining the phylogenetic data and the expression data (Figure 2-7): only PtHCTA1/2 show strong consistent expression in xylem, consistent with lignin function. PtHCTC1/2 show expression in young leaf tissue that varies in the microarray treatments, consistent with secondary metabolite production. PtHCTE2 only shows expression in male catkin in the qPCR analysis; however, the microarray data is not consistent.

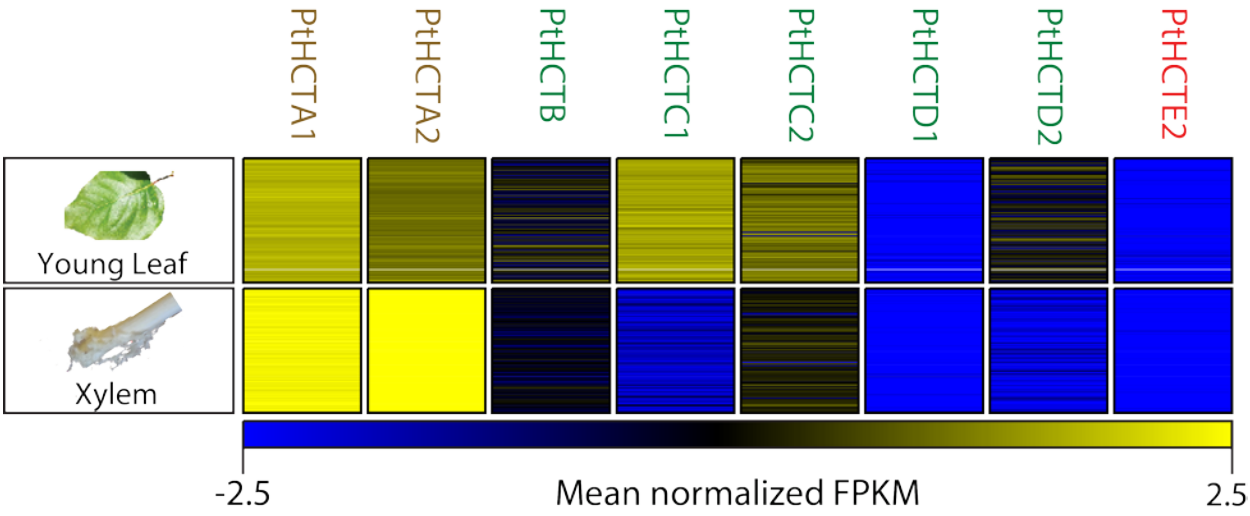


Figure 2-6. RNASeq Expression of HCTs in Mature Leaf and Xylem Tissues
RNASeq expression of the HCTs shows that PtHCTA1/PtHCTA2 were most highly expressed in xylem, and that expression of PtHCTA1/PtHCTA2 and PtHCTC1/PtHCTC2 are highly expressed in mature leaf. Each line represents a different *Populus* accession.

Networks of tightly co-regulated genes can be determined based on their co-expression in natural populations (Porth *et al.*, 2013). In order to obtain further indications of putative functions, gene-co-expression analyses were performed comparing transcript abundance variation across accessions in a given tissue of each *HCT* with all other transcripts in the *Populus* genome. With the exception of PtHCTA2, none of the HCT gene transcript abundances were well correlated with any other transcripts in the *Populus* transcriptome in the xylem or leaf RNASeq data (Pearson correlation coefficient [r^2] > 0.6). Although I was not able to design a successful primer for PtHCTA2, it was found to be highly expressed in the xylem and co-expression analysis of this transcript indicated that it was tightly co-expressed with other cell-wall related genes (Table 2-4).

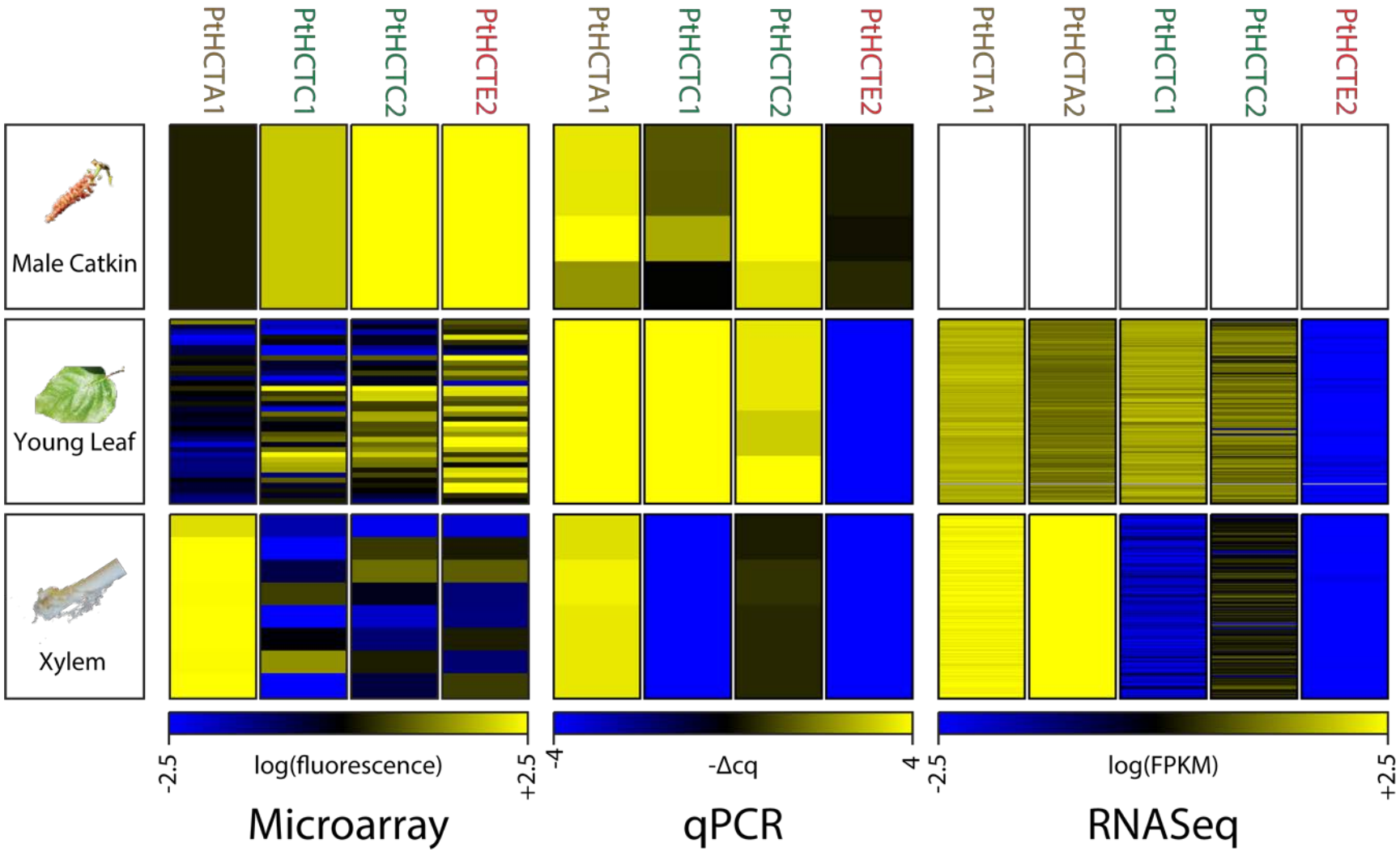


Figure 2-7. Combined Microarray, qPCR, and RNASeq Expression Data for PtHCTA1/2, PtHCTC1/2, and PtHCTE2
Of the candidates only PtHCTA1/2 show strong consistent expression in xylem, consistent with lignin function. PtHCTC1/2 show expression in young leaf tissue that varies in the microarray treatments which is consistent with secondary metabolite production in leaves. PtHCTE2 only shows expression in male catkin in the qPCR analysis which expected tissue based on phylogenetic placement; however, the microarray data is not consistent.

Of the transcripts that were co-expressed with PtHCTA2, several of the transcripts are annotated as enzymes involved in aromatic amino acid metabolism and the general pathway and include: an arogenate dehydratase (Potri.004G188100), two phenylalanine ammonia lyases (Potri.008G038200, Potri.016G091100), two cinnamate 4-hydroxylase (Potri.013G157900, Potri.019G130700), and two 4-coumarate--CoA ligase isoforms (Potri.001G036900, Potri.003G188500). Likewise, a ferulic acid 5-hydroxylase (Potri.007G016400), involved in G- and S-monolignol biosynthesis is tightly co-expressed. Furthermore, other enzymes and regulatory proteins involved in secondary cell-wall biosynthesis and xylem differentiation are among the co-expressed transcripts. Galacturonosyl transferase 12 (Potri.001G416800001G416800) is involved in the biosynthesis of xylan and pectin in woody tissues, and the gene Irregular Xylem 15 (Potri.007G047000) functions in xylan biosynthesis and deposition (Brown *et al.*, 2011). Four proteins are transcriptional regulators with known functions in xylem differentiation and secondary cell-wall biogenesis: the myb-domain protein 43 (Potri.017G130300), and three NAC (NAM (no apical meristem), ATAF1 and −2, and CUC2 (cup-shaped cotyledon)) domain-containing protein 73s (Potri.007G135300, Potri.011G058400, Potri.017G016700). These NAC domain-containing protein 73s are similar to the *A. thaliana* SNDs involved in the regulation of fibre formation (Zhong *et al.*, 2007). The remainder have unclear function and are not similar to well-characterised plant proteins; however, based on enzymes with characterised functions in plants with a pattern of expression similar to PtHCTA2, it appears that this gene transcript is correlated with the shikimate, phenylpropanoid, and monolignol pathways.

Table 2-4. Top 25 Genes Co-expressed with PtHCTA2 in Xylem Across 195 Natural Accessions of *P. trichocarpa*

Gene transcript abundances most highly correlated (Pearson's) with PtHCTA2 across 195 natural accessions of *P. trichocarpa*. FPKM values were provided by POPCAN (Corea *et al.*, 2017). A custom R script generated pairwise Pearson correlations coefficient for each of the annotated gene transcripts in the P. trichocarpa genome V3.1 (Tuskan et al., 2006) vs every other annotated transcript in order to determine highly co-regulated genes in the natural population. The transcript pairs with Pearson correlation coefficient (r2)> 0.6 were included in this analysis and identifiers were manually retrieved from Phytomine and MapMan databases. Secondary cell-wall formation transcripts appear to be over-represented.

Potri ID	r^2	Mapman	Phytomine
Potri.003G183900	1.00	Hydroxycinnamoyl transferase	Shikimate O-hydroxycinnamoyltransferase
Potri.001G416800	0.87	Galacturonosyl transferase 12	Galacturonosyl transferase 12
Potri.010G095900	0.87	Protein kinase family protein	Interleukin-1 receptor-associated kinase 4
Potri.004G188100	0.86	Arogenate dehydratase 6	Prephenate dehydratase family protein
Potri.008G038200	0.86	Phenylalanine ammonia-lyase 1	Phenylalanine ammonia-lyase.
Potri.007G016400	0.86	Ferulic acid 5-hydroxylase 1	Cytochrome P450
Potri.017G130300	0.86	Myb domain protein 43	Myb family transcription factor
Potri.013G157900	0.86	Cinnamate 4-monooxygenase	Trans-cinnamate 4-monooxygenase
Potri.005G141300	0.85		
Potri.001G036900	0.85	4-Coumarate--CoA ligase 2	4-Coumarate-CoA ligase
Potri.007G135300	0.85	NAC domain-containing protein 73	No apical meristem family protein
Potri.006G167200	0.85	Cytochrome P450 reductase 2	Cytochrome P450 reductase
Potri.011G058400	0.84	NAC domain-containing protein 73	NAC domain-containing protein
Potri.017G016700	0.84	NAC domain-containing protein 73	No apical meristem family protein
Potri.008G198100	0.84	Endo/exonuclease/phosphatase	Endo/exonuclease/phosphatase
Potri.012G023400	0.84	Auxin-responsive family protein	SAUR family protein
Potri.003G188500	0.83	4-Coumarate-CoA ligase 2	4-Coumarate-CoA ligase
Potri.010G076200	0.83		Targeting protein for Xklp2
Potri.005G211000	0.83		
Potri.015G075300	0.83	Kinesin light chain-r elated	Kinesin light chain
Potri.016G091100	0.83	Phenylalanine ammonia-lyase 1	Phenylalanine ammonia-lyase
Potri.004G001800	0.82		
Potri.019G130700	0.82	Cinnamate 4-hydroxylase	
Potri.007G047000	0.82		Protein Irregular Xylem 15 related
Potri.016G141900	0.82		
Potri.010G088200	0.82		

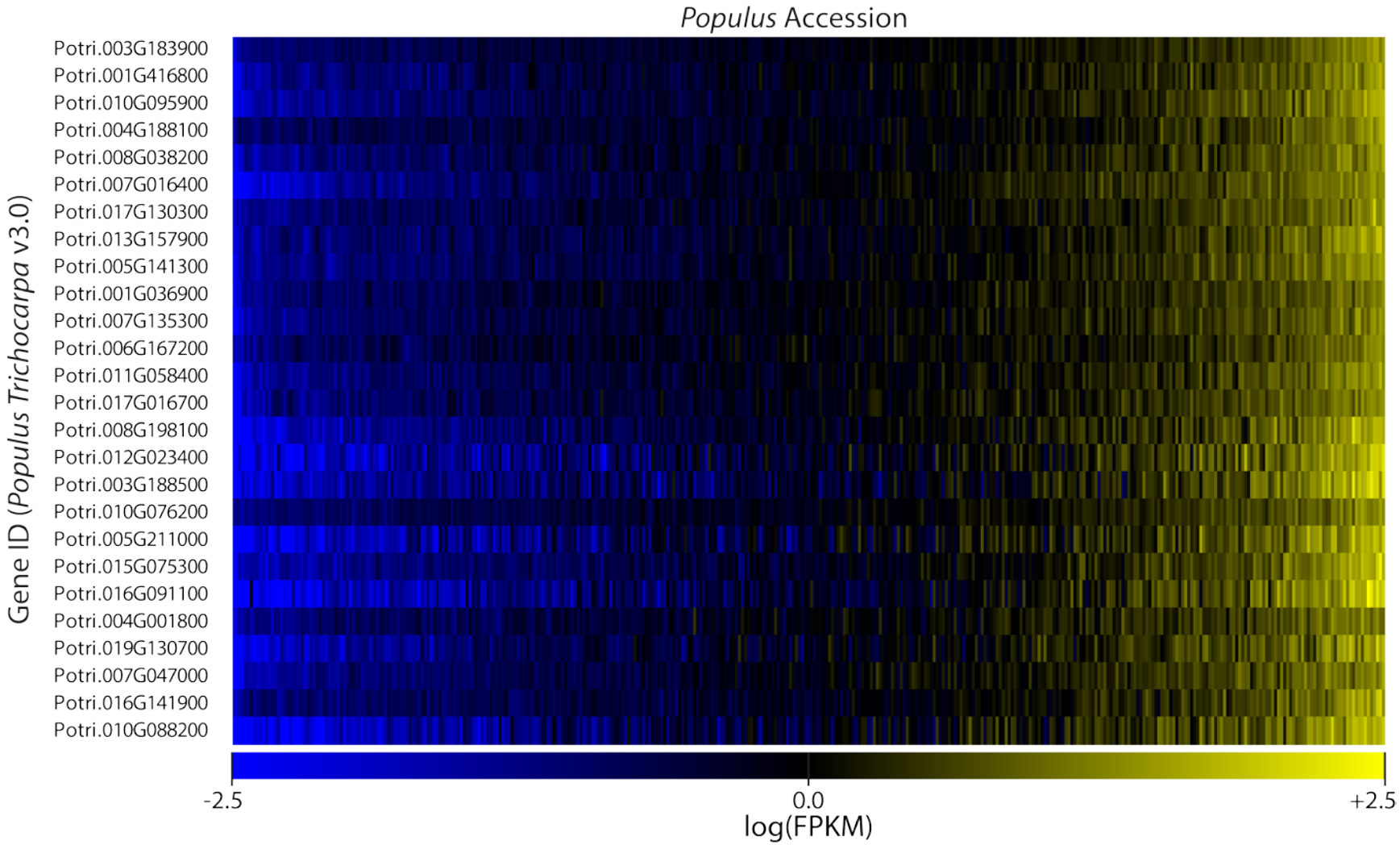


Figure 2-8. Co-expression Heat Map of the 25 Genes Most Highly Correlated with PtHCTA2
The transcript abundances determined by RNASeq are shown for the transcripts shown in Table 2-4. Linear colour gradient indicates mean centred and normalised $\log_2$(FPKM). Transcripts (Table 2-4) are sorted (top to bottom) by Pearson correlation with PtHCTA2 and each column represents a *Populus* accession sorted by average expression level with each column representing a different accession.

2.4. Conclusions

The large majority of characterised HCTs are within the BAHD superfamily of enzymes. By generating phylogenies based on the amino acid sequences of all BAHD enzymes, candidates were identified based on evolutionary relationships to characterised members. This placed BAHDs with HCT function primarily into clade G. This provided an evolutionary framework for which genes to include or exclude in further analyses. Clade G contained nine HCT candidates in *Populus*, and phylogenetic position within clade G provided my first hypothesis for functional diversity, which was further supported by transcript-expression profiling. Two candidate genes (Potri.001G042900, PtHCTA1; Potri.003G183900, PtHCTA2) were hypothesised to be responsible for caffeoyl-shikimate formation during secondary xylem (i.e. wood) formation (caffeoyl-shikimate is a precursor in the formation of lignin, the biopolymer that imparts mechanical stability to wood). One candidate gene (Potri.018G109900, PtHCTE2) was connected to the formation of caffeoyl-spermidine in male catkins. This compound functions in pollen coat formation. One candidate gene (Potri.006G165200, PtHCTE1) was not expressed in any tissue according to any of the methods that I had attempted potentially indicating a loss of function due to redundancy. Two candidate genes (Potri.018G104700, PtHCTC1; Potri.018G104800, PtHCTC2) were identified as candidates for the formation of soluble HCCs in leaves and bark based on transcript abundance in these tissues. Phylogenetics, coupled with expression profiling, proved to be an appropriate framework for generating a functional hypothesis which can now be addressed through targeted biochemical or reverse genetic experiments.

2.5. Contributions

Cuong Le: performed the sequence collection, annotation and phylogenetic reconstruction, analysed the microarray data, designed the qPCR experiments generated and analysed the qPCR data, assisted in the annotation and analysis of RNASeq data and performed the correlation analysis. **Anneke Hylkema, Jacqui Duggan:** extracted RNA, validated primers and assisted in the qPCR experiments. **Charles Hefer, Oliver Corea:** provided mapped RNAseq data and generated the expression values. **YingYu Huang:** performed the microarray compilation and normalization. **Carol Parker:** provided copy editing. **Christoph Borchers** and **Jürgen Ehlting:** provided supervision, funding, copy editing, and laboratory materials and equipment.

3. Hydroxycinnamoyl Transferases Are Necessary for Lignin Biosynthesis in Poplar

3.1. Introduction

A large proportion of the planet's carbon biomass that is readily available for human use consists of wood. Wood, also known as secondary xylem, is a tissue that contains extensively thickened secondary cell-walls. In 1838, Payen indicated that wood does not represent a uniform substance but is cellulose and a ligneous material (Payen, 1838). This ligneous material was later termed lignin and is the major structural component of plant secondary cell-walls, particularly in wood-forming tissues (Vanholme *et al.*, 2010). It is responsible for the structural integrity of the cell wall and provides mechanical stability to woody structures stiffening and strengthening the stem. Lignin serves to waterproof cell walls, enables high negative-pressure water transport, and allows the extension of the plant's vascular system (Boerjan *et al.*, 2003). Lignin also serves to protect plants against biotic and abiotic stresses, and the structure of lignin adapts to various stimuli including wounding (e.g., by herbivores), infection, and ions in the cell-wall structure (Vanholme *et al.*, 2010). The stress of gravity can also directly influence the composition of lignin. "Reaction wood" is formed in response to physical forces, primarily gravity. The formation of high-cellulose tension wood resists the downward bend that would occur due to gravity as branches extend. In angiosperms, tension wood is formed in response to tensile stresses pulling the branch away from gravity. In contrast, gymnosperms produce high H-lignin compression wood that resists compressive

forces. In this way that plant can change the composition of lignin in response stimuli to provide wood with differential properties.

Lignin is primarily composed of three subunits which differ in the hydroxylation and methoxylation of their phenyl group: p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) subunits, although a small number of other subunits have been identified (Vanholme *et al.*, 2010).These subunits are derived from phenylpropanoids which are a diverse group of plant secondary metabolites which are derived from the shikimate pathway and which have a wide array of ecological functions and 3, 4-hydroxylation). Phenylalanine ammonia lyase (PAL) is the key entry point into the phenylpropanoid pathway and the first enzyme in a series of conserved biochemical reactions that convert phenylalanine to 4-coumaroyl-CoA, linking aromatic amino acid metabolism and secondary metabolism. 4-coumaroyl-CoA can then undergoes a series of reactions to produce p-coumaryl, coniferyl, and sinapyl alcohol, which produce H-lignin, G-lignin, and S-lignin, respectively (Boerjan *et al.*, 2003). The composition and structure of the polymer varies greatly among species, cell types, and even cell wall layers. Gymnosperm lignin consists primarily of G subunits with a small amount of H subunits. Angiosperms generally have mostly S subunits with varying levels of G subunits. Monocots have higher H-lignin content, and eudicots typically have almost none (Boerjan *et al.*, 2003). As the H-subunit component is generally very small or almost non-existent in eudicots, a common measure is the S:G ratio with the assumption the H is not significant. In some species, lignin can be up to 85% S with an S:G ratio of 5.7 (Vanholme *et al.*, 2010), but in most species the S:G ratio is lower -- for example S:G ratios for *Populus* species are generally between 1.8 and 2.3 (Davison *et al.*, 2006). In addition to across-species differences, the

compositions of the G, S, and H subunits in lignin differs by cell type and plant age and can even be dissimilar at the level of the individual cell-wall layers, as revealed by Raman, IR, and UV micro-spectrophotometry, or by immunolabeling of the secondary cell-wall with antibodies reacting with specific lignin substructures (Vanholme *et al.*, 2010).

As previously indicated, the hydroxylation/methoxylation of the phenyl ring determines the identity of the three major lignin subunits: H-lignin (4-hydroxy), G-lignin (4-hydroxy, 3-methoxy), and S-lignin (4-hydroxy, 3,5-methoxy). When lignin is being forms it occurs *via* a radically polymerised end-wise reaction leading to primarily β-coupled subunits including β–O–4, β –5, β – β C, 5–5, 5–O–4, and β –1 linkages with the β–O–4 linkage being the most common and also the most easily degraded (Boerjan *et al.*, 2003). Depending on the subunits that are used during the polymerization different chemical linkages are possible due to the reactivity of the moieties at the 5 and 3 positions, which in turn affect the degradability and physio-chemical properties of wood.

H-lignin is derived directly from 4-coumaroyl-CoA via reduction to the corresponding monolignol 4-coumaryl alcohol. In contrast, the shikimate specific hydroxycinnamoyl transferase uses the 4-coumaroyl-CoA produced in the general phenylpropanoid pathway to produce 4-coumaroyl-shikimate, which has been proposed to be the precursor of both S-lignin and G-lignin. Although there have been indications that there is an alternative pathway (Vanholme *et al.*, 2013b), HSTs have been shown to be involved in lignin biosynthesis in several plant species, as previously indicated. HSTs position after the branch point leading to H-lignin biosynthesis but upstream of S-lignin and G-lignin makes it an attractive target for

modifying the composition of lignin. Employing the reverse reaction, HST has also been proposed to be involved in the conversion of caffeoyl-shikimate to caffeoyl-CoA for G-lignin and S-lignin biosynthesis, but an alternative pathway involving a caffeoyl-shikimate esterase (CSE) has also been proposed (Ha *et al.*, 2016). However, *Populus* extracts do not show CSE activity which is evident in *Arabidopsis* (Wang *et al.*, 2014) although the poplar genome contains *CSE* homologs.

HSTs have been shown to be involved in G- and S-lignin biosynthesis in several plant species, and I have shown in Chapter 2 that the HST gene transcript I identified were co-expressed with lignin-related gene transcript and were highly expressed in lignin-rich tissues. In this chapter, I show that *Populus* INRA 353-38 (*Populus tremula x Populus tremuloides*) HCTA (PttHCTA) -- which I had identified as a likely HST candidate -- is involved in the biosynthesis of lignin in *Populus,* and I also show that a knock-down alters the content and composition of lignin in wood. A strong decrease in the PttHCTA expression leads to dwarfing in *Populus,* which is similar to that observed in other HCT knock-downs in *Arabidopsis thaliana* (Hoffmann *et al.*, 2005; Li *et al.*, 2010) and *Medicago sativa* (Shadle *et al.*, 2007). Some mutants, however, appear to have a change in lignin composition *without* a change in growth rate. As stated above, these intermediate mutants should be assayed for wood properties that are industrially beneficial, including cellulosic biofuel production.

3.2. Materials and Methods

3.2.1. Cloning of HCTA RNAi-Construct

RNAi knock-down vectors were designed as previously described (Wesley *et al.*, 2004) using the pHANNIBAL vector (Supplementary Figure 15) The transcript sequences for genes that had previously been identified as HST genes were compiled (Chapter 2). When genes occurred as paralogous pairs (PtHCTA, PtHCTC, PtHCTD, and PtHCTE) transcript sequences were aligned and a consensus sequence was generated for each pair. Primer3 software (Untergasser *et al.*, 2012) was used to generate primers producing 250-350 bp fragments of the transcript sequences. Primers were designed in order to generate a cassette targeting both transcripts by manually excluding regions that were not similar between the two homologs. Appropriate restriction-enzyme sites were added to the ends of each primer as indicated in Supplementary Table 5 -- for example, XhoI and KpnI were added to the 5'-ends of sense-orientation-fragment forward and revers primers, respectively, for PtHCTA; and HindIII and ClaI were added to the antisense-orientation fragment forward and reverse primers, respectively for PtHCTA (Figure 3-1.).

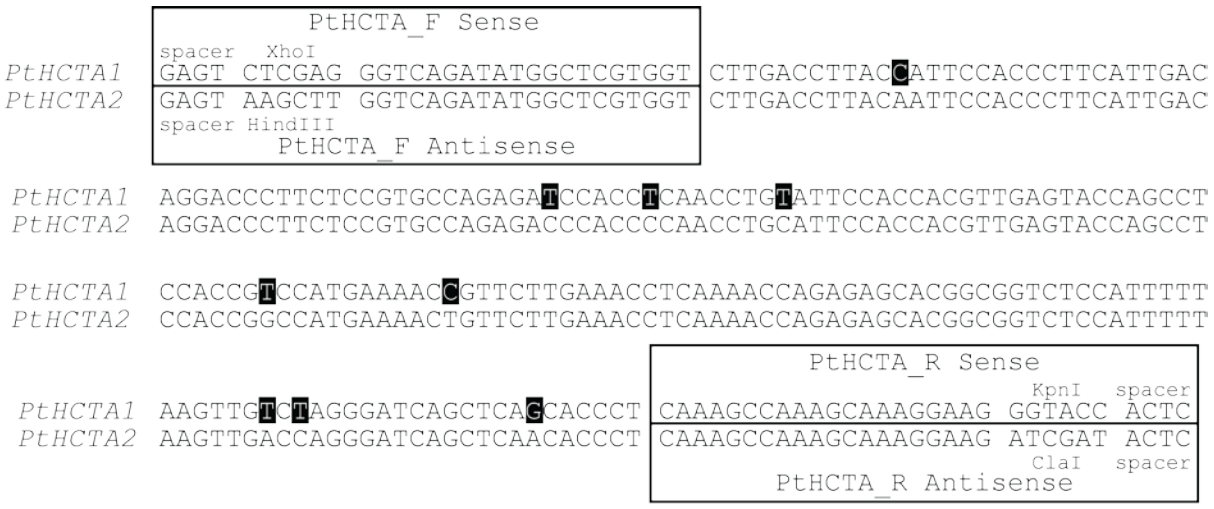


Figure 3-1. Schematic of the Cassettes Designed to Generate RNAi Knock-Downs
Primers were designed to flank the target sequence and provide the desired restriction sites for the PtHCTA RNAi knock-down cassette (XhoI on the forward sense-orientation fragment primer, KpnI on the reverse sense-orientation fragment primer, HindIII on the forward antisense-orientation fragment primers, and ClaI on the reverse antisense-orientation fragment primer.) Primer sequences are available in Supplementary Table 3-3. The fragment cloned into pHANNIBAL was identical to *PtHCTA2* (Potri.003G183900). *PtHCTA1* (Potri.001G042900) is provided for reference. Differences between the sequences of *PtHCTA1* and *PtHCTA2* are shaded in the *PtHCTA1* sequence.

DNA was extracted from leaves from greenhouse-grown *P. trichocarpa* 'Nisqually-1', and fragments were amplified using Q5 DNA polymerase (New England Biolabs), using the manufacturer's recommended conditions and an annealing temperature of 60°C. PCR products were used directly. DNA was digested using the appropriate enzyme (XhoI, KpnI, HindIII, ClaI, and/or NotI) according to the manufacturer's instructions (New England Biolabs). Plasmids were chemically transformed into and propagated in *E. coli* XL1 Blue. Plasmids were purified using an EZ-10 Spin Column Plasmid DNA Miniprep Kit (Biobasics). The digested sense-orientation PCR products were ligated into a linearized pHANNIBAL vector (Wesley *et al.*, 2001) which was provided by C. Peter Constabel, University of Victoria. After confirmation of the insertion of the sense-orientation fragment via Sanger sequencing (P-5: GGGATGACGCACAATCC, I-5: ATAATCATACTAATTAACATCAC) the pHANNIBAL vector

containing the sense-orientation fragment was linearized with the appropriate enzymes and the corresponding antisense-orientation PCR product was ligated into the vector. Insertion of the anti-sense fragment was also confirmed via Sanger sequencing (P-3: GAGCTACACATGCTCAGG I-3: TGATAGATCATGTCATTGTG). The final pHANNIBAL cassette containing both the sense-orientation and the antisense-orientation fragments was extracted from the vector using NotI and cloned into linearized pART27 binary vector (Gleave, 1992). The final construct was purified and transferred, via electroporation, into *Agrobacterium tumefaciens* strain GV3101 cells which were provided by C. Peter Constabel, University of Victoria.

3.2.2. **Plant**

3.2.3. **Transformation**

In vitro tissue-culture-grown *Populus* plants were grown on Murashige and Skoog (M&S) media (Murashige and Skoog, 1962)) solidified with 14 g/L agar. M&S media contains 4.33 g/L M&S Salts (Phytotech laboratories), 200 mg/L-glutamine (Sigma), 100 µg/L indole butyric acid (IBA), 2 µg/L biotin, 100 mg/L myo-inositol, 20 g/L sucrose, and M&S vitamins and is adjusted to pH 5.8. M&S vitamins contain 1 mg/L nicotinic acid, 1 mg/L pyroxidine HCl, 1 g/L thiamin HCl, and 1 g/L calcium pantothenate. Plant transformations were performed as previously described (Meilan and Ma, 2007) with adaptations (Yoshida *et al.*, 2015). Leaves were excised from the plants, were wounded using a scalpel blade, and were co-cultivated on callus induction media (CIM) solidified with 14 g/L agar at 22 ^{o}C in the dark for 2 days with engineered *A. tumefaciens* strain GV3101. CIM contains M&S media, 10 µM 1-

naphthaleneacetic acid (NAA), 10 µM 2ip, 50 mg/L kanamycin, 200 mg/L timentin, and 200 mg/L cefotaxime. The explants were washed with sterile water four times, transferred to fresh CIM plates, and subcultured by transferring to fresh media every 2 weeks for 6 weeks. The explants were then transferred to shoot induction medium (SIM) solidified with 14 g/L agar and subcultured on fresh SIM plates every 3 weeks for 4 months. SIM contains M&S media, 0.2 µM thidiazuron (TDZ), 50 mg/ml kanamycin, 200 mg/L timentin, and 200 mg/L cefotaxime. Explants with multiple shoots were transferred onto shoot elongation medium (SEM) solidified with 14 g/L agar and subcultured every 3 weeks for 4 months. SEM contains M&S media, 0.1 µM 6-benzylaminopurine (BAP), 50 mg/L kanamycin, 200 mg/L timentin and 200 mg/L cefotaxime. Regenerated shoots were transferred into M&S media with 50 mg/L kanamycin, which was solidified with 14 g/L agar. Plants were subcultured at three- to six-month intervals; longer growth periods were used to improve survival of the fragile *Ptthcta* knock-down lines. Lines that had successfully rooted for three generations were selected for further analysis.

3.2.4. **Plant Growth**

Transgenic plants were micro-propagated *in vitro* on M&S media with 50 mg/L kanamycin, which was solidified with 14 g/L agar, in growth chambers under long-day conditions (16 h of light/8 h dark). These plants were the source of material for physiological experiments. Three-month-old plants were removed from *in-vitro* tissue culture, rinsed in water, and planted in Sunshine #4 soil mix (Sungro Horticulture). Plants were maintained in a high-humidity misting chamber (70% humidity; 22°C) in a Venlo-style greenhouse for 4 weeks, and were then transferred to 1 gallon pots containing growth medium containing 4 L Sunshine Mix #4

(Sungro Horticulture), 21.4 g Prohort Standard (21-7-14; Crop Production Services), 2.9 g Micromax Micronutrients (Scotts-Sierra Horticultural Products Company), 11.4 g Dolomite Lime, 1.1 g -Super Phosphate (Horticultural Grade; 0-20-0).). Plants were maintained in a greenhouse and watered daily using a drip irrigation system. Plant height was measured weekly from the base of the plant to the apex.

3.2.5. **Sample Collection**

Ten weeks after transplanting into soil, the plants were harvested. The whole stem -- from between the 7th and 9th internode -- was flash frozen in liquid nitrogen for qPCR analysis. The stem from between the 9th and 10th internode was harvested, dried, and cut into 1-mm cubes using a scalpel blade for electron microscopy. The stem from between the 11th and 12th internode was harvested and air dried for lignin analysis. The stem from between the 12th and 13th internode was harvested, and flash frozen in liquid nitrogen for sectioning.

3.2.6. **qPCR**

For the phloem and xylem samples, the stems were partially thawed. Bark was stripped from a branch and the layer of developing cells was scraped from the wood sample using a blade. Likewise, the innermost layer of the developing bark (i.e., the phloem) was harvested by scraping the peeled bark. Ten mg (wet weight) of xylem scrapings were manually ground under liquid nitrogen using a mortar and pestle. RNA was extracted as previously described (Kolosova *et al.*, 2004). Samples were resuspended in TURBO DNase digestion buffer, and digested with TURBO DNase according to the manufacturer's instructions (Life Technologies). Phenol (pH 4.5):chloroform (1:1; v:v) cleanup was performed to remove DNase. Samples were

analysed using a NanoDrop 2000C (Thermo Fisher Scientific) to determine the yield. The RNA was visualised on a TAE/formamide gel (Masek *et al.*, 2005). cDNA synthesis was performed from 5 µg of RNA in a 20-µL reaction with Oligo(dT)20 primers, using SuperScript III reverse transcriptase, according to the manufacturer's instructions (Life Technologies).

SYBR Select Master Mix for CFX was used according to the manufacturer's instructions (Invitrogen) on a BioRad CFX96 thermocycler under the control of BioRad CFX Manger v3.1. Gradient PCR between 55°C and 65°C was performed to determine the acceptable annealing temperature ranges for each primer set (Supplementary Table 4). A temperature of 58°C was used as the annealing temperature for all primers. Samples were manually pipetted into a 96-well plate (Thermo Fisher Scientific). Primers for *elongation factor 1β (EF1β)* (Potri.009G018600), *ubiquitin (ubi)* (Potri.014G115100.1) and *ribosomal protein (RP)* (Potri.001G342500) were used as reference transcripts. Primers were evaluated using RNA extracted from wild-type *Populus* INRA 353-38. All samples were analysed in quadruplicate. BioRad CFX v3.1 was used for the initial data analysis and normalization, and a single threshold for cycle of quantitation (Cq) was used per target. Average Cq values for technical replicates were determined and the ΔCq were calculated for each matched pair using the geometric mean of the three reference transcripts for normalization. The sum of the squared standard error was calculated for each sample set where appropriate.

3.2.7. **Lignin H-S-G Ratio**

Wood samples were stripped of bark and pith, and were air dried. The dry wood samples were ground into a fine power using a Model 4 Wiley Mill (40 mesh) and processed for lignin

analysis as described previously (Robinson and Mansfield, 2009). Samples were hydrolysed with 2.5% boron trifluoride etherate and 10% ethanethiol, in dioxane (v/v) at 100°C for 4 h, with hourly agitation. Samples were adjusted to pH 3-4 with sodium bicarbonate, and tetracosan (5 mg/ml in methylene chloride) was added as an internal standard. Samples were mixed with 2 ml of water and 1 ml of methylene chloride, and allowed to settle, and the organic phase was collected for lignin analysis. Samples were dried and resuspended, and 20 µL of each sample was derivatised by combining 20 µL of resuspended sample with 20 µL of pyridine and 100 µL of N,O-bis(trimethylsilyl) acetamide, and incubating for 2 hours at 25 °C. Separation was performed on a 30-m RTX5ms 0.25-mm ID column on a Hewlett Packard 5890 series II GC/MS. Inlet and detector temperatures were set to 250°C, using a temperature gradient of 130°C, 0 min; 130°C, 3 min; 250°C, 123 min; 250°C 128 min.

3.2.8. **Electron Microscopy**

One-to-two mm cubes of wood were cut using a razor blade. An adapted protocol was used for embedding in the epoxy resin Quetol 651, as previously described (Abad *et al.*, 1988). Wood cubes were rinsed in distilled water for 60 minutes. They were then dehydrated in 3:1 Quetol 651: water (v/v) for 1 day. Samples were transferred to 100% Quetol 651 and dehydrated for an additional day. Samples were infiltrated with the Quetol formulation for 3 days and then transferred to fresh Quetol formulation for an addition day in order to fully impregnate the samples. The samples in Quetol were then allowed to polymerise at 60°C for 1 day.

All polymerised blocks were trimmed for TEM sections. Silver/gold coloured sections were cut, picked up onto 100 mesh copper grids containing a carbon-coated Formvar film and stained for 10 minutes in 5% uranyl acetate and 4 minutes in 5% lead citrate.

The sections were visualised using a Jeol JEM-1011 Transmission Electron Microscope at 80 kV. Digital images were captured with a Gatan Erlangshen ES1000W CCD camera. Images were taken at 2,500-fold magnification initially, and high-magnification images were taken at 10,000 fold. Cell wall thickness was measured using ImageJ 1.48k. All cells visible on each image were measured.

3.3. Results

The *P. trichocarpa* genome contains a total of nine HCT-gene family members. Based on phylogenetic position and expression profiling, two were identified as HCT-candidates related to lignin biosynthesis (Chapter 2). Due to a recent whole genome duplication, most genes in *P. trichocarpa* have a paralogue with high sequence similarity -- including PtHCTA1 and PtHCTA2 (92.4% nucleic acid similarity; 96.1% amino acid similarity). As these genes are nearly identical and may have redundant functions, knockout constructs were designed manually from an alignment of all PtHCTs to target regions where the genes pairs were highly similar to each other, but distinct from other HCTs. In initial analysis, knock down was attempted in *Populus* INRA 717-1B4 (*P. tremuloides* x *P. alba*) using RNAi and -- although transgenics were successfully generated for all of the other HCT pairs (Supplementary Table 5), none of the plants that regenerated from a transformation with the PtHCTA construct were successfully transformed, indicating the possibility of a severely lethal phenotype

(Supplementary Table 5). Although transgenic lines in the *Populus* INRA 717-1B4 background could not be recovered, transformation of *Populus* INRA 353-38 (*P. tremuloides* x *P. tremula*) with the PtHCTA construct yielded a similar number of transformants for the HCTA construct compared to the other lines (Supplementary Table 5)

3.3.1. *HCTA* Down Regulation via RNA Interference

Putative RNAi knock-down lines were generated and regenerated plants were screened for transcriptional down-regulation in developing xylem using qRT-PCR. Primers for PtHCTA1 and PtHCTA2, as well as three previously published reference gene primers, which also have been used for other species, were re-evaluated for use in *Populus* INRA 353-38 and showed acceptable linearity (Supplementary Figure 14) and acceptable efficiency during PCR (Supplementary Table 4). A total of five transgenic lines originating from independent transformation events targeting both *PtHCTA1 and PtHCTA2* (and thus referred to as Ptthcta), and one transgenic line with an empty vector without an RNAi target, were analysed, each including four clonal replicate plants grown in a greenhouse. Based on a t-test ($p < 0.01$), two lines (Ptthcta-1 and Ptthcta-2) showed down-regulation of both Ptthcta1 and Ptthcta2, and three lines (Ptthcta-4, Ptthcta-8, and Ptthcta-9) showed down-regulation of Ptthcta2 only. Based on a Tukey's range test ($p <0.01$), Ptthcta-9 did not show down-regulation of either PttHCTA1 or PttHCTA2, and Ptthcta-4 did not show down-regulation of PttHCTA2 (Figure 3-2). PttHCTA2 showed little variation in expression among the empty vector control replicates, but PttHCTA1 expression was much more variable (Figure 3-2). Based on RNAseq and microarray data, Pt*HCTA2* expression, but not Pt*HCTA1* expression, was highly correlated with other lignin-associated gene transcripts, indicating that PtHCTA2 may be the primary HCT

involved in lignin metabolism (Chapter 2). The lower variation of expression levels of HCTA2 compared to HCTA1 seen in the qPCR may be indicative of its "house-keeping" function. The lines chosen for further study showed a strong average knock down of PttHCTA2 ranging from 99% in Ptthcta-2 to 86% in Ptthcta-4 (Ptthcta-1, 88.9%; Ptthcta-2, 99.1%; Ptthcta-4, 85.6%; Ptthcta-8, 89.2%, Ptthcta-9, 91.8%); however, the variation in residual *HCTA2* expression among clonal replicates for knock-downs Ptthcta-2, Ptthcta-8, and Ptthcta-9 was substantial.

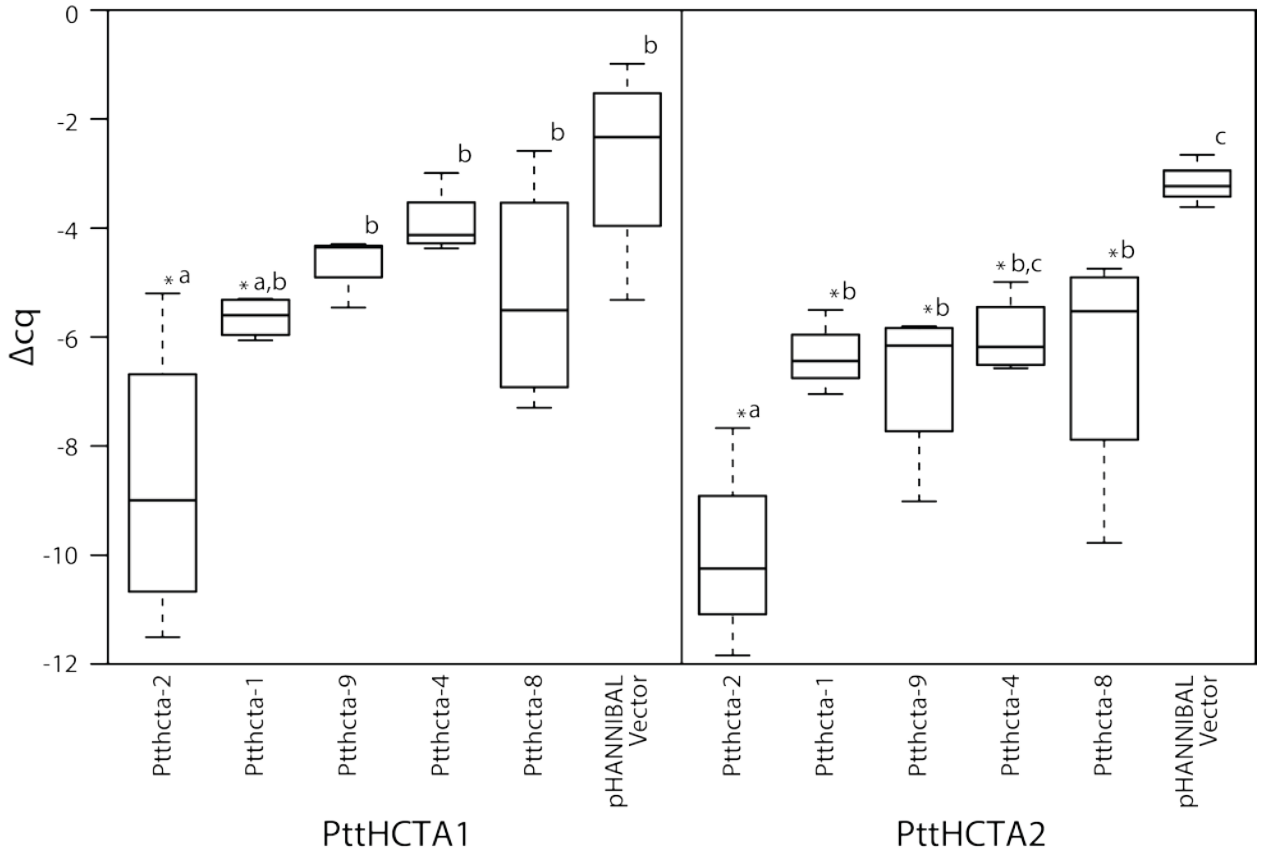


Figure 3-2. Relative Expression of PttHCTA1/PttHCTA2 in Transgenic RNAi Knock-Down Lines
The relative expression of the putative lignin associated HCTs were determined by qPCR using primers specific for each target in order to determine the decrease in transcript abundance of the HCTs in the RNAi knock-down mutant lines. pHANNIBAL vector represents plants transformed with pART27 with the pHANNIBAL cassette without sense or anti-sense sequences. The data shown represent transcript abundance normalized to the references genes. (n=4; letter grouping represents Tukey's range test with the same letters representing groups

which members are not significantly different from each other at $p < 0.01$; asterisk represents a significant difference compared to the pHANNIBAL vector determined by t-test at $p<0.01$).

3.3.2. **HCTA Down Regulation Corresponds with Gross Morphological Changes**

Two of the *Populus* INRA 353-38 HCTA knock-down plants (Ptthcta-1, and Ptthcta-2) showed an impaired-growth phenotype that was apparent already early in the tissue culture procedure, accompanied by a red colouration of the stem (Figure 3-3), which may indicate a general stress response or anthocyanin accumulation as seen in other HCT knock-downs (Besseau *et al.*, 2007). The five transgenic lines were transferred to soil and grown in a greenhouse. Ptthcta-2, which had the lowest expression of PttHCTA1 and PttHCTA2, also had the lowest growth rate (Figure 3-4). Ptthcta-1 showed less severe down-regulation of PttHCTA1 and PttHCTA2, but still showed a strong growth phenotype. Ptthcta-9, and Ptthcta-8 were statistically different from empty vector in terms of PttHCTA2 expression, but did not show a reduced-growth phenotype. Ptthcta-4 did not show significant down-regulation of PttHCTA1 or PttHCTA2 and also did not have any apparent reduced-growth phenotype. There appears to be a correlation between the average degree of down-regulation and growth rate (A1: r^2=0.84; A2 r^2=0.68), but – restraint in interpretation is warranted due to the observed variability of transcript abundance, the variability of the growth rate, and the limited sample size.

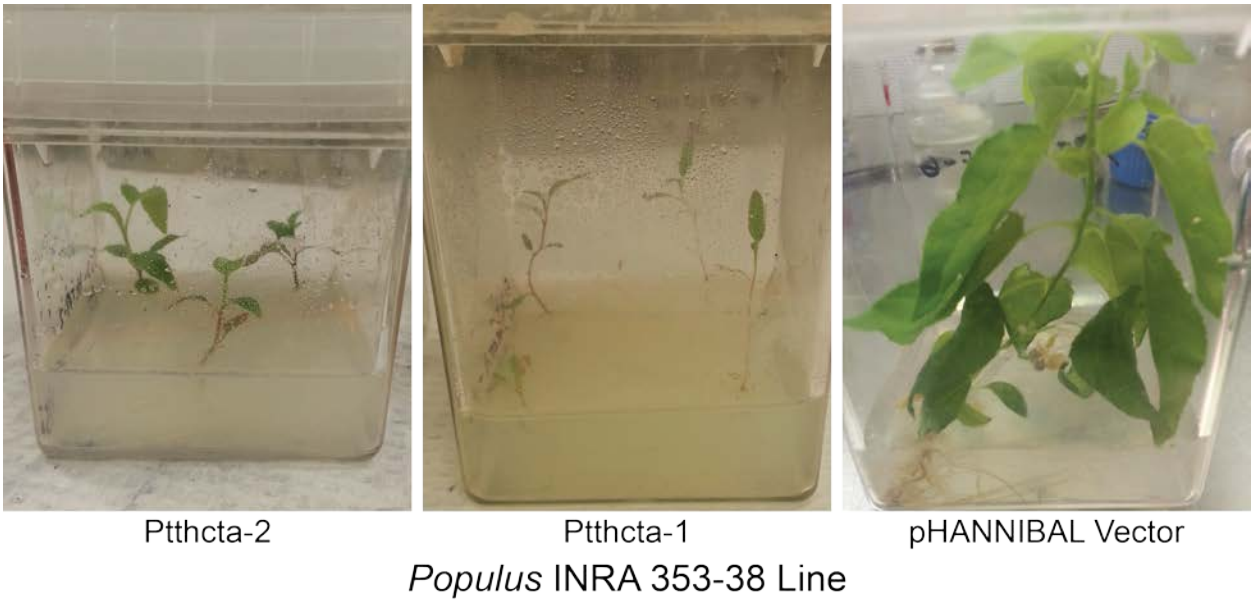


Figure 3-3. Three-Month-Old *Populus* Plants Grown in *in-vitro* Tissue Culture
Plants were grown in M&S media and show the early dwarfing phenotype apparent in the RNAi knock-down lines Ptthcta-2 and Ptthcta-1.

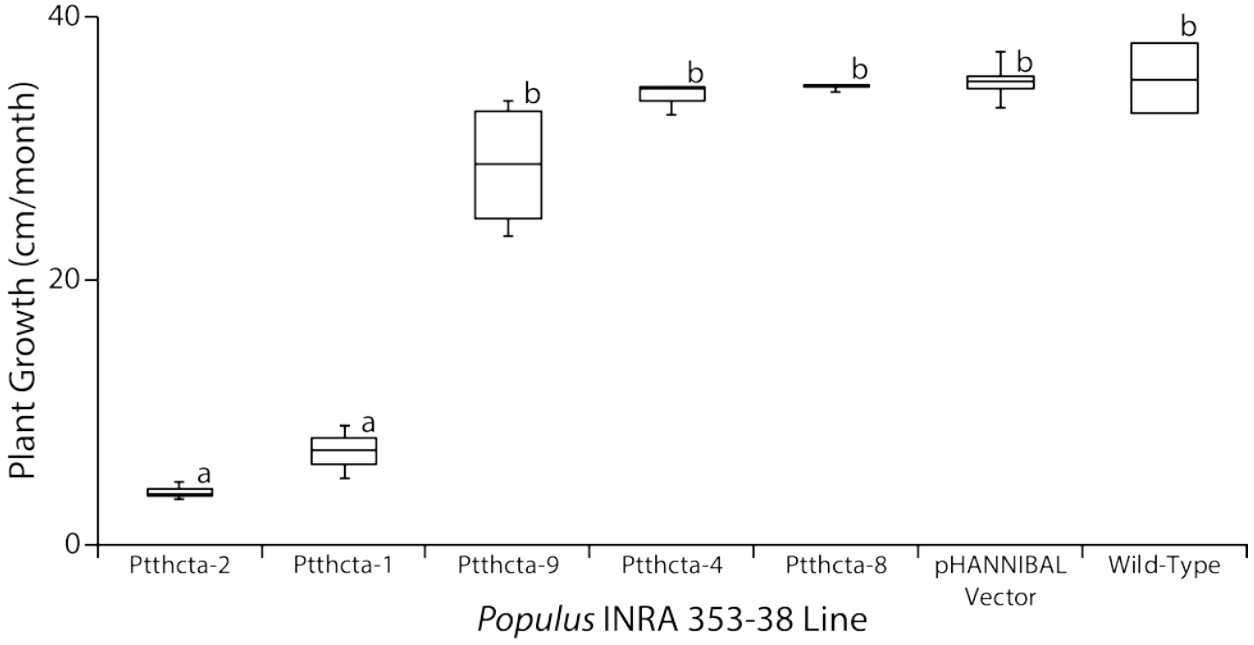


Figure 3-4. Growth Rate of Greenhouse-Grown HCT-A Knock-Down Plants Lines.
In order to compare the relative growth rate as a proximate for potential biomass the length of each *Populus* INRA 353-38 plant was measured from the base of the plant to the apex measured weekly over the course of two months. (Grouping represents Tukey's range test with the same letters representing groups which members are not significantly different from each other, $p < 0.01$).

3.3.3. **HST Knock-Down has a Subtle Effect on Cell Wall Thickness**

Stem cross-sections from Ptthcta plants were examined using transmission electron microscopy (TEM) in order to observe the secondary cell-wall structure. All transgenic lines tested were statistically different from the empty vector control in terms of cell wall thickness, based on a Tukey's range test. Although the difference is statistically significant, it is subtle and the Ptthcta-2 RNAi knock-down plants appeared to have a *decreased* cell wall thickness while the other lines, unexpectedly, appear to have an *increased* cell wall thickness (Figure 3-5).

3.3.4. **HCTA RNAi Lines Produce Less Lignin that is Enriched in H-Units**

The empty vector control line produced wood that was primarily composed of S-lignin, with an S:G ratio of 2.3±0.3, and with barely detectable H-lignin levels (Figure 3-6), consistent with previous reports (Davison *et al.*, 2006). The lignin composition of the HCTA RNAi knock-downs indicated a change in lignin composition. The knock-down mutants that showed a strong growth phenotype also had a large apparent change in the quantity and composition of lignin. S:G ratios of Ptthcta-1 and Ptthcta-2 were 3.7 and 3.4, respectively, which was significantly higher than the wild-type's ratio of 2.3. Importantly, H-lignin made up a significant proportion in Ptthcta lines with an H:G ratio of 1.8 in both Ptthcta-1 and Ptthcta-2, compared to 0.0 in control plants. The total content of thioacidolysis-released monolignols, estimated based on the total peak area, also decreased by up to 45% in both lines compared to wild type. Ptthcta-9 had an intermediate growth phenotype, and also contained an intermediate amount of total released monolignols, including significant amounts of H-lignin (Figure 3-6). Lignin analyses were performed by Faride Vide and Shawn Mansfield (UBC).

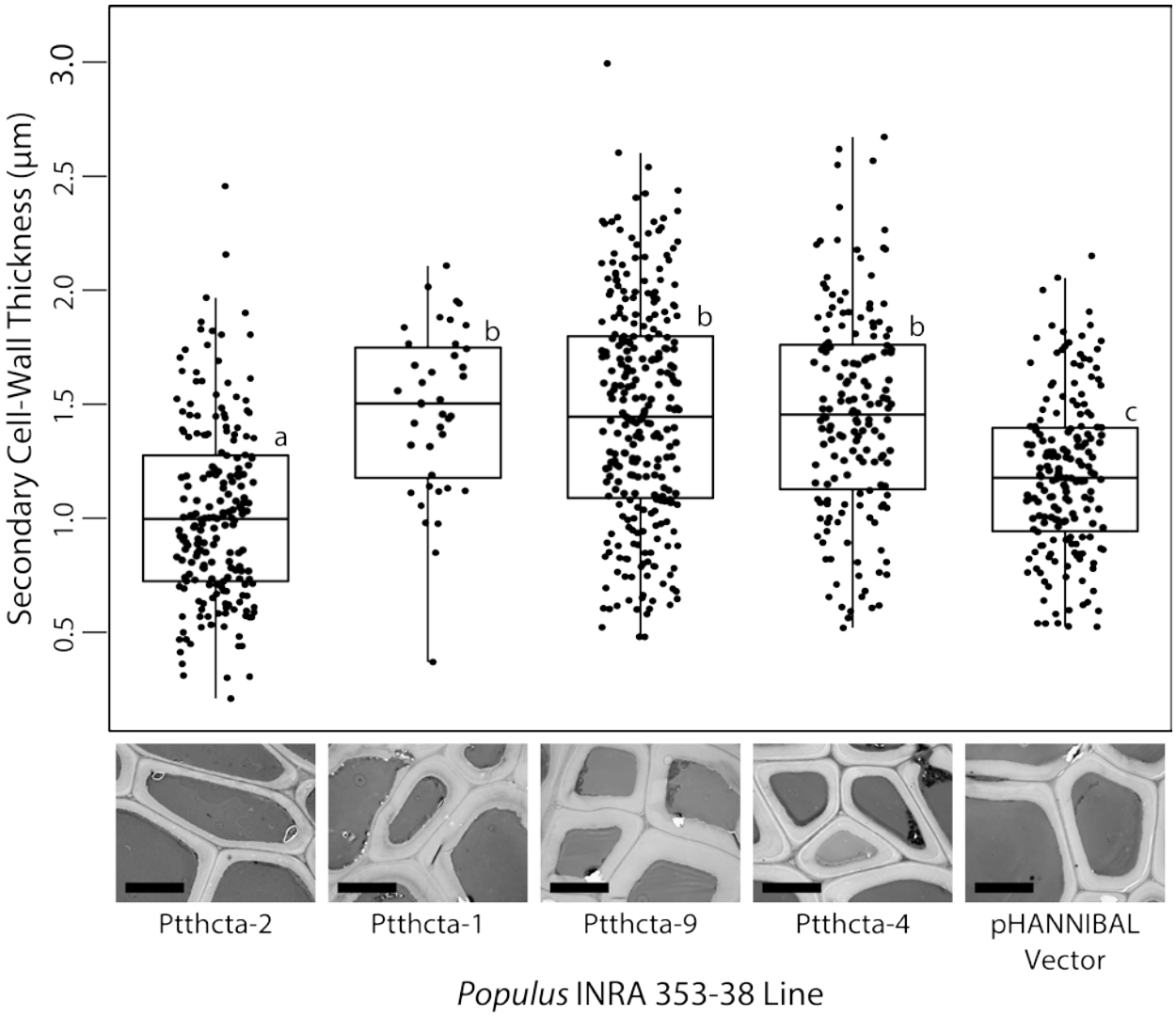


Figure 3-5. Cell Wall Thickness in PtHCTA RNAi Knock-Down Lines
TEM was performed on wood samples in order to determine the impact of lignin modification on cell wall thickness of secondary xylem. The thicknesses of all of the cell walls were measured for each image in ImageJ using the measurement tool using the measurement tool. TEM images from each of the lines are provided for reference with 5µm black scale bars shown. (four biological replicates, six samples per replicate, > 50 cells; grouping represents Tukey's range test with the same letters representing groups which members are not significantly different from each other, $p < 0.01$)

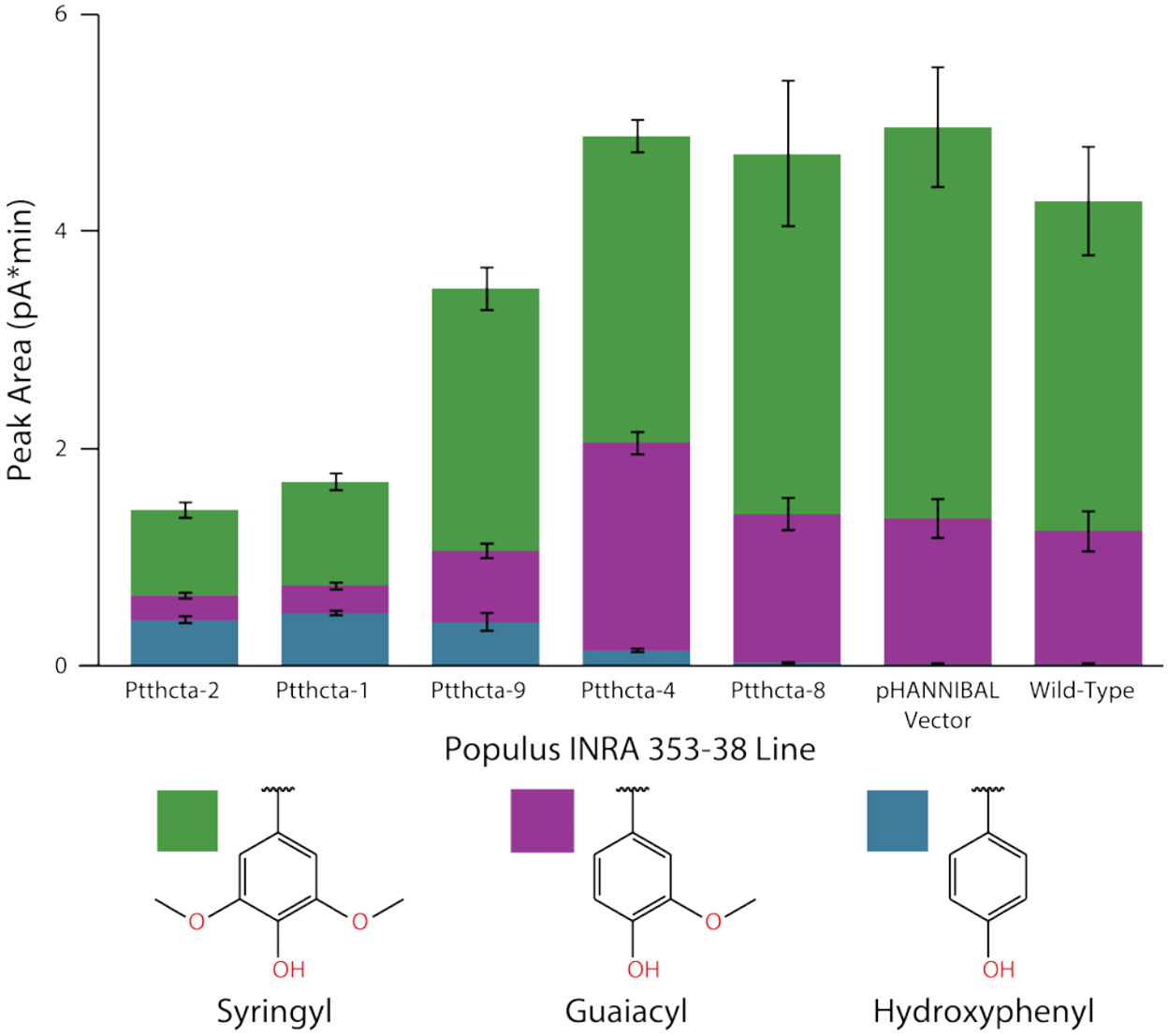


Figure 3-6. H-lignin, S-lignin, and G-lignin Content of *Populus* INRA 353-38 Transgenics Showing Changes in Lignin Composition Upon Knock-Down of *PttHCTA*

In order to determine the impact of HCTA downregulation on lignin content and composition the concentration of H-lignin, S-lignin, and G-lignin were determined by thioacidolysis GC-MS. Values represent instrument response normalized to an internal standard as previously described (Robinson and Mansfield, 2009). Full percentage values and calculated ratios are available in the supplemental material (Supplementary Table 7) (n=3: Ptthcta-1, Ptthcta-2, Ptthcta-8; n=4, Ptthcta-9, Ptthcta-4, vector, wild-type) Error bars represent S.E.M.

3.4. Discussion

3.4.1. **HCTA is Required for G- and S-Lignin Biosynthesis**

The data presented here indicated that PttHCTA2 and possibly PttHCTA1 is required for the synthesis of S-lignin and G-lignin. These genes are homologs of the PtHCTA gene (the gene that I identified as a candidate for shikimate specific HCT activity in *P. trichocarpa* based on phylogenetic and expression data (Chapter 2)). In the *Populus* INRA 353-38 Ptthcta knock-downs, a severe knock-down (Ptthcta-1 and Ptthcta-2) led to a substantial dwarf phenotype and a strong decrease in G-lignin and S-lignin, and lines with modest reduction in HCTA expression had a modest decrease in G-lignin and S-lignin. Initial attempts to transform *Populus* INRA 717-1B4 with the same cassette targeting PtHCTA, showed substantially inhibited regeneration compared to cassettes targeting other HCT transcript. This resulted in no transgenic plants being recovered. (Supplementary Table 5). This may indicate that the knock-down of the HCTA transcript severely disrupts plant growth, leading to a lethal phenotype in this hybrid. In the INRA 353-38 Ptthcta knock-downs, a strong knock-down (Ptthcta-1 and Ptthcta-2) led to a substantial dwarf phenotype and a strong decrease in G-lignin and S-lignin, and the lines with modest knock down had a modest decrease in G-lignin and S-lignin. Together, this indicates that strong knock-down of the HCTA transcript severely disrupts lignin biosynthesis and plant growth. This is consistent with results obtained in *Arabidopsis thaliana*, *Nicotiana benthamiana*, and *Medicago sativa*, where HCT mutants have decreased lignin, altered lignin composition, and decreased growth rate (Hoffmann *et al.*, 2005; Li *et al.*, 2010; Shadle *et al.*, 2007). For *Populus*, however, direct conversion of 4-coumaric acid to caffeic acid by a cinnamic acid 4-hydroxylase/4-coumaroyl-3-hydroxylase membrane

protein complex has been proposed (Chen *et al.*, 2011).This pathways which had been supported by kinetic data and metabolic flux studies would involve bypassing both the HCT and CSE steps (Lin *et al.*, 2015; Wang *et al.*, 2014). In the model proposed, reduction of HCT to 50% of the wild type would show a reduced quantity of lignin and an increase in S/G ratio; however, further decline in protein abundance below 10% of the wild type would lead to a change in metabolic flux restoring wild-type lignin content (Wang *et al.*, 2014). Although *Populus* apparently lacks CSE activity, the reduction of G- and S-lignin and the accumulation of H-lignin subunit in the PttHCTA knock-down lines indicates that this alternative pathway is not sufficient for G and S-monolignol biosynthesis. Instead, my results are consistent with a pathway as proposed in other species, where HCT is required for G- and S-monolignol biosynthesis via conversion of 4-coumaryl-CoA into caffeoyl-CoA (Vanholme *et al.*, 2010).

A single-nucleotide polymorphism (SNP) found in *Populus nigra* in an HCT (PnHCT1; most similar to PtHCTA1) led to a truncated protein product. This genotype led to a subtle phenotype with no statistically-significant down-regulation of total lignin production, but there was an increase in the S:G ratio (Vanholme *et al.*, 2013a). Vanholme *et al.* (2013a) also identified an alternate HCT (i.e., PnHCT6, which is most similar to PtHCTA2), which was highly expressed in young tissue. Both enzymes in *P. trichocarpa* have been shown to synthesize caffeoyl-shikimate (Wang *et al.*, 2014). Since their knockout only affects one of the two isoforms, this may explain why a strong phenotype was not observed. PnHCT1 is also most closely associated with PtHCTA1 which, based on my co-expression analyses, is not closely correlated with lignin biosynthesis. The RNAi knock-down used here targets both HCTA isoforms, and has led to a strong lignin phenotype only when both HST isoforms were

knocked out indicating that -- although there may be an alternate pathway for the decoupling reaction -- the HCTA paralogs are required for S-lignin and G-lignin biosynthesis also in *Populus*, but act, at least partially, redundantly.

p-Coumaroyl-CoA 3′-hydroxylase (C3′H) is the enzyme that has been proposed to carry out the steps between the two HCT reactions: the transesterification of p-coumaroyl-CoA and shikimate to form p-coumaroyl-shikimate and the transesterification of p-caffeoyl-shikimate and CoA to form p- caffeoyl-CoA (Coleman *et al.*, 2008). RNAi mediated suppression of this enzyme generated an extremely misregulated line which saw a decrease in total lignin content and a significant shift in lignin monomer composition. The change in the composition of lignin was primarily due to an increase in H-lignin and decrease in G-lignin with no change in S-lignin, which is similar to the effect seen in PttHCTA-9. In the severe knock-down lines, PttHCTA-2 and PttHCTA-1, we also see a decrease in S-lignin. This indicates that the p-caffeoyl-CoA substrate is preferentially channelled into S-lignin and S-lignin composition only changes once there is insufficient p-caffeoyl-CoA.

In addition to a decrease in G-lignin and S-lignin in PttHCTA knock-down lines, there is an accompanying accumulation of the H-lignin subunit. This was to be expected if HCTA is involved in lignin biosynthesis, since the H subunit is formed directly from coumaroyl-CoA, through a pathway branching off just upstream to HCT action. H-lignin subunit in my knock-downs indicates that the pathway in which free 4-coumaric acid is hydroxylated is not sufficient for G and S-monolignol biosynthesis. Taken together, these data place the poplar HCTA enzyme at the branch point separating H-lignin from G- and S-lignin biosynthesis

representing the first committed step of G- and S-lignin formation, as it is the first enzyme that is required for G-lignin and S-lignin but is not required for H-lignin synthesis.

3.4.2. **Changes in Lignin Composition Can be Decoupled from Dwarfing**

Plants, from diverse species, with mutations in lignin biosynthetic genes lead to dwarfing or other deleterious growth abnormalities. HCT (Hoffmann *et al.*, 2004; Li *et al.*, 2010), cinnamoyl-CoA reductase (Piquemal *et al.*, 1998), CSE (Ha *et al.*, 2016; Vanholme *et al.*, 2013b), cinnamyl alcohol dehydrogenase (O'Connell *et al.*, 2002), and p-coumaroyl-shikimate 3’-hydroxylase (Franke *et al.*, 2002) . The growth phenotype in an *A. thaliana* HCT knock-out was shown to be dependent on chalcone synthase (CHS) expression implicating flavonoid accumulation as the cause of the dwarf phenotype (Besseau *et al.*, 2007). A double knockout of both p-coumaroyl shikimate 3'-hydroxylase, which is downstream of the lignin HCT, and CHS led to a plant with a dwarf phenotype with flavonoid accumulation (Li *et al.*, 2010) indicating that lignification itself may be required for normal plant growth.

Lignin production is required for the normal growth of a plant. RNAi leads to a post-transcriptional down-regulation or transcripts, which is beneficial for studying genes that are essential for survival because a method which results in a *complete* knock-out (such as CRISPR) would lead to death. Because RNAi only limits transcript levels of its targets, it may be more suitable for the study of essential genes *via* intermediate phenotypes. The physiology and growth of the strong HCTA knock-downs (Ptthcta-1, and Ptthcta-2) have been substantially perturbed and -- due to the many changes in the plant -- it becomes difficult to dissect possible distinct functions of the genes, as both growth and lignin are affected, which

may suggest a causal relationship; however, mutants with an intermediate phenotype have been generated. These show a significant change in lignin amount and composition with no obvious change in growth. This suggests that the change in lignin composition itself may not fully explain the dwarfing. The change in lignin composition may not necessarily result in a decrease in biomass and therefore there is likely some other control mechanism may be perturbed by the knockout. Alternatively, a certain threshold of lignin alteration / reduction must be met to cause the developmental dwarfing phenotype, since the HCTA knock-down lines not showing a developmental phenotype also have a less pronounced lignin phenotype compared to lines showing a dwarf phenotype.

3.4.3. **Beneficial Phenotype for Biofuel Feed Stock.**

The resistance of wood to decay is essential for longevity of trees and is beneficial for its use as lumber as it promotes durability of the product. But the recalcitrance of wood to degradation is a major hindrance during pulping and biofuel production. Chemical pre-treatment and/or hydrolytic enzymes are often required to liberate the saccharides required for microbial fermentation from lignocellulose biomass, as lignin's physiochemical properties inhibit these processes. The chemical and enzymatic treatment of lignocellulose material are two of the most economically burdensome steps in the conversion of wood to a usable biofuel. Because the treatments required for saccharide release are some of the most expensive steps, engineering wood to contain less or more easily degradable lignin has been proposed in order to develop wood into an economically viable biofuel feedstock for example in poplars (Furtado *et al.*, 2014; Vanholme *et al.*, 2010).

The intermediate RNAi lines show that a significant change in lignin composition does not necessarily result in a decrease in biomass. RNAi leads to a post transcriptional down-regulation, which is beneficial for studying genes that are essential for survival or have severe or complex phenotypes in complete loss-of-function mutants. Because RNAi only reduces the expression of its target, but to various levels in independent lines, it makes it suitable to the study of our mutants (Hoffmann *et al.*, 2004; Li *et al.*, 2010). The *Populus* RNAi-mediated knock-downs developed here, which show a change in lignin composition, may be suitable for use as a biofuel feedstock. The quantity and composition of lignin is paramount in the commercial use of wood products, but, as previously indicated, attempts to alter lignin composition by interfering with biosynthetic enzymes have resulted in a dwarf phenotype, similar to the phenotype we observed in some of our mutants (Hoffmann *et al.*, 2004; Li *et al.*, 2010). An increase in S-lignin (Stewart *et al.*, 2009) but without a change in total lignin content, has been shown to decrease lignin polymerization and chain length; may be an important determinant of cell-wall digestibility. Changes in lignin composition, particularly an increase in H-lignin (Lapierre and Rolando, 1988) and also S:G ratios (Studer *et al.*, 2011) has been shown to impact sugar release. Thus, increasing the H-lignin and S-lignin composition, without decreasing the total lignin content -- and more importantly, growth --may ultimately lead to more efficient production of cellulosic ethanol.

Manipulation of other genes in the monolignol pathway have been shown to change lignin composition. Down-regulation of the caffeic acid/5-hydroxyferulic acid O-methyltransferase (COMT) leads to no lignin reduction, but there is an apparent decrease in G content and a new 5-hydroxyguaiacyl lignin monomer accompanied by a change in wood colour (Tsai *et al.*,

1998; Van Doorsselaere *et al.*, 1995). Down-regulation of cinnamoyl-CoA reductase (CCR) led to a reduced lignin content (primarily due to an S-lignin decrease) and orange-brown colouration of the xylem. The CCR knock-down improved wood pulping characteristics, indicating a benefit also to biofuel production, but, as with most lignin transgenics, a severe limited-growth phenotype was seen in 5-year-old plants (Beckers *et al.*, 2016; Leple *et al.*, 2007; Skyba *et al.*, 2013).

In this study, Ptthcta-4 and showed a significant increase in H-lignin content. More importantly, there is a change in the wood properties and there is no observable change in growth rate. These RNAi-mediated knock-downs may thus be suitable for use as an efficient biofuel or paper feedstock. RNAi-mediated knock-downs have already been generated for many plants crops, with a RNA-mediated knock-down in polyphenol oxidase in apples (Holderbaum *et al.*, 2010) having received regulatory approval (Waltz, 2015) paving the way for future RNAi crops. The promising initial phenotype on lignin composition in young greenhouse-grown plants needs to be validated in mature field-grown plants in order to determine if there are longer-term changes in composition. As a biofuel feed crop the growth of these plants in the field for several years will be required in order to evaluate the robustness of this phenotype. Lignin is a complex biopolymer and the extent of changes to the structure of lignin in these mutants has not yet been determined. Although plants with increased H-lignin deposition have increased saccharide release, which in turn is correlated with improved biofuel production (Vargas *et al.*, 2016), the exact properties of the wood composition of the Ptt*HCTA* knock-down mutants have yet to be determined. Wood from

more mature trees would need to be assayed in order to determine if the wood that is produced has properties that make it more amenable to biofuel production.

3.5. Conclusions

In *Populus*, HCT is involved in the biosynthesis of lignin and directly affects the chemical composition of wood. It carries out the first committed step in the synthesis of S-lignin and G-lignin that is not required for the synthesis of H-lignin. A severe decrease in PttHCTA expression leads to dwarfing in *Populus*, which is similar to other species with impaired monolignol biosynthesis. Some mutants generated in my studies have a change in lignin composition *without* a change in growth rate. These intermediate mutants should be assayed for wood properties that are industrially beneficial, including cellulosic biofuel production.

3.6. Contributions

Cuong Le: performed the cloning, mutant generation, plant culturing and growth, qPCR, gross morphological assessment, histological analysis, and data analysis. **Brent Gowen**: acquired the transmission electron microscopy images. **Faride Unda**: acquired the lignin composition data. **Bob Chow, Shawn Mansfield, and Peter Constabel**: provided laboratory materials and equipment. **Carol Parker**: edited the manuscript. **Christoph Borchers, Jürgen Ehlting**: edited the manuscript, provided supervision, funding, and laboratory materials and equipment.

4. Establishment of CRISPR-Mediated Gene Knock-Outs in Hairy Root Cultures of *Populus* Targeting a Hydroxycinnamoyl-Transferase Putatively Involved in Chlorogenic Acid Synthesis

4.1. Introduction

4.1.1. Hydroxycinnamoyl Conjugates

HCCs are an abundant class of phenylpropanoids, which are the major class of phenolic compounds derived from the shikimate pathway. Phenylpropanoids and the HCC subgroup were introduced previously (Chapter 1.1). HCCs are widely distributed in plants and lignin and HCC derived product can comprise a large portion of their biomass. HCCs can, for example, constitute up to 10% of leaf dry mass, and more than 30 different HCCs have been identified in diverse *Populus* species (English *et al.*, 1991, 1992; Greenaway and Whatley, 1990, 1991; Greenaway *et al.*, 1987, 1988; Greenaway *et al.*, 1990; Greenaway *et al.*, 1991a, b; Whatley *et al.*, 1989).

There are four HCC subclasses based on the phenylpropanoid acid moiety -- the class defined by caffeate (3, 4-hydroxylation) is the most prevalent in the plant kingdom, as well as being the most abundant in many plants, with the 4-coumarate, ferulate, and sinapate classes being found much less frequently. A general overview of HCCs in plants was previously provided in Section 1.1. The alcohol moiety of the ester can be composed of i) shikimate pathway derivatives such as quinate or shikimate itself, ii) aromatic alcohols such as benzyl alcohol derivatives, phenylethanol, or monolignols, iii) simple alcohols or enols, including isoprenoid

precursors such as prenyl-alcohol and geraniol, or iv) glycerol derivatives. All species within the genus *Populus* produce quinate esters and shikimate esters of caffeate; other esters may be limited to specific *Populus* species. Little is known about the biosynthesis of this class of natural products in the genus *Populus*. Based on the data analysis described in Chapter 2, however, several potential HCT candidates catalysing their biosynthesis have been identified.

4.1.2. **Chlorogenic Acid**

The most common HCCs are chlorogenic acids, which are esters of caffeate and quinate. Several different isoforms exist, with the most common isoform, 3-caffeoyl-quinate, being commonly referred to as chlorogenic acid. Chlorogenic acid is found in a variety of plants, including coffee, potato, tobacco, apple, and poplars (Clifford, 2000). It has diverse *in-vitro* bioactivities, including antioxidant (Kikuzaki *et al.*, 2002), antibacterial (Huang *et al.*, 2009), antifungal (Huang *et al.*, 2009), and antiviral (Kishimoto *et al.*, 2005) properties. It is an astringent, providing the bitter taste found in coffee and the unpleasant taste of many plant leaves, suggesting that it may act as a feeding deterrent. Its anti-herbivory effect has been demonstrated in willows where it is a defence against leaf beetle (Grace *et al.*, 1998; Ikonen *et al.*, 2001) and in chrysanthemum where it is a defence against thrips (Leiss *et al.*, 2009). In addition to its demonstrated function as an antifeeding compound, CGA is proposed to function in iron chelation, defence against pathogens, and protection from UV radiation or environmental stress (Mondolot *et al.*, 2006), as well as a storage form of hydroxycinnmamoyl-CoAs leading to lignin and suberin biosynthesis (Valinas *et al.*, 2015). This highlights the multi-functionality frequently seen for plant secondary metabolites.

4.1.3. **Chlorogenic Acid Biosynthesis (Clade G-IV-β)**

HCCs (Section 1.1) are derived from phenylpropanoids, a major class of phenolic compounds derived from the shikimate pathway. A key step in the biosynthesis of many HCCs, such as chlorogenic acid, is a transesterification reaction performed by an HCT, which are commonly members of the BAHD family of CoA-dependant acyl-transferases. In the case of chlorogenic acid, an HCT couples the end product of the general phenyl propanoid pathway, coumaryl-CoA to an quinate, a dehydrated derivative of shikimate. The HCT catalysed reaction is commonly the final step in the formation of HCCs, but modifications may occur afterwards, adding to the diversity of HCCs found. Based on phylogenetic reconstructions presented in section 2.3.1, several HCT candidates have been identified. The focus here is on HCTC, which I connected to soluble HCC biosynthesis based on expression profile in sections 2.3.3 and 2.3.4.

4.1.4. **Hairy Roots**

Root systems in general are known to secrete chlorogenic acid, and the exudation of chlorogenic acid has been proposed to aid in iron uptake (Hether *et al.*, 1984). Hairy roots are plants tissues naturally formed after an infection with the soil bacterium, *Agrobacterium rhizogenes*, which manipulates the plant's genome to cause proliferation of fine root from the infection site. Hairy roots are neoplastic tissues that grow quickly and do not require hormones for sustained growth, which have been proposed as a commercial source of secondary metabolite (Giri and Narasu, 2000). Inclusion of a transgene construct in *A. rhizogenes* allows for the comparably fast generation of stable transgenic root tissue cultures in many plants, including *Populus*. These roots generally proliferate from a few or only a single cell, and basic screening using florescence markers can provide a simple marker for

transgenic tissues. Chlorogenic acid has also been shown to be a contributing factor in root hair formation in lettuce (Narukawa *et al*., 2009). Plant hairy roots are also known to be high in soluble phenolic compounds, and a previous report has indicated that soluble phenolics, including chlorogenic acid, are the primary component of hairy roots in several species in *Populus* INRA 717-1B4 (Tang, 2015). Also in other species, such as *Stevia* sp, the chlorogenic acid concentration may reach up to 10% (Fu *et al*., 2015b). Due to the abundance of phenolic compounds, including chlorogenic acid, *Populus* hairy roots were chosen as a model system for investigating the role the enzymes which were determined in Chapter 3 to be candidates for secondary metabolite production (HCTC). Using a targeted gene knock-out approach the involvement of these enzymes in the biosynthesis of chlorogenic acid and other soluble hydroxycinnamic acid conjugates was interrogated.

4.1.5. **Clustered Regularly Interspaced Short Palindromic Repeats**

Chlorogenic acid is a secondary metabolite which is not required for normal growth and function -- because of this, there is little risk of growth or developmental abnormalities in genetic manipulation of these genes, so true knock-outs are desirable in order to maximize any phenotypic change. Clustered regularly interspaced short palindromic repeats (CRISPR)s are sequences that were recently identified as an essential feature of immunity in bacteria and archaea (Marraffini and Sontheimer, 2010); by degrading foreign DNA, they can prevent bacteriophage infection, conjugation, and transformation. DNA fragments from viruses that infect bacteria are incorporated into the bacterial genome to generate an additional CRISP repeat unit. These sequences are then transcribed and guide a specific endonuclease (Cas9) to newly-invading viruses, causing double-strand cleavage of the viral DNA or in our case

plant genomic DNA (Ali *et al.*, 2015). Expressing an artificial guide that targets a specific gene of interest, together with Cas9 in eukaryotic cell, will cause double-strand breaks at the target location. Subsequent non-homologous end joining (NHEJ) leads to a small insertion or deletion in the target sequence, which may result in a frame shift mutation (Puchta, 2005). Making use of the Cas9 CRISPR system thus allows for the knockout of target genes. In this part of my research, I utilised the Cas9 CRISPR system to specifically knockout genes in *Populus* hairy roots which I had previously identified (Chapter 2) as candidate genes involved in chlorogenic acid biosynthesis.

4.2. Materials and Methods

4.2.1. **Cloning of CRISPR Constructs**

Potential targets for all annotated genes in *Populus* INRA 717-1B4 had been previously designed (Xue *et al.*, 2015). Targets close to the 5' end of the coding sequence were selected (Table 4-1). MtU6 and scaffold sequences were cloned from a pUC gRNA shuttle vector containing the MtU6 promoter sequence and the CRISPR scaffold, and CRISPR constructs were prepared as previously described (Jacobs *et al.*, 2015), with minor modification as described below.

Table 4-1. Previously Identified CRISPR gDNA Targets for Potri.018G104800 (Xue *et al.*, 2015) Which Were Selected for CRISPR Knock-Outs

Target Name	CRISPR sequence	Position on CDS
PtaHCTC1-A	GAGTCATTTTATCCTGTAGC	271-293
PtaHCTC1-B	GCTGAAACAGGTTCTGCTAT	364-386

Targets were amplified using Q5 HF DNA polymerase according to the manufacturer's protocol (New England Biolabs); primer sequences are given in Table 4-2. MtU6_R and 35SSpeI-MtU6F primers were used to amplify the MtU6 promoter with a region overlapping the p201G vector (Supplementary Figure 17), provided by C. Peter Constabel, University of Victoria. The p201G vector contains Cas9 (the key protein in the CRISPR process) as well as GFP which is used as a plant selection marker. Fragments which contain the genomic DNA target (gPtaHCTC1-A or gPtaHCTC1-B) were used with ApaI_Scaffold_R to generate a fragment (MtU6 overlap - CRISPR gDNA target - Scaffold - p201G vector overlap). p201G was sequentially digested with SpeI and ApaI, and purified using a QIAquick PCR Purification Kit (Qiagen). The resulting fragments were combined using Hot Fusion assembly (Fu *et al.*, 2015a): the PCR products and the purified linearized plasmid were incubated with an isothermal reaction mixture containing Q5 HF DNA polymerase and T5 exonuclease at 50°C for 1 hour and then gradually cooled. An arbitrary amount of this mixture, which contains a 'nicked' plasmid, was transformed into *E. coli* XL1-blue *via* electroporation and colonies were screened for insertion using identified *via* colony PCR with 35SSpeI_MtU6_F and ApaI_Scaffold_R. Purified plasmid was sequenced in order to confirm correct assembly. Constructs were transformed *via* electroporation into *Agrobacterium rhizogenes* ARqua1, and positive transformants were identified *via* colony PCR with 35SSpeI_MtU6_F and ApaI_Scaffold_R.

Table 4-2. Primers Used to Create Fragments for Hot Fusion Assembly

Primer Name	Use	Sequence
gPtaHCTC-A	MtU6 – **PtaHCTC-A**- e	TCAAGCGAACCAGTAGGCTT**GAGTCATTTTATCCTGTAG CT**GGGTTTTAGAGCTAGAAATAGC
gPtaHCTC-B	MtU6 – **PtaHCTC**-B- Scaffold	TCAAGCGAACCAGTAGGCTT**GCTGAAACAGGTTCTGCTA TG**GGGTTTTAGAGCTAGAAATAGC
ApaI_Scaffold_R	Scaffold - **P201G**	GTGCTCCACCATGTTGGGCCAAAAAAAGCACCGACTCGGT G
MtU6_R	MtU6R	AAGCCTACTGGTTCGCTTGAAG
35SSpeI_MtU6_F	P201G - **MtU6**	CGTGCTCCACCATGTTGGGAATGCCTATCTTATATGATCAA TGAGG

4.2.2. **Hairy Root Transformation**

Populus INRA 717-1B4 plants were maintained as sterile axenic cultures on Murashige and Skoog (M&S) media [0.25 g/L MES, 0.1 g/L myo-inositol, 30 g/L sucrose, 4.33 g/L M&S salts, M&S Vitamins (1 mg/L nicotinic acid, 1 mg/L pyroxidine HCl, 1 g/L thiaminne HCl, 1 g/L calcium pantothenate), pH 5.8], supplemented with 0.5 m/ml indole butyric acid (IBA) and solidified with 7 g/L agar, in growth chambers, under long-day conditions (16 h of light/8 h of dark, 28°C). Transformations were performed as described previously (Yoshida *et al.*, 2015). Leaves were excised from the plants and placed on M&S media solidified with 3 g/L agar and 1 g/L gellan gum for 16 h in the dark at 28°C. Transformed colonies were grown on mannitol-glutamic acid:Luria-Bertani (MGL) medium [5 g/L tryptone, 2.5 g/L yeast extract, 5.2 g/L NaCl, 2.0 g/L glutamic acid, 10 g/L mannitol, 0.2 g/L $MgSO_4*7H_2O$ and 0.5 g/L K_2HPO_4], supplemented with 50 mg/L kanamycin and solidified with 15 g/L agar, at 28°C. Liquid cultures were grown in MGL medium, supplemented with 50 mg/L kanamycin, for 16 h before centrifugation (20 min; 3500 g; RT). The bacterial pellet was resuspended in M&S medium containing 200 mM acetosyringone, to give an optical density of 0.6 to 0.8. Leaves were then

manually wounded with a scalpel blade and inoculated with induction broth containing engineered *A. rhizogenes* and grown on solid woody plant medium (WPM) [2.3 g/L WPM Salts (PhytoTechnology Laboratories), 20 g/L sucrose, 0.1 g/L myo-inositol, M&S Vitamins, at pH 5.5] for 2 days. Plants were transferred to hairy root medium (HRM) [WPM medium, 500 mg/ cefotaxime, and 0.5 mg/L IBA], solidified with 3 g/L agar and 1 g/L gellan gum, until hairy roots formed (subcultured every week for four weeks). Hairy roots arising from transformed leaves were excised and maintained in the dark at 28°C on HRM, solidified with 3 g/L agar and 1 g/L gellan gum (subcultured monthly).

4.2.3. **Transgenic Identification**

Roots were assayed for GFP fluorescence on an SZX7 stereo microscope (Olympus) equipped with a SZX-MGFP filter set (Olympus; excitation: 460-490 nm; dichroic: 505 Emission: 510 long pass) was used for higher resolution images (18.6 times magnification). A 120Q fluorescent illuminator (X-Cite) was used as a light source and a Retiga 2000R Scientific CCD Camera (QImaging) was used to capture digital images., equipped with, and green roots were isolated and transferred to fresh sterile media. An SZX16 stereo microscopes (Olympus) equipped with a SZX2-FGFPA filter set (Olympus; excitation: 460-490 nm, emission: 510-550 nm) was used for higher resolution images (18.6 times magnification). A 120Q fluorescent illuminator (X-Cite) was used as a light source and a Retiga 2000R Scientific CCD Camera (QImaging) was used to capture digital images. Roots were screened by fluorescence for three subculturing periods. Genomic DNA was extracted from hairy roots using established plant DNA extraction protocol (Allen *et al.*, 2006) with modifications. Samples were ground in liquid nitrogen, incubated with extraction buffer (30 min, 65°C), centrifuged (12,000 g; 10 min; RT), and extracted twice with

phenol:chloroform:isoamyl alcohol (25:24:1; v/v). The aqueous phase was mixed with an equal volume of isopropanol, incubated (10 min; RT), and centrifuged (12,000 g; 10 min; RT). The pellet was resuspended in Tris-EDTA buffer with RNAse A (10 mM Tris, 1 mM EDTA, and 1g/L RNAse A, at pH 8.0), and incubated for 30 min at 37°C. Then, 25 µL of 3 M sodium acetate, and 600 µL of ethanol were added. The solution was then mixed, incubated (20 min at -20°C), and centrifuged (12,000 g for 10 min at RT). The supernatant was discarded, the samples were allowed to air dry, and then the samples were resuspended in nuclease-free water (Invitrogen). Primers flanking the CRISPR site were designed (PttHCTC_for: CTCAAGGAGGCACTGAGCAA; PttHCTC_rev: TTCGCATCCCTTGCCAATCT) to generate a 71 bp product. Reactions were run with Q5 HF DNA polymerase as per manufactures instruction (New England Biolabs); reactions were 20 µL final volume and contained 0.5 µM PttHCTC_for, 0.5 µM PttHCTC_rev, 200 µM dNTPs, 0.02 U Q5, 10 ng genomic DNA,PCR reaction protocol was 98 °C for 30 seconds, then 45 cycles of 98 °C for 10 s and 60 °C for 30 s, followed by 73 °C for 30 s, and a final extension at 72 °C for 120 s. Post amplification, EvaGreen was added to each reaction tube and fluorescence was measured in 0.1°C increments from 95-75°C for 10s on a BioRad CFX96 instrument controlled by CFX manager V3.1. Data were analysed using a custom Python script: data were normalised to the maximum value, and the average of the empty-vector controls were subtracted from all samples. K-means clustering using scikit-learn 0.18.1 in Python was then used to divide the samples into two groups, one representing wild-type and one representing melt-curves distinct from wild-type. Only samples from the group not containing wild-type amplicons were used for further analysis. As CRISPR may insert a

different indel into each chromosome, heteroduplexes are expected for successfully transformed tissues regardless whether one or both alleles are affected.

4.2.4. **Phytochemical Identification**

Plant material taken from *in-vitro* tissue culture was ground in 100% MeOH and incubated at 60°C for 2 hours. Extracts were dried using a stream of nitrogen, resuspended in 20 µL of MeOH and diluted 1:10 with water. Samples were centrifuged and 20 µL of the supernatant was separated on a Waters Acquity UPLC CSH Phenyl-Hexyl column (2.1 mm i.d. x 150 mm long, packed with 1.7 µm particles, pore size 130 Å) using a linear gradient (A: 0.1% formic acid in water; B: 0.1% formic acid in 1:1 acetonitrile:isopropanol (v:v); 5% B, 0 min; 5% B 5 min; to 100% B; 40 min). A preliminary on-line LC/MS analysis was performed in the positive ionization mode on a Waters Synapt HDMS to determine the retention-times of the peaks of interest. In a subsequent run, the LC was disconnected from the mass spectrometer, and the eluent was collected manually at the appropriate retention times. Pooled fractions were concentrated and infused on an Orbitrap Velos Pro in order to obtain MS/MS/MS spectra. Elemental composition was determined using the Seven Golden Rules software (Kind and Fiehn, 2007).

4.2.5. **LC-MS Quantitation**

5 mg of tissue was extracted in 1000 µL of 100% MeOH and incubated at 60°C for 2 hours. Samples were dried and resuspended in 100 µL of 10% MeOH. Samples were separated on an Eclipse Plus C18 RRHD column (2.1 mm i.d. x 100 mm long, packed with 1.7 µm particles) on a Dionex 3000 ultimate with a linear gradient (A: 0.1% formic acid in water; B: 0.1% formic acid

in 1:1 acetonitrile:isopropanol (v:v); 5% B to 100% B, 40 minutes). ESI- MS data were collected on a Velos Pro (XIC; 353.0878; ± 2ppm) and the previously isolated compound was used as a standard.

4.3. Results and Discussion

4.3.1. **Transgenic Hairy Root Cultures**

Two candidate genes were selected, based on previous phylogenetic inference, as likely to be involved in the formation of soluble hydroxycinnamoyl esters in *Populus* (Potri.018G104700, *PtHCTC1*; Potri.018G104800, *PtHCTC2*). Initially, RNAi knock-downs were generated for *PttHCTC* and *PtaHCTC* based on the sequence of *PtHCTC* (Supplementary Table 4); however, transcriptional down-regulation was not detected in any of the plants in *in vitro* tissue culture and was not present in the lines selected for green-house growth (data not show). Primary metabolic genes are more likely to be conserved than secondary metabolic pathway in plants (Mukherjee *et al.*, 2016). As the constructs for RNAi were initially designed for *P. trichocarpa,* they may be more likely to knock down primary metabolite genes due to greater sequence similarity.

Chlorogenic acid has been reported to be the primary component of *Populus* INRA 717-1B4 hairy roots (Tang, 2015), therefore a hairy-root CRISPR construct was generated to knock out the *PtaHCTC* pair in *Populus* INRA 717-1B4. The high similarity in sequences of the paralogs of *PtaHCTC* provided simplicity in the design of a CRISPR construct, because many possible CRISPR sites targeted both isoforms thereby overcoming the problem of potential redundancy. Based on this information, I initially designed constructs to target both paralogs

of *PttHCTC*, but as previously indicated the early stop codon makes *PttHCTC*2 an unlikely candidate for HCCs biosynthesis.

Although CRISPR has been used in hairy root transformations for many plants including *Glycine max* (Cai *et al.*, 2015; Sun *et al.*, 2015; Tang *et al.*, 2016), *Medicago truncatula* (Michno *et al.*, 2015; Wang *et al.*, 2017), *Solanum lycopersicum* (Ron *et al.*, 2014), *Lotus japonicas* (Wang *et al.*, 2016), *Zea mays* (Liang *et al.*, 2014; Shi *et al.*, 2017)*, Taraxacum kok-saghyz* (Iaffaldano *et al.*, 2016)*,* and *Brassica carinata (Kirchner et al., 2017)*, there are currently no published reports using CRISPR in *Populus* hairy roots; thus a CRISPR construct was designed based on a procedure described for CRISPR hairy root transformation of *G. max*, combined with previously published *Populus* hairy root transformation protocols. The initial transformation yielded 29 calli, many with hairy roots from 30 explants. The high efficiency of hairy root transformation coupled with the high mutation rate of CRISPR has allowed for the genome editing in *T. kok-saghyz* with no selection, obtaining an 88.9% mutation rate in hairy roots and a 80.0% mutation rate in plants regenerated from hairy roots (Iaffaldano *et al.*, 2016). The efficiency of transformation is species dependant and the efficiency of CRISPR mutation is sequence dependant so care must be taken when comparing results. The efficiency of *Populus* hairy root transformations has been shown to vary dramatically with one study showing hairy root regeneration efficiencies of between 0% and 92% depending on the species used (Han *et al.*, 1997).

4.3.2. **Validation of Transformation**

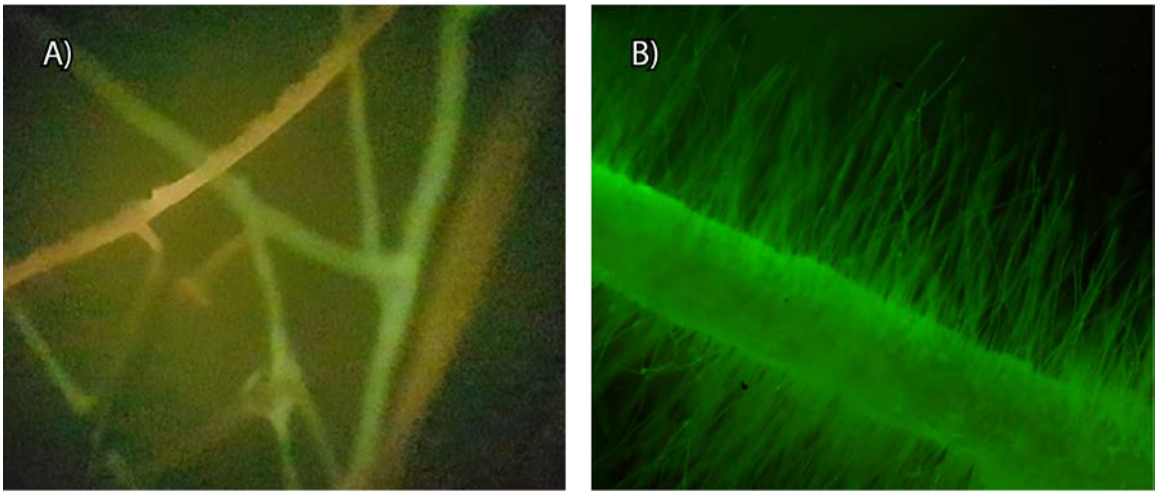


Figure 4-1. Optical Images of PttHCTC CRISPR Knock-Out Lines
Putative transgenic lines were screened using GFP fluorescence, which is included as a selection marker on the p201G plasmid. Hairy roots were visualized using a GFP filter set and the hairy roots that were transformed with the plasmids containing GFP fluoresced green. A) optical image showing putative Ptthctc hairy roots under GFP illumination scope, transgenic hair roots show a clear green colouration while non transformed roots appear orange B) higher magnification (18.6 times) of a single isolated GFP positive root showing the multitude of fine root hairs (image acquired using a black-and-white camera and shown in false colour) under a GFP filter.

After transformation with a hairy-root construct, roots were screened for expression of the GFP selectable marker in the p201G plasmid. Hairy roots are ideally formed from a single cell and thus a single transformation event, but there is also the possibility of forming a chimera, with a mixture of cells forming an off-shooting root. Therefore, three sub-culturing steps, each derived from a single root tip, were performed in order to increase the likelihood of a homogenous population. Twenty-nine individual transformed lines were selected based on the presence of GFP with only one root per step selected to increase the likelihood of independent transformation events.

CRISPR typically generates mutants with small insertions or deletions (indels) at a specific locus. If both alleles have been modified, this is expected to have happened independently,

thus likely generating different indels. Denaturation and annealing of the amplified target region may form a heteroduplex between the alleles when mutated, and subsequent high-resolution melt (HRM)-curve analysis may allow rapid and cost-effective screening of putative knock-out lines (Samarut *et al.*, 2016). Here, primers were designed to flank the CRISPR target site and form a short < 71-bp product. A HRM assay was developed and DNA extracted from roots. Out of a total of 27 lines, 18 showed a clear deviation from empty-vector controls, based on k-means clustering (Figure 4-2). The difference here appears to be quite dramatic in many of these samples. A single deletions or insertions would be expected to produce a subtle change and previous HRM assays showed a small shift in melt curve for indels (Samarut *et al.*, 2016). The initial screen from this assay would indicate a 62% mutagenicity rate similar to the 54% found in *G. max* hairy root transformation with the same constructs (Cai *et al.*, 2015); however, as the PCR products were too small to sequence, sequencing is still required in order to confirm these mutations and whether any of the other samples contained small mutations.

The lines that shows a clear deviation were selected for further growth and propagation. Five lines that were identified by HRM as putative transformants, and which maintained high GFP fluorescence and grew hairy roots rapidly after multiple rounds of sub culturing, were selected for MS analysis.

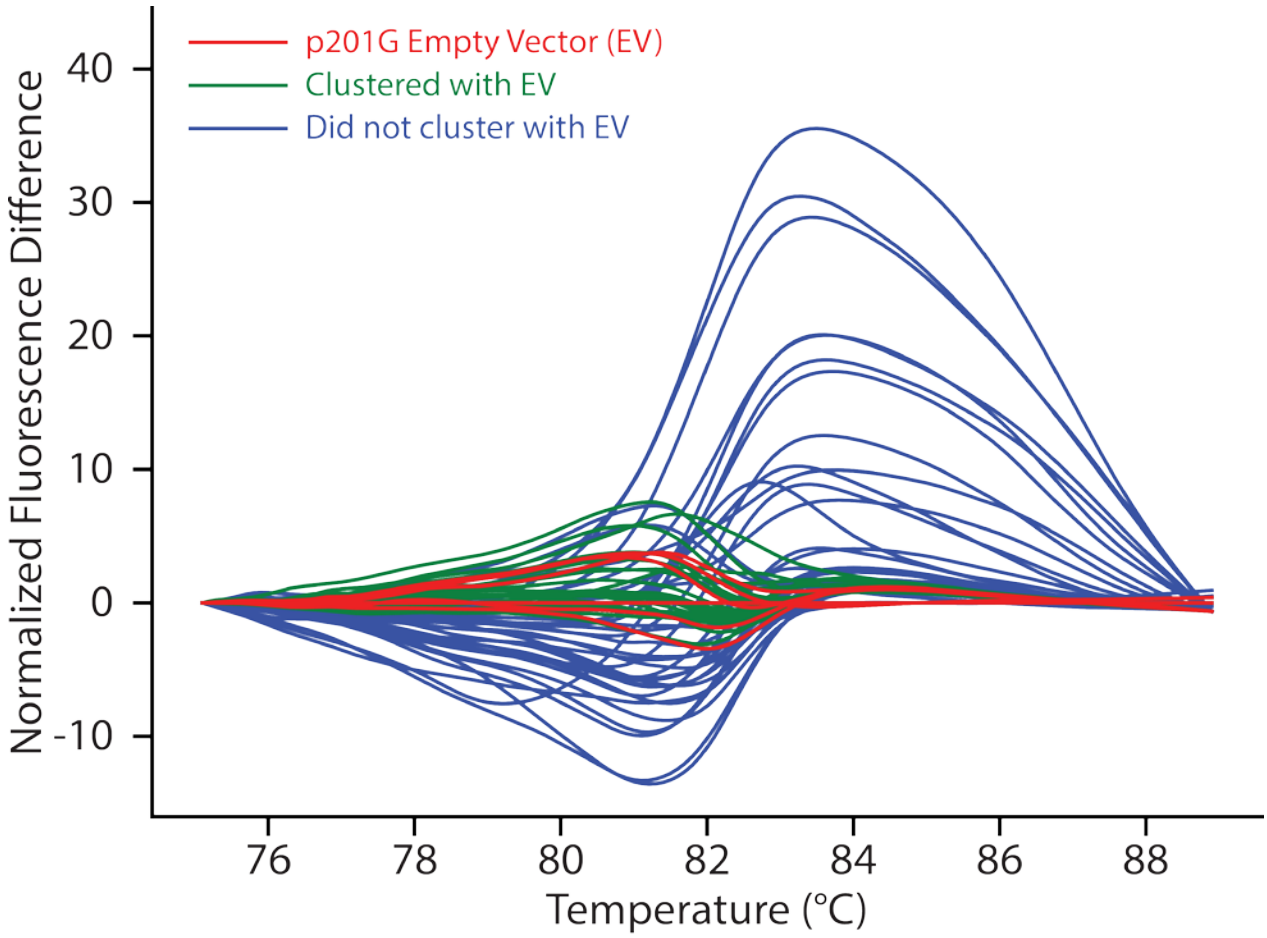


Figure 4-2. Difference Curve From a High-Resolution Melt (HRM) Assay of Putative Transgenic Lines
PCR primers designed to flank the X1 CRISPR site were used to generate small (<100 bp) PCR products. Each sample is coloured based on k-means clustering, and samples which did not cluster with the empty-vector controls were selected for further analysis (n=3).

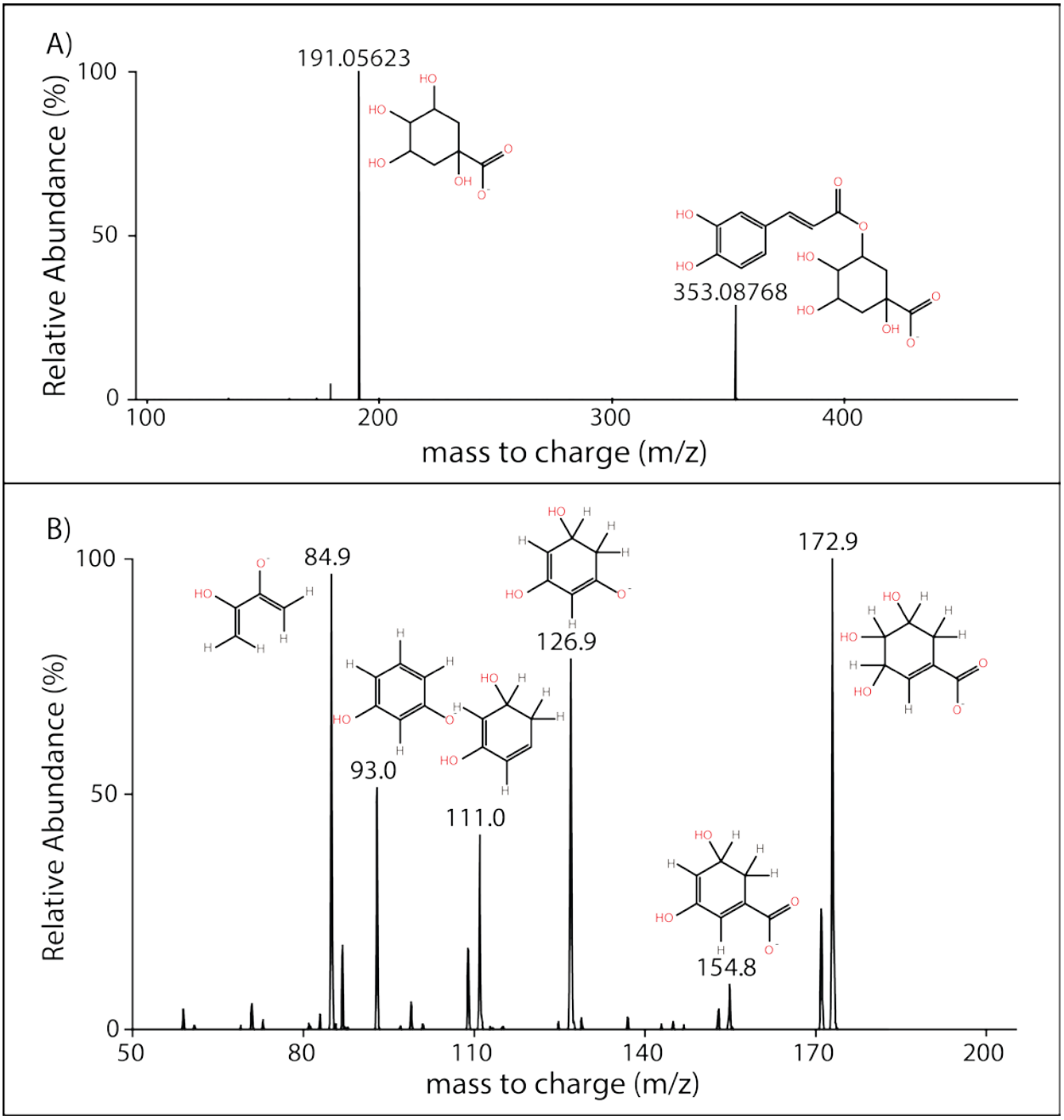


Figure 4-3. Structural Confirmation of a Compound Extracted from *Populus* Leaves as Chlorogenic Acid
Based on MS/MS and MS/MS/MS fragmentation pattern on a Velos Pro mass spectrometer. The compound was purified by HPLC. A) an MS/MS spectrum of a manually purified fraction from *Populus* leaves showing a clear m/z 353.08768 ion for chlorogenic acid, as well an m/z 191.05623 ion, which is characteristic of the quinate ion (Orbitrap; (-)ESI) B) MS/MS/MS of the m/z 191.05623 peak (from panel A) shows fragmentation that is characteristic of quinate (Ion trap; (-)ESI.

4.3.3. **Phenolic Profile**

Contrary to previous reports, in our initial screen of empty-vector *Populus* INRA 717-1B4 hairy roots, chlorogenic acids was not found to be an abundant component. The method's ability to detect chlorogenic acid was confirmed using *Populus* leaf tissue, as this is known to accumulate chlorogenic acid (Tsai *et al.*, 2006). *Populus* leaf extracts were pooled, and a compound with the same mass and retention time as chlorogenic acid was identified. The eluent containing this compound was manually collected and MS^3 confirmed the identity of the compound as chlorogenic acid (Figure 4-3). High resolution MS^2 allowed for the determination of the elemental composition of the ions providing further support for the identification of chlorogenic acid (Table 4-3).

Table 4-3. Theoretical Elemental Composition of Detected Ions
Elemental composition was determined using Seven Golden Rules. Only one chemical formula was possible for each ion. The theoretical mass and the deviation from theoretical are show.

m/z	Chemical Composition	Theoretical mass	Δppm
353.08768	C16H18O9	353.08780	0.34
191.05623	C7H12O6	191.05611	-0.63

Because the amount of material in the hairy root samples was limited, and the absolute abundance of chlorogenic acid was low, only relative quantitation based on an extracted ion chromatogram (XIC; 353.0878 m/z ± 2 ppm; RT: 4.9 ± 0.2 min) could be conducted. Based on my preliminary results analysing the CRISPR *PtaHCTC* knock-outs (*Ptahctc*), PtaHCTC2 appears to have an effect on chlorogenic acid production. In three of the *Ptahctc* transgenic hairy roots lines, there is a clear decrease in chlorogenic acid (Figure 4-4), but this preliminary result

has not been validated in independent biological replicates. The PCR product that was used for HRM analysis was too short to be successfully sequenced, and initial attempts to design larger flanking primers for sequencing have been unsuccessful. Knockout of PtaHCTC2 using the CRISPR construct led to a decrease in chlorogenic acid concentration in hairy root, but confirmation of an indel should be carried out on these lines in order to determine if there is a heterogeneous population of mutant cells in the transgenic cells. The low levels of chlorogenic acid in the roots, and the variability in the tissue indicate that further confirmation is needed. A larger biological sample would provide sufficient tissue for improved quantitation and to obtain a phenolic profile in order to determine the utility of hairy root tissues for phytochemical and functional analysis of secondary metabolite producing enzymes. In addition, a more sensitive detection method, such as a multiple reaction monitoring (MRM) based assay with a stable isotope-labelled internal standard would allow more accurate quantitation.

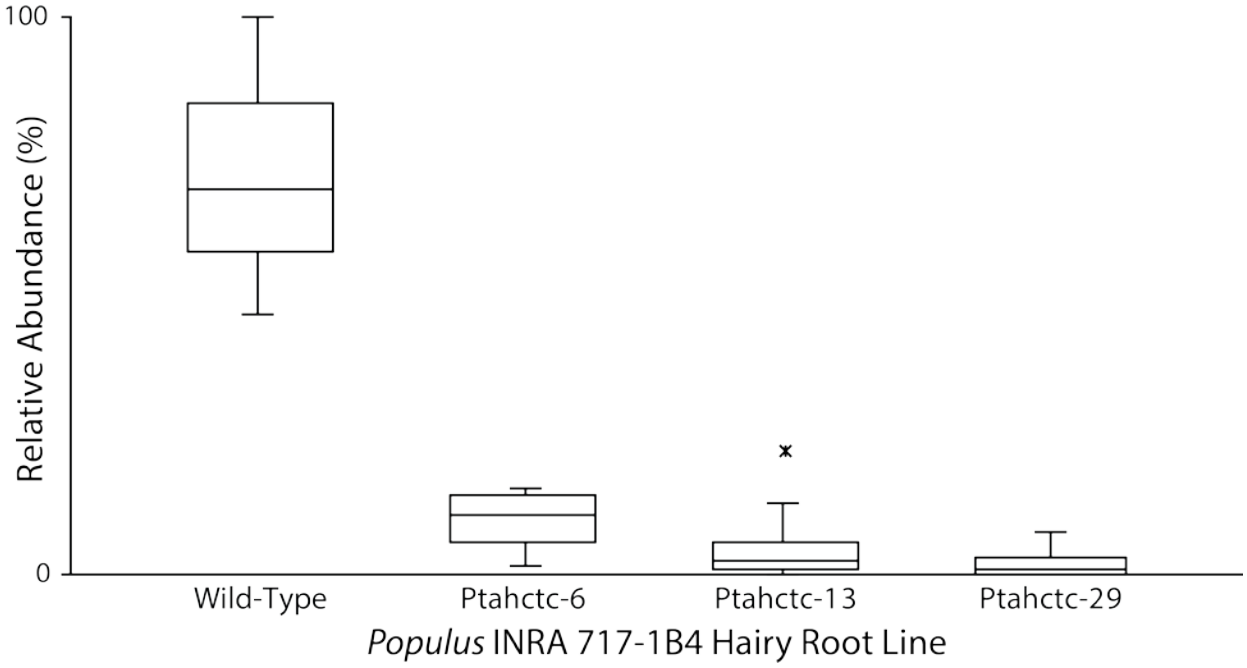


Figure 4-4. Relative Quantitation of Chlorogenic Acid in Hairy Root Lines of *Populus* INRA 717-1B4
Chlorogenic acid concentration in Ptahctc *Populus* INRA 717-1B4 hairy root lines was determined with LC-MS. 95% confidence intervals are indicated with outliers shown as asterisks. Ptahctc knock-outs harbour a CRISPR construct targeting PtaHCTC1 (Potri.018G104800) and PtaHCTC2 (Potri.018G104700). LC-MS quantitation was based on XIC (m/z 353.0878 ± 2 ppm; RT: 4.9 ± 0.2 min; n = 4) and MS data were acquired on a Velos Pro.

4.4. Conclusions

A method for using CRISPR-mediated gene knock-down was established for use in *Populus* INRA 717-1B4 hybrids. -The use of CRISPR technology in hairy roots provides an advantage in studying genes involved in secondary metabolism in plants, because this method generates true loss-of-function mutants without the requirement for whole plant regeneration. In *Populus,* the time frame for generating hairy root mutants is several months to a year, compared to potentially several years for whole plants. Although, with hairy-root transgenics, the biological function and expression of target genes may be an issue, many aspects of biochemistry, physiology, and cell biology can be addressed using this hairy root system. The use of hairy roots also may reduce the genetic complexity of knock-outs, because, in general,

each root originates from a single cell which may lead to a lesser degree of mosaicism. The relative simplicity and speed of hairy root transformation coupled with a possibility of reduced mosaicism, makes whole-plant regeneration system from transgenic hairy roots in *Poplar,* as has been done in some other plants, an attractive possibility. Transformation methods developed for *Populus* hairy roots allowed the successful introduction of a hairy root CRISPR construct with GFP as a transformation marker. GFP screening, coupled with a polymerase chain-reaction high-resolution melt (PCR-HRM) assay, was used for preliminary characterization of knock-outs.

4.5. Contributions

Cuong Le: performed gene cloning for CRISPR knockout constructs, mutant generation, plant culturing and growth, microscopy, PCR experiments, mass spectrometric experiments, and data analysis. **Steve Perlman, Bob, Chow, John Taylor, Peter Constabel**: provided laboratory materials and equipment. **Carol Parker:** provided copyediting. **Christoph Borchers, Jürgen Ehlting:** provided copyediting, supervision, funding, laboratory materials, and equipment.

5. Future Directions

I employed a functional genomics approach in order to expand our understanding of hydroxycinnamoyl-conjugate biosynthesis and function in *Populus*. Starting with a large candidate gene family with sequences for approximately 5300 putative BAHD enzymes, phylogenetic reconstruction provided a broad outline of the evolutionary history of the family and a basis for the inferences of potential functional relationships. This guide reduced the number of candidate genes to five pairs of putative *HCT* genes. Functional hypotheses were subsequently refined for these enzymes based on their expression in various tissues using a mixture of public microarray data, qPCR and RNAseq approaches. The transcript abundance of HCTA in xylem, HCTC in leaves, and HCTE in catkins support the hypotheses that these gene pairs are related to wood formation, secondary metabolite production, and pollen coat formation, respectively. Targeted gene co-expression analyses further supported the hypotheses that HCTA function is related to lignin biosynthesis. These data make the identified genes obvious targets for biochemical and genetic characterization and facilitates the understanding of HCC and secondary metabolism as a complete unified system. While comparative genomics, phylogenetics, and expression profiling are powerful tools for hypothesis generation and reinforcement, targeted biochemical or reverse genetic approaches are necessary for validation.

In order to provide functional support for my hypotheses, I targeted the candidate HCT gene pairs for RNAi mediated posttranscriptional downregulation and established CRISPR mediated gene knock-down or knock-out methodology for *Populus* hairy roots. HCTA- knock-

down plants showed a lignin phenotype which is consistent with the developed hypothesis regarding HCTA function. This establishes a previously disputed role of HCT in lignin biosynthesis and confirms that monolignol biosynthesis in *Populus* follows the same biochemical route in poplar as found in other species. These genes directly impact lignin quantity and quality as the knock-downs had a decrease in lignin content and a decrease in G-lignin and an increase in H-lignin content, confirming that HCT is required for G-, but not H-lignin biosynthesis. Interference with the monolignol biosynthetic pathway frequently leads to severe dwarf phenotypes and it is still a matter of debate if this is a direct consequence of impaired lignin formation, or an indirect effect by interfering with the biosynthesis of or re-directing flow into other phenylpropanoids. Interestingly, some HCTA-knock-down lines showed a change in lignin composition with no change in growth rate. This appears to suggest that lignin and growth phenotypes can be separated, i.e. are independent. Further characterization of these lines will distinguish a threshold of lignin reduction that is necessary to cause a growth phenotype. Increases in S:G ratio and increased H-lignin content have been directly connected to improved saccharification efficiency during bioethanol production from second generation feedstocks, such as poplar wood; however, severe growth reductions in mutants affected in lignin biosynthesis obviously outweigh the benefits of altered lignin structure, which make these plants more amenable to biofuel production. The combination of a desirable lignin change with no observable biomass depression as observed in some HCTA knock-down lines may provide a plant feedstock that is industrially beneficial. In order to determine the suitability of these plants for use as a crop, the properties of this wood and the plant as a whole need further testing. Clonal replicates of these HCTA knock-down lines need

to be grown under field conditions to generate mature trees which have their lignin composition determined. Wood from these trees can subsequently be converted to bioethanol to determine whether the change in lignin improves saccharification for biofuel production.

Providing that framework for a long-term study, RNAi lines of the HCT, which my work proposes to be involved in pollen coat biosynthesis, have been generated. Analysing reproductive tissues is a challenge in poplar as it is in most trees, due to the time scale required for the development of reproductive structures (six to ten in the case of poplar). For this reason, it will be left to other researchers to characterize male flower development and pollen coat structure using the knock-down lines I have generated.

CRISPR mediated targeted genome manipulation is an emerging, powerful tool for reverse genetic approach that is just beginning to be developed in plants. A CRISPR construct targeting *HCTC* was successfully introduced into *Populus* 717-1B4 hairy roots. CRISPR was chosen as a method for interrogating the *in vivo* role of HCTC expressed in leaves in chlorogenic acid production as this editing method provides a knock-out of the gene on both chromosomes simultaneously as opposed to a just a knock-down achievable with other methods such as RNAi. A complete gene knock-out should lead to a more pronounced phenotype on secondary metabolite production, which is not expected to impose a lethal phenotype. The method of genetic manipulation needs to be carefully considered; however, as in some cases moderate impairment of a gene's function may be preferable over a complete loss of function, as may be seen in the case of the lignin related HCTA described

above. Hairy roots putatively knocked out for HCTC function showed a soluble ester phenotype and had a decreased concentration of chlorogenic acid indicating that these genes are necessary for chlorogenic acid formation. Unexpectedly, the HCC concentration was low in the hairy root tissue and only chlorogenic acid could be detected and quantitated. The knock-out appears to have a significant effect on chlorogenic acid concentration without an obvious physiological effect on the hairy root cultures. This suggest that HCTC has a function in soluble HCT production distinct from HCTA which is responsible for G- and S-lignin biosynthesis. Further work is needed to validate that this enzyme is also responsible for chlorogenic acid and other HCCs produced abundantly in other tissues. Although one of the benefits of hairy root culture is that it does not require whole plant regeneration it does not preclude it.

A whole plant regeneration system from hairy roots, as has been established for other plant species, would combine several of the advantages of the hairy root system, such as short time scale and the less genetic mosaicism by root tip sub-culturing, with the advantages of having the complete plant system available for analyses. Alternatively, whole plant transformation employing standard *A. tumefaciens* transformation may be considered to allow development of plant structures known to be rich in soluble phenolics such as bark and young leaves. This may provide a more complete understanding of the natural substrates of these enzymes and would also allow for the characterization of other compounds in connected pathways that may be affected for example via carbon flux alterations to other soluble phenolics or into lignin.

Overall, the work presented shows that the HCT family has diverse roles in phenolic metabolism. I confirmed that HCTA is essential for G- and S-lignin biosynthesis cementing the monolignol pathway is occurring via a shikimate conjugate also in Populus, and that other HCTs have non-complementary functions in soluble HCC biosynthesis. Having established the CRISPR mediated gene knockout approach is setting the stage to study the role of these compounds and their biosynthetic genes in chemical ecology and plant development.

6. Appendix

Supplementary Table 1. qPCR Primers Used for Expression Profiling in *Populus* Tissues

Name	Target	Sequence
EF1β _F	Potri.009G018600	TGGGGCCTCTATTTAGCATGGAT
EF1β _R	Potri.009G018600	CTGCACCCGAAATGGGATTGACC
Ubi_F	Potri.014G115100	ACCAAGCCCAAGAAGATCAAGCA
Ubi_R	Potri.014G115100	CCAGCACCGCACTCAGCA
RP_F	Potri.001G342500	TGTTGTGACCGCTGATTGTTTG
RP_R	Potri.001G342500	CCACCTGTTCTTGCCTGTCTTA
HCTA1_F	Potri.001G042900	TCCATCGACACCTGTTCACATGC
HCTA1_R	Potri.001G042900	CACTCCCAGACCAAGTTCTCTTTG
HCTB_F	Potri.018G105500	TGGTGTTACAGGTGACCCACTTC
HCTB_R	Potri.018G105500	TGATGCAAGCCAACTCCAAGAGAG
HCTC1_F	Potri.018G104800	GCAGGGATAGATTTCGTCCTCCAC
HCTC1_R	Potri.018G104800	AGCAATTGGTGTTGCTAGGAAGG
HCTC2_F	Potri.018G104700	TCTTCCAAGCTACACCCATTGCTC
HCTC2_R	Potri.018G104700	TGTGAATTCTCTCCGCTGTGTGC
HCTD1_F	Potri.005G028100	AGCGAGGGCCACACACATTTAAG
HCTD1_R	Potri.005G028100	ATGTATGCCATGCCCTCGAAGC
HCTD2_F	Potri.005G028000	TCTCTACATAGGGCCGCAAGATG
HCTD2_R	Potri.005G028000	TTCGGCAGGTCGCACAATAG
HCTE2_F	Potri.018G109900	GGTCTGTTTGTAAGGCTCGTGGAC
HCTE2_R	Potri.018G109900	ACACAAACACCTAGAGCAGTTGGC

Supplementary Table 2. Characterised Enzymes with Demonstrated HCT Activity That Are Not Members of the BAHD Superfamily
Although the majority of enzymes with HCT activity belong to the BAHD superfamily of acyl-transferase there are several that do not. The characterized non-BAHD HCTs are provided here for reference. Only enzymes with biochemically characterized function and sequence information are included in this table.

#	Name	Genbank ID	Species	Function	Reference
1	Hydroxycinnamoyl-CoA: serotonin N-(Hydroxycinnamoyl)transferase	AAK15312.1	*Capsicum annuum*	Biosynthesis of amides involved in cell wall fortification	(Back *et al.*, 2001)
2	Tyramine N-feruloyltransferase	Q9SMB8.1	*Nicotiana tabacum*	Biosynthesis of amides involved in cell wall fortification	(Farmer *et al.*, 1999)
3	Tyramine N-feruloyltransferase	P80969.2	*N. tabacum*	Alternate allele of #3	(Farmer *et al.*, 1999; Negrel and Javelle, 1997)
4	Tyramine N-feruloyltransferase	NP_001234022	*Solanum lycopersicum*	Biosynthesis of amides involved in cell wall fortification	(Von Roepenack-Lahaye *et al.*, 2003)
5	N-hydroxycinnamoyl-CoA:tyramine N-hydroxycinnamoyl transferase	NP_001234877	*S. lycopersicum*	Variant of #5	(Von Roepenack-Lahaye *et al.*, 2003)
6	Serine carboxypeptidase-like 19	AED91420.1	*A. thaliana*	Sinapoylcholine biosynthesis	(Shirley and Chapple, 2003; Shirley *et al.*, 2001)
7	Serine carboxypeptidase-like 19	AAQ91191	*B. napus*	Sinapoylcholine biosynthesis	(Milkowski *et al.*, 2004)
8	Sinapoylglucose:malate sinapoyltransferase	AAF78760	*A. thaliana*	Production of sinapoyl malate	(Lehfeldt *et al.*, 2000)
9	Serine carboxypeptidase-like 17	AEE75166	*A. thaliana*	Benzoyl or sinapoyl glucosinolates	(Lee *et al.*, 2012)
10	1-O-sinapoylglucose:choline sinapoyltransferase	AY383718	*B. napus*	Sinapine biosynthesis	(Weier *et al.*, 2008)

Supplementary Table 3. Characterised Enzymes in the BAHD Superfamily Curated in this Study Including HCTs

Identification numbers, species, clade, function, and literature reference are provided as well as the clade designated in previous work. Substrate specificity and sequence are available in the supplemental material. Clade groupings are based on the phylogenetic reconstruction in show in Figure 3-1.

#	Sequence ID	Genbank Protein ID	Clade	D'Auria Clade[1]	Tuominen Clade[2]	Organism	Function	Reference
1	Fgr01800	BAA24430	A-II			*Fusarium graminearum*	Trichothecene acylation	(Garvey et al., 2008; Kimura et al., 1998; McCormick et al., 1999)
2	Fsp01889	AAD19745	A-II			*Fusarium sporotrichioides*	T-2 synthesis: isotrichodermol accumulates when knocked out	(Garvey et al., 2008; McCormick et al., 1999)
3	Ath00437	AEE84918	B-IV	II	II	*A. thaliana*	Elongation of C28 wax	(Haslam et al., 2012; Negruk et al., 1996)
4	Zma06082	CBC84335	B-IV	II	II	*Zea mays*	Elongation of C28 wax	(Tacke et al., 1995)
5	Ath00412	F18A530	B-IV			*A. thaliana*	Elongation of C30 wax	(Haslam et al., 2012; Pascal et al., 2013)
6	Sly05054	AFM77971	C-I			*S. lycopersicum*	Formation of acyl sugars in tomato trichomes	(Schilmiller et al., 2012)

[1] D'Auria, J.C. 2006. Acyltransferases in plants: a good time to be BAHD. Current Opinion in Plant Biology **9**(3): 331-340.

[2] Tuominen, L.K., Johnson, V.E., and Tsai, C.-J. 2011. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. BMC Genomics **12**(1): 236-236.

7	Can00891	AAV66311	C-III	III	IIIa	*C. annuum*	Capsaisin biosynthesis: gene silencing reduces capsaisin biosynthesis	
8	Cbr01244	CBC84333	C-I	III	IIIa	*C. breweri*	Floral scent production (Benzylacetate)	(Dudareva *et al.*, 1998; Nam *et al.*, 1999)
9	Cme01304	AAW51126	C-IV	III	IIIa	*Cucumis melo*	acetate biosyntheis (cinnamoyl acetates)	(El-Sharkawy et al., 2005)
10	Cro00965	Q9ZTK5	C-III	III	IIIa	*Catharanthus roseus*	Deacetylvindoline biosynthesis	(Power et al., 1990)
11	Cro00966	AAO13736	C-III	III	IIIa	*C. roseus*	Minovincinine synthesis: based on crude root extract purifed minovincinine was not tested.	(Laflamme et al., 2001)
12	Nat03199	AET80687	C-I			*Nicotiana attenuata*	Spermidine diacylation: monoacetyl spermidine accumulated in silenced plants	(Onkokesung et al., 2012)
13	Nat03200	AET80686	C-I			*N. attenuata*	Spermidine hydroxycinnamoylation	(Onkokesung et al., 2012)
14	Fan01636	AAG13130	C-IV	III	IIIa	*Fragaria ananassa*	flavour development during ripening (volatile esters)	(Aharoni et al., 2000; Beekwilder et al., 2004)
15	Fve01718	AAN07090	C-IV	III	IIIa	*Fragaria vesca*	flavour development during ripening (volatile esters)	(Beekwilder et al., 2004)

16	Pso03988	Q94FT4	C-II	III	IIIa	*Papaver somniferum*	Morphine biosynthesis	(Grothe et al., 2001; Lenz and Zenk, 1995)
17	Rhy04701	AAW31948	C-IV	III	IIIa	*Rosa hybrid cultivar*	Rose aroma volatile synthesis(geranyl acetate)	(Shalit et al., 2003)
18	Rse04610	CBC84337	C-II	III	IIIa	*Rauvolfia serpentina*	Vinorine synthesis	(Bayer et al., 2004; Gerasimenko et al., 2004; Ma et al., 2004; Pfitzner et al., 1986)
19	Ssp04720	AAR26385	C-I	III	IIIa	*Salvia splendens*	salvianin colouration: anthocyanin malonyltransferase	(Suzuki *et al.*, 2004c)
20	Phy04013	BAF93857	D-III			*Torenia hybrida*	Hydroxycinnamoyl flavonoid synthesis	(Katsumoto et al., 2007)
21	Ath00432	AED97429	D-III			*A. thaliana*	Hydroxycinnamoyl amide synthesis	(Muroi et al., 2009)
22	Ath00433	AEE27636	D-III		Ia	*A. thaliana*	Hydroxycinnamoyl flavonoid synthesis	(Luo et al., 2007)
23	Ath00434	AEE27579	D-III		Ia	*A. thaliana*	Hydroxycinnamoyl flavonoid synthesis	(Luo et al., 2007)
24	Gtr01941	BAA74428	D-III	Ia	Ia	*Gentiana triflora*	Hydroxycinnamoyl flavonoid synthesis	(Fujiwara et al., 1998)
25	Pfr04008	BAA93475	D-III	Ia	Ia	*Perilla frutescens*	Hydroxycinnamoyl flavonoid synthesis	(Yonekura-Sakakibara et al., 2000)
26	Ath00435	AEE77592	D-III		Ia	*A. thaliana*	Hydroxycinnamoyl flavonoid synthesis	(D'Auria et al., 2007a; Luo et al., 2007)

27	Cmo00972	AAQ63615	D-III		Ia	*Chrysanthemum morifolium*	Malonyl flavonoid synthesis	(Suzuki et al., 2004a; Unno et al., 2007)
28	Cmo00973	AAQ63616	D-III		Ia	*Chrysanthemum morifolium*	Malonyl flavonoid synthesis	(Suzuki et al., 2004a; Unno et al., 2007)
29	Cmo00974	AB290338	D-III			*Chrysanthemum morifolium*	Malonyl flavonoid synthesis	(Unno et al., 2007)
30	Dpi01432	Q8GSN8	D-III	Ia	Ia	*Dahlia pinnata*	Malonyl flavonoid synthesis	(Suzuki *et al.*, 2002)
31	Gma01969	ABY59019	D-III		Ia	*Glycine max*	Malonyl flavonoid synthesis (AQY54374, IMaT3; alternate allele characterised)	(Ahmad et al., 2017; Dhaubhadel et al., 2008; Suzuki et al., 2007)
32	Gma02135	XP_003552580	D-III			*Glycine max*	Malonyl flavonoid biosynthesis (AQY54373, IMaT1; alternate allele characterised)	(Ahmad et al., 2017)
33	Lpu02392	AAS77404	D-III		Ia	*Lamium purpureum*	Malonyl flavonoid synthesis	(Suzuki *et al.*, 2004b)
34	Mtr02830	ADV04046	D-III			*M. truncatula*	Malonyl flavonoid synthesis	(Zhao et al., 2011)
35	Mtr02831	XP_003599128	D-III			*M. truncatula*	Malonyl flavonoid synthesis	(Zhao et al., 2011)
36	Mtr02832	XP_003601997	D-III			*M. truncatula*	Malonyl flavonoid synthesis	(Zhao et al., 2011)

37	Mtr02833	ABY91220	D-III		Ia	*M. truncatula*	Malonyl flavonoid synthesis	(Yu et al., 2008)
38	Mtr02834	ABY91222	D-III		Ia	*M. truncatula*	Malonyl flavonoid synthesis	(Yu et al., 2008)
39	Mtr02835	ABY91221	D-III		Ia	*M. truncatula*	Malonyl flavonoid synthesis	(Yu et al., 2008)
40	Nta03205	2XR7_A	D-III		Ia	*Nicotiana tabacum*	Malonyl flavonoid synthesis	(Manjasetty et al., 2012; Taguchi et al., 2005)
41	Ath00443	AED93236	D-II			*A. thaliana*	Triacylglycerol synthesis	(Rani et al., 2010)
42	Pfr04009	AAL50565	D-III		Ia	*Perilla frutescens*	Not chemically characterised (deposited with #47)	(Suzuki et al., 2001)
43	Ath00445	AED94391	D-III			*A. thaliana*	Malonyl flavonoid synthesis	(Taguchi et al., 2010)
44	Ath00446	AEE77599	D-III			*A. thaliana*	Malonyl flavonoid synthesis	(Taguchi et al., 2010)
45	Pcr04007	AAO38058	D-III		Ia	*Pericallis cuneata*	Malonyl flavonoid synthesis	(Suzuki *et al.*, 2002)
46	Ssp04719	Q8W1W9, AAL50566	D-III		Ia	*S. splendens*	Malonyl flavonoid synthesis	(Suzuki et al., 2003; Suzuki et al., 2001)
47	Ssp04721	AAR26386	D-III			*S. splendens*	Malonyl flavonoid synthesis	(Suzuki *et al.*, 2004c)
48	Ghy01952	AAS77403	D-III		Ia	*Verbena hybrida*	Malonyl flavonoid synthesis	(Suzuki *et al.*, 2004b)

49	Phy04015	ABG75942	E-II		IIIb	*Petunia hybrida*	Isoeugenol synthesis (coniferyl alcohol acyltransferase)	(Dexter et al., 2007)
50	Nat03198	AET80688	F-II			*N. attenuata*	Hydroxycinnamoyl amide synthesis (caffeoylputrescine)	(Onkokesung et al., 2012)
51	Hvu02358	A9ZPJ6	F-II	IV	IV	*Hordeum vulgare*	Hydroxycinnamoyl amide synthesis	(Burhenne et al., 2003)
52	Hvu02359	A9ZPJ7	F-II			*H. vulgare*	Hydroxycinnamoyl amide synthesis	(Burhenne et al., 2003)
53	Asa00554	BAC78633	G	V	Vb	*Avena sativa*	Avenanthramide phytoalexin synthesis	(Yang et al., 2004)
54	Ath00439	AED95744	G	V	Vb	*A. thaliana*	Lignin synthesis	(Hoffmann et al., 2005)
55	Ath00441	AEC06845	G		Vb	*A. thaliana*	Hydroxycinnamoyl spermidine synthesis involved in pollen coat formation	(Grienenberger et al., 2009)
56	Cca01420	ABK79690	G		Vb	*Cynara cardunculus var. altilis*	Chlorogenic acid synthesis	(Comino et al., 2009)
57	Cca01274	ABO77957	G			*C. canephora*	Chlorogenic acid synthesis	(Lallemand et al., 2012b)
58	Cca01275	EF137954	G			*C. canephora*	Lignin synthesis	(Lallemand et al., 2012b)
59	Cca01421	AAZ80046	G			*Cynara cardunculus var. scolymus L.*	Lignin synthesis	(Comino et al., 2007)

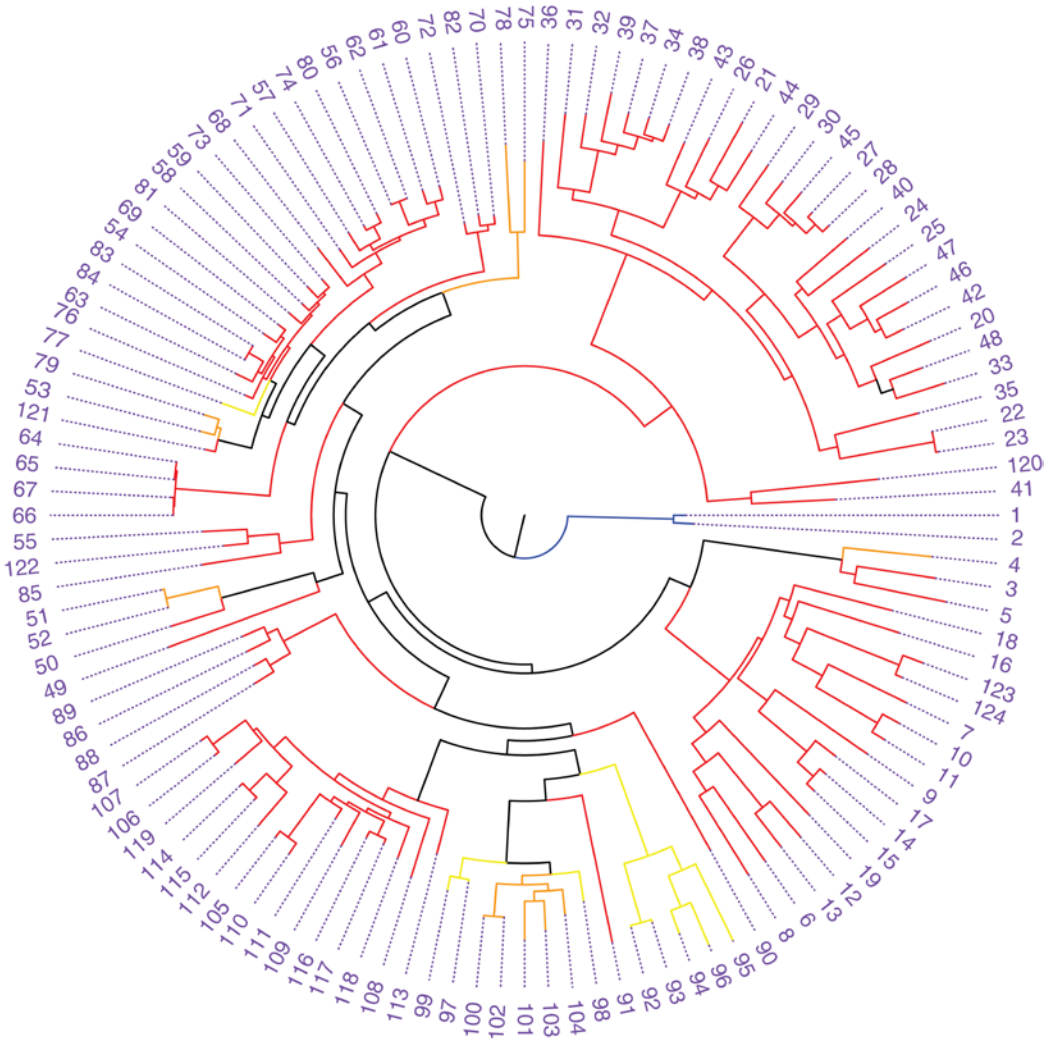
60	Cca01422	CAM84302	G			*Cynara cardunculus var scolymus L.*	Chlorogenic acid synthesis	(Sonnante et al., 2010)
61	Cca01423	ACJ23164	G			*Cynara cardunculus var. scolymus L.*	Variant of #60	(Sonnante et al., 2010)
62	Cca01424	ABK79689	G		Vb	*Cynara cardunculus var altilis DC*	Chlorogenic acid synthesis	(Comino et al., 2009)
63	Csa01311	AEJ88365	G			*Cucumis sativus*	Lignin synthesis	(Varbanova et al., 2011)
64	Dca01436	CAB06430	G	V	Vb	*Dianthus caryophyllus*	Dianthramide synthesis (N-benzoyl anthranilate)	(Yang et al., 1997)
65	Dca01437	O24645	G			*Dianthus caryophyllus*	Variant of #64	(Yang et al., 1997)
66	Dca01438	O23917	G			*Dianthus caryophyllus*	Variant of #64	(Yang et al., 1997)
67	Dca01439	O23918	G			*Dianthus caryophyllus*	Variant of #64	(Yang et al., 1997)
68	Eco01452	AFF19203	G			*Erythroxylum coca*	Chlorogenic acid synthesis	(Pardo Torre et al., 2012)
69	Hca02268	AFN85668	G			*Hibiscus cannabinus*	Lignin synthesis	(Chowdhury et al., 2012)
70	Lan02395	ABI48360	G			*Lavandula angustifolia*	Rosmarinic acid sythensis	(Landmann et al., 2011)
71	Lja02585	ACZ52698	G			*Lonicera japonica*	Chlorogenic acid synthesis	(Peng et al., 2010)
72	Mof02939	CBW35684	G			*Melissa officinalis*	Rosmarinic acid synthesis	(Weitzel and Petersen, 2011)

73	Nta03203	Q8GSM7	G	V	Vb	*Nicotiana tabacum*	Lignin synthesis	(Hoffmann et al., 2003)
74	Nta03204	CAE46932	G	V	Vb	*Nicotiana tabacum*	Chlorogenic acidsynthesis	(Niggeweg et al., 2004)
75	Osa03439	Q5SMM6	G			*O. sativa*	Hydroxycinnamoyl glucose synthesis: physiological function unknown	(Kim et al., 2012)
76	Pni04215	JF693234	G			*Populus nigra*	Lignin synthesis: natural frameshift knock-out)	(Vanholme et al., 2013a)
77	Pra04155	ABO52899	G		Vb	*Pinus radiata*	Lignin synthesis	(Wagner et al., 2007)
78	Pvi03692	AFY17066	G			*Panicum virgatum*	Chlorogenic acid synthesis	(Escamilla-Trevino et al., 2014)
79	Pvi03693	AGM90558	G			*Panicum virgatum*	Chlorogenic acid synthesis	(Escamilla-Trevino et al., 2014)
80	Sly05055	NP_001234850	G		Vb	*S. lycopersicum*	Chlorogenic acid synthesis	(Niggeweg et al., 2004)
81	Ssc05309	CBI83579	G		Vb	*Solenostemon scutellarioides*	Lignin synthesis	(Sander and Petersen, 2011)
82	Psc04194	CAK55166	G			*S. scutellarioides*	Rosmarinic acid synthesis	(Sander and Petersen, 2011)
83	Tpr05723	ACI16630	G		Vb	*Trifolium pratense*	Lignin synthesis	(Sullivan, 2009)
84	Tpr05724	ACI28534	G		Vb	*T. pratense*	Lignin synthesis: variant of #83)	(Sullivan, 2009)
85	Tpr05725	ACI16631	G		Vb	*T. pratense*	Phaselic acid synthesis	(Sullivan, 2009)
86	Ath00436	AED94628	I-II		Va	*A. thaliana*	Suberin synthesis	(Gou et al., 2009)
87	Ath00442	BAD43042	I-II			*A. thaliana*	Cutin synthesis	(Rautengarten et al., 2012)
88	Ath00444	AED97769	I-II			*A. thaliana*	alkyl hydroxycinnamates synthesis of in root waxes	(Kosma et al., 2012)

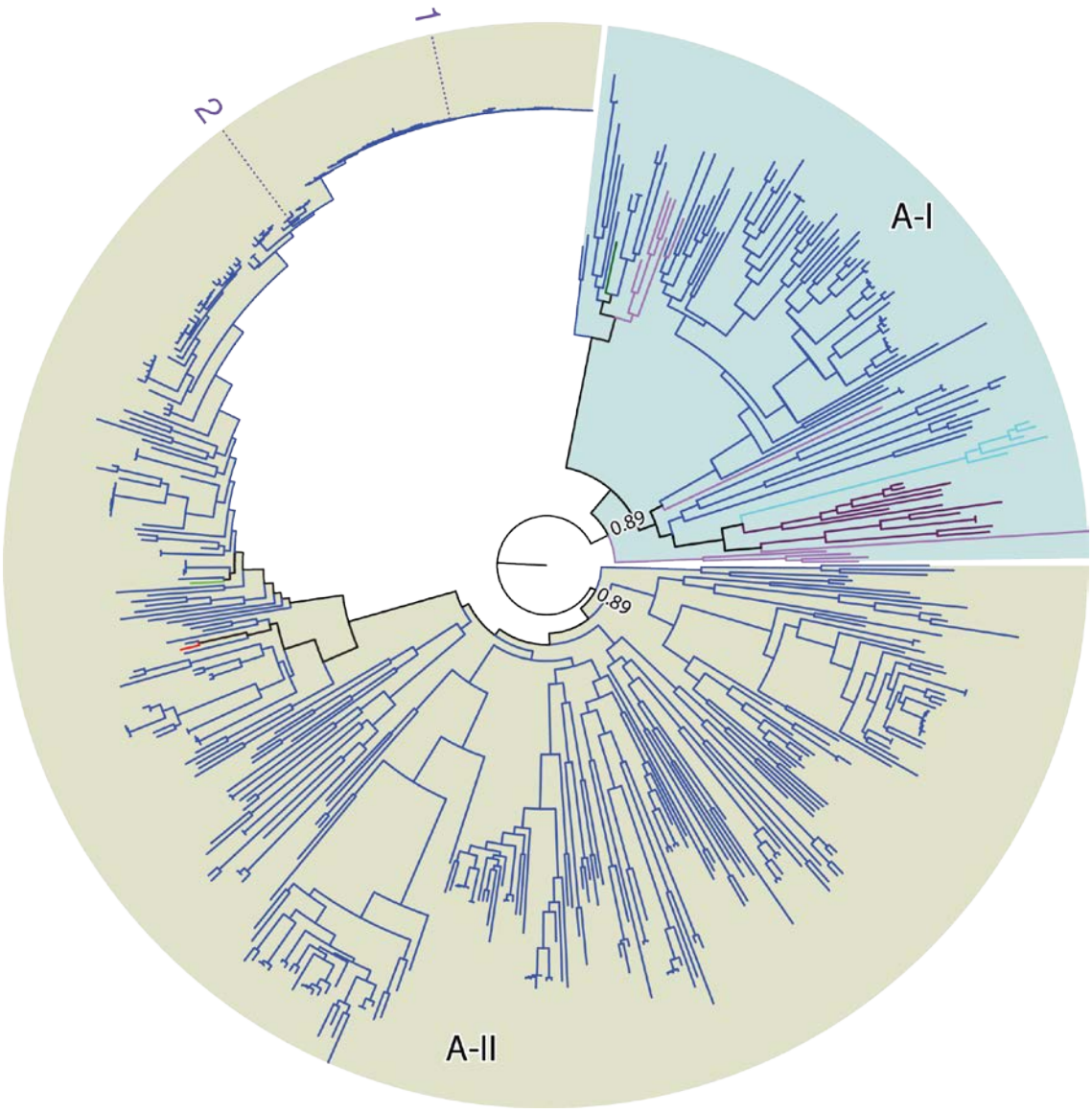
89	Stu05167	ACS70946	I-II			*Solanum tuberosum*	Suberin synthesis	(Serra et al., 2010)
90	Aer00008	X	J-III			*Actinidia chinensis* *Actinidia deliciosa* *Actinidia eriantha*	Volatile ester synthesis (butyl acetate and butyl propionate)	(Günther et al., 2011)
91	Ath00440	AEC07460	K-I		Va	*A. thaliana*	disinapoyl spermidine synthesis	(Luo et al., 2009)
92	Iho02389	CBD10869	K-II		Va	*Iris hollandica*	Hydroxycinnamoyl anthocyanin synthesis	(Yoshihara et al., 2006)
93	Iho02390	CBD10859	K-II			*I. hollandica*	Hydroxycinnamoyl anthocyanin synthesis	(Yoshihara et al., 2006)
94	Msa03170	AX025506	K-II	V	Va	*Musa sapientum*	Flavour development during ripening (alcohol acyltransferase)	(Beekwilder et al., 2004)
95	Osa03438	BAD33123	K-II			*O. sativa Japonica Group*	Glucuronoarabinoxylan hydroxycinnamoylation in grass cell walls: overexpression increased paracoumaric acid and decreased ferulic acid	(Bartley et al., 2013)
96	Osa03440	NP_001042791	K-II			*O. sativa Japonica Group*	Hydroxycinnamoyl hydroxycinamaete syntheis in monocot cell wall	(Withers et al., 2012)
97	Tcu05576	CBD10894	K-I			*Taxus cuspidata*	Paclitaxel synthesis	(Chau et al., 2004)
98	Tca05486	CBD10937	K-I	V	Va	*Taxus canadensis*	Paclitaxel synthesis	(Nevarez et al., 2009)

99	Tcu05577	CBD10897	K-I			*T. cuspidata*	Paclitaxel synthesis	(Chau et al., 2004)
100	Tba05465	CBD10927	K-I			*Taxus baccata*	Paclitaxel synthesis	(Fang and Ewald, 2004)
101	Tcu05578	CBD10923	K-I	V	Va	*T. cuspidata*	Paclitaxel synthesis	(Walker et al., 2002)
102	Tcu05579	CBD10933	K-I	V	Va	*T. cuspidata*	Identical sequence cloned from (Taxus x Media; AAS13684.1) (paclitaxel biosynthesis)	(Guo *et al.*, 2007; Walker and Croteau, 2000b)
103	Tcu05580	Q9FPW3	K-I	V	Va	*T. cuspidata*	Paclitaxel synthesis	(Walker and Croteau, 2000a)
104	Tcu05581	CBD10942	K-I	V	Va	*T. cuspidata*	Paclitaxel synthesis	(Walker et al., 2000)
105	Cme01308	ACA04742	L			*C. melo*	RNAi knock-down has decrease acyl-ester Variant of #110	(Shan et al., 2012)
106	Ach00006	X	L			*Actinidia chinensis*	Volatile ester synthesis (butyl benzoate)	(Günther et al., 2011)
107	Ade00007	X	L			*Actinidia deliciosa*	Volatile ester synthesis (butyl benzoate); Variant of #106	(Günther et al., 2011)
108	Ath00438	AEE73949	L	V	Va	*A. thaliana*	Green leaf volatile synthesis ((Z)-3-hexen-1-yl acetate)	(D'Auria et al., 2007b)
109	Cbr01245	CBD10921	L	V	Va	*C. breweri*	Volatile esters syntheis (benzyl benzoate)	(D'Auria et al., 2002)

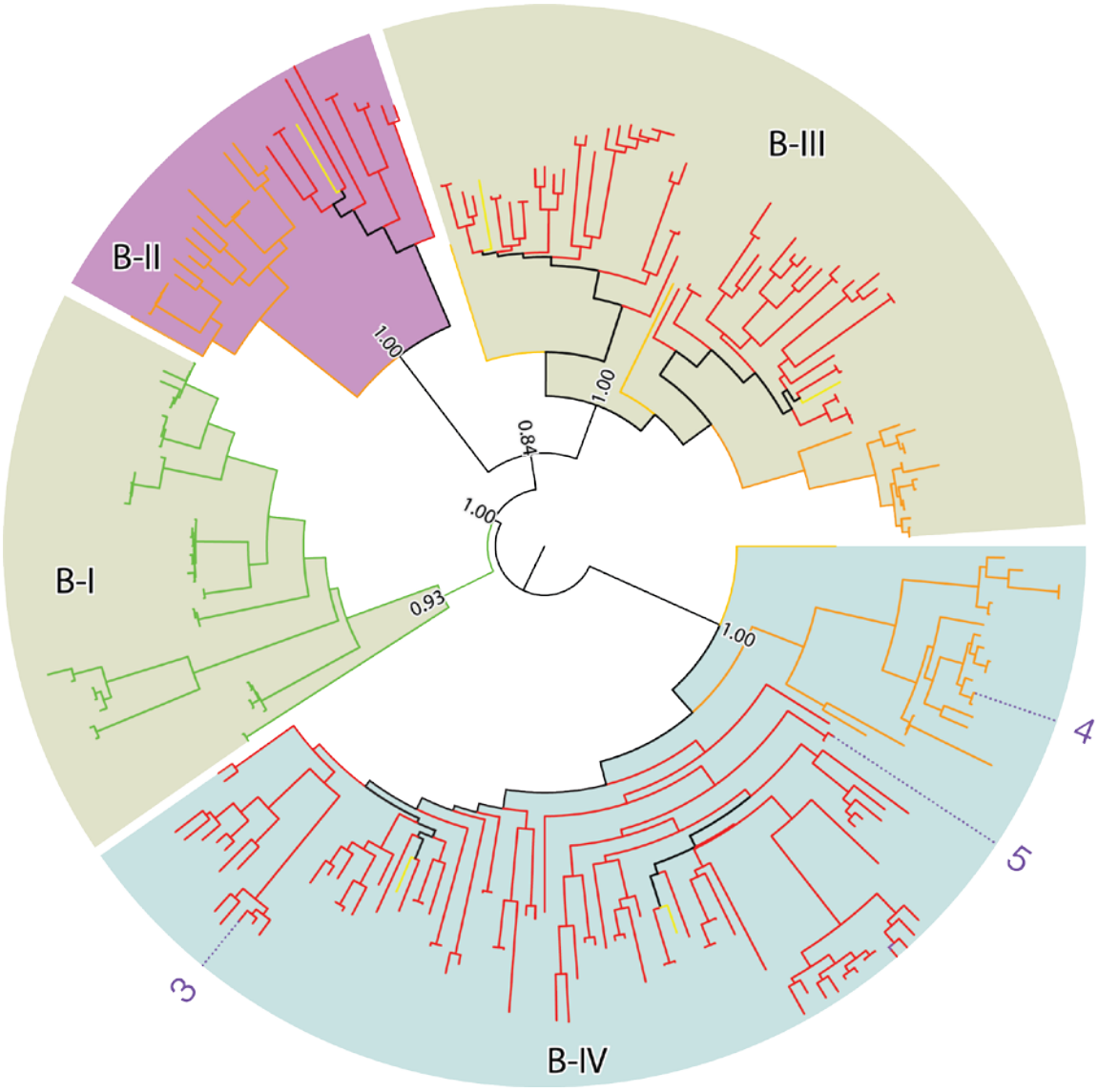
110	Cme01302	CAA94432	L	V	Va	*C. melo*	Acyl ester synthesis (E-2-hexenyl acetate and hexyl hexanoate)	(El-Sharkawy et al., 2005; Yahyaoui et al., 2002)
111	Cme01303	AAW51125	L	V	Va	*C. melo*	Benzyl acetate synthesis	(El-Sharkawy et al., 2005)
112	Fan01635	AEM43830	L			*F. ananassa*	Strawberry volatile ester biogenesis	(Cumplido-Laso et al., 2012)
113	Lal02594	CBD10882	L	V	Va	*Lupinus albus*	Tigloylation of tetracyclic quinolizidine alkaloids biosynthesis	(Suzuki et al., 1994)
114	Mdo02606	CBD10901	L	V	Va	*M. pumila*	Volatile ester biosynthesis (hexyl esters, butyl acetate and 2-methylbutyl acetate)	(Souleyre et al., 2005)
115	Mdo02607	CBD10915	L		Va	*M. pumila*	Volatile ester biosynthesis (Butyl acetate)	(Li et al., 2006)
116	Nta03202	CBD10919	L	V	Va	*N. tabacum*	Volatile esters biosyntheis (benzyl benzoate)	(D'Auria et al., 2002)
117	Phy04014	AAT68601	L	V	Va	*Petunia hybrida*	Volatile esters biosyntheis (benzyl benzoate)	(Boatright et al., 2004; Dexter et al., 2008)
118	Vcu05841	ACT82248	L			*Vasconcellea pubescens*	Volatile esters biosyntheis (benzyl acetate)	(Balbontín et al., 2010)
119	Vla05846	CBD10871	L	V	Va	*Vitis labrusca*	Biosynthesis of grape fruit esters including methyl anthranilate	(Wang and Luca, 2005)



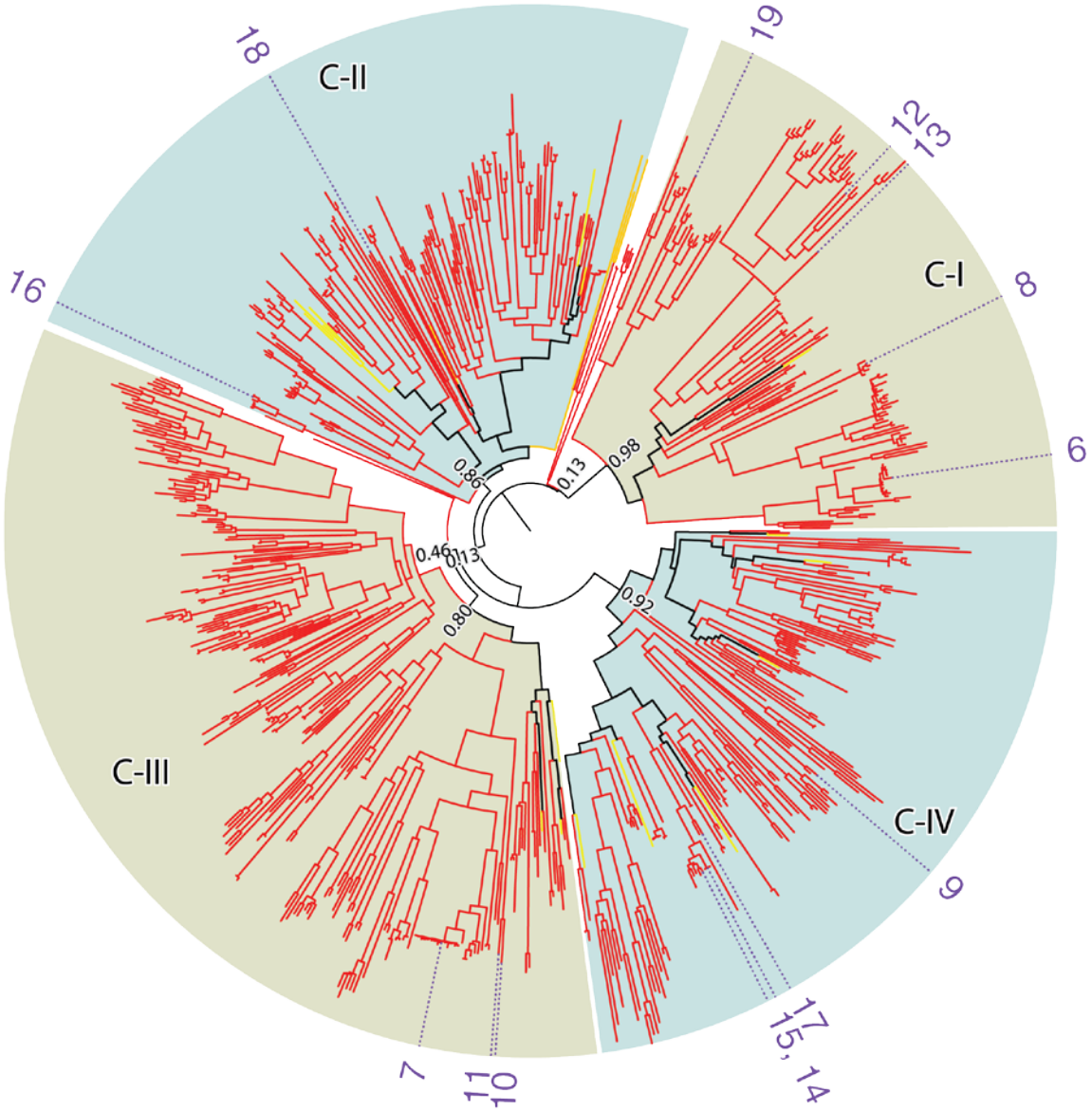
Supplementary Figure 1. Phylogenetic Reconstruction of Characterized BAHD Acyltransferases
Characterized BAHD acyltransferases (Supplementary Table 3) were aligned in order to generate a profile used for the construction of the BAHD acyltransferases (Figure 3-1. Every line represents a unique protein sequence and branch colour coding is as indicated in (Table 2-1).



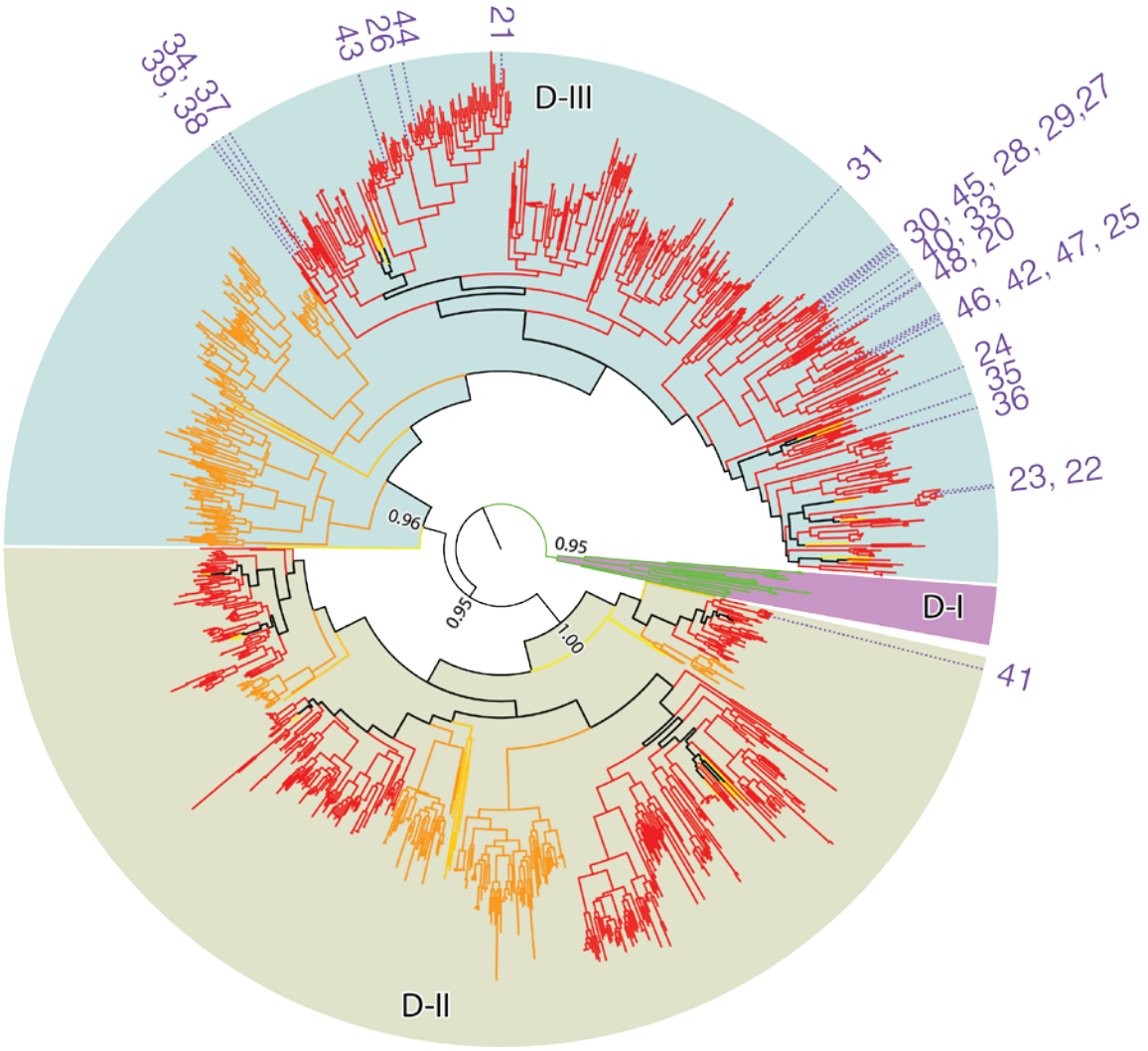
Supplementary Figure 2. Phylogenetic Expansion of BAHD Clade A
This clade is can be broken into sub clades A-I and A-II. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



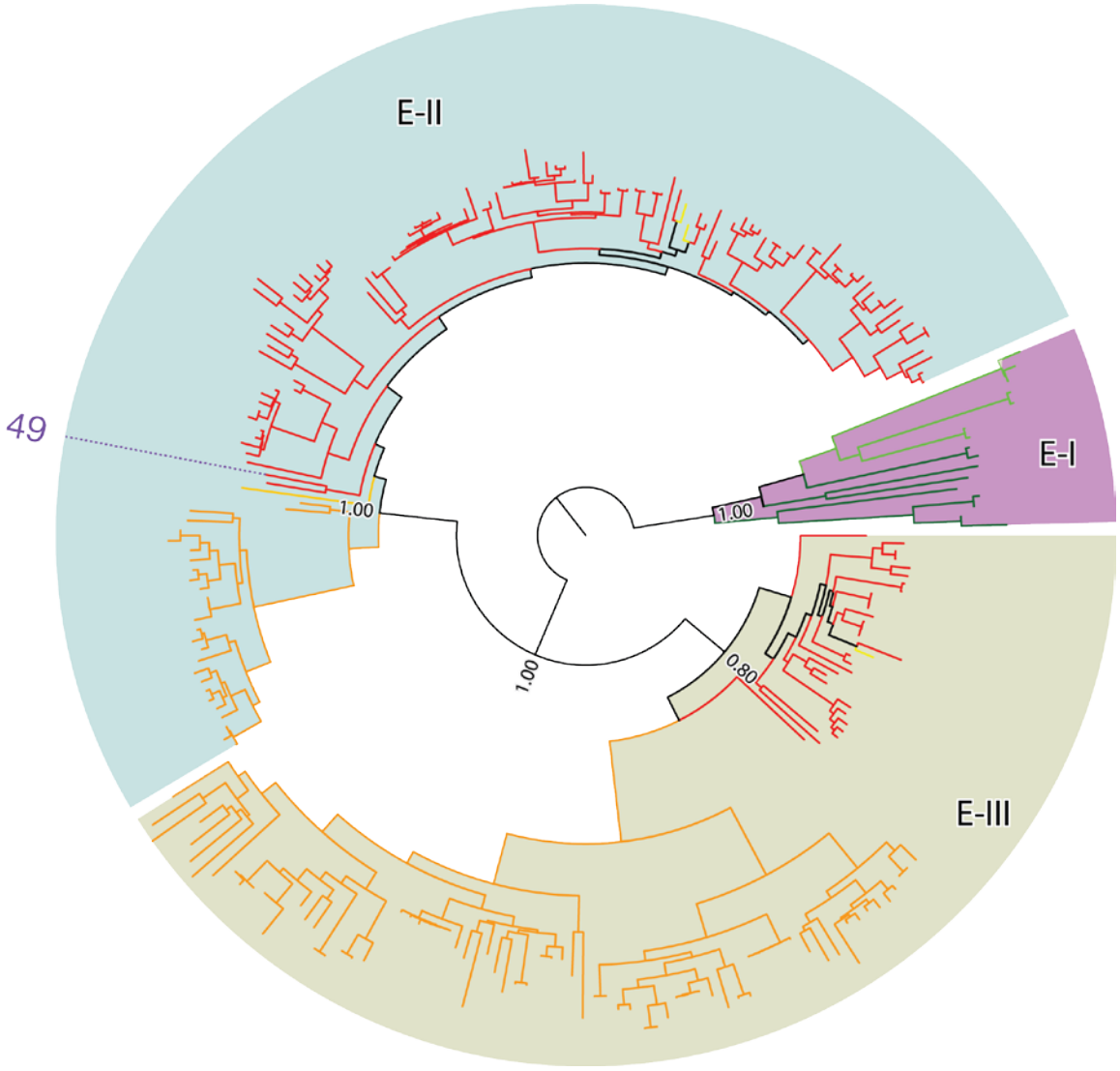
Supplementary Figure 3. Phylogenetic Expansion of BAHD Clade B
This clade is can be broken into sub clades B-I, B-II, B-III, and B-IV. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



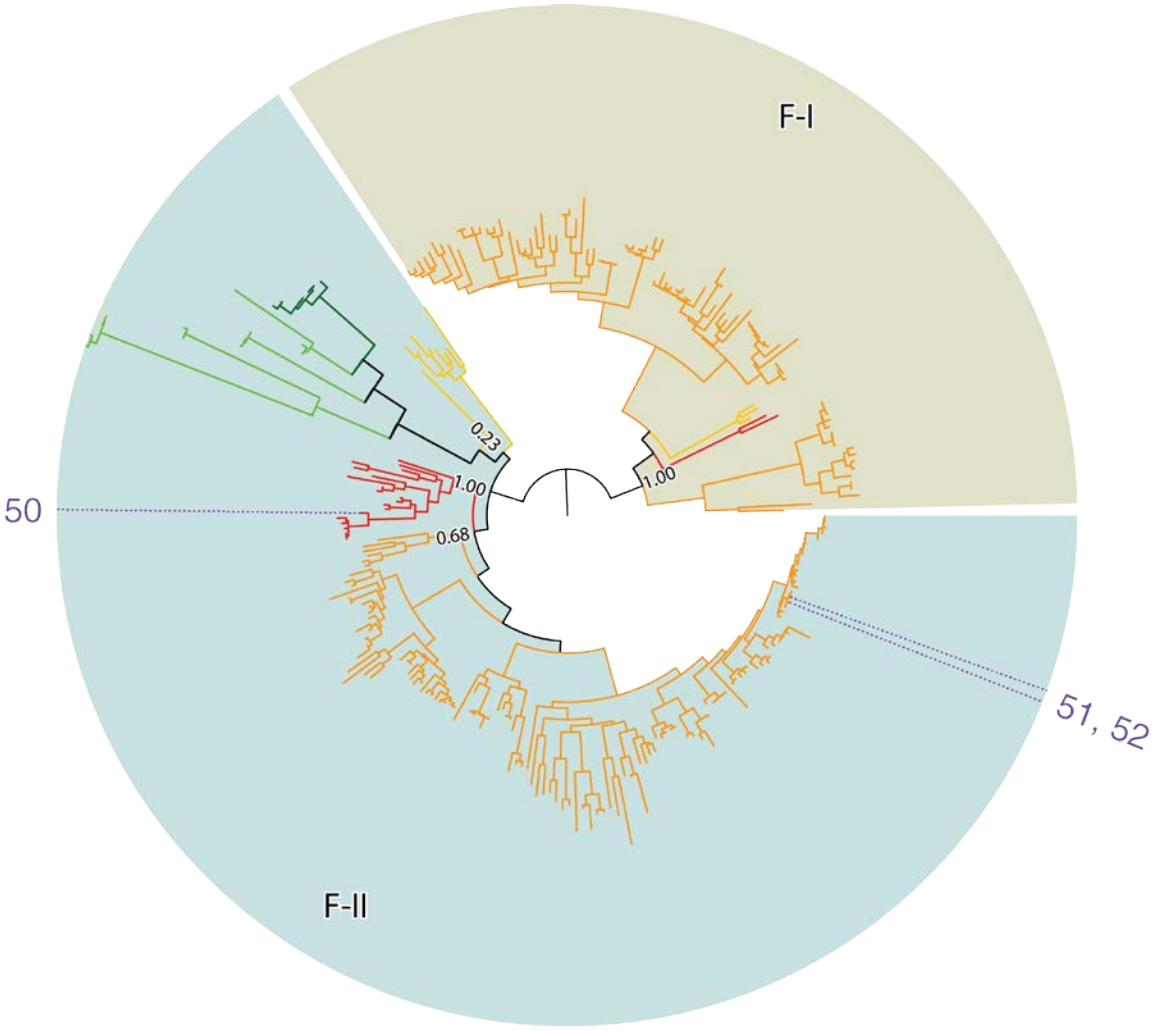
Supplementary Figure 4. Phylogenetic Expansion of BAHD Clade C
This clade is can be broken into sub clades C-I, C-II, C-III, and C-IV. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



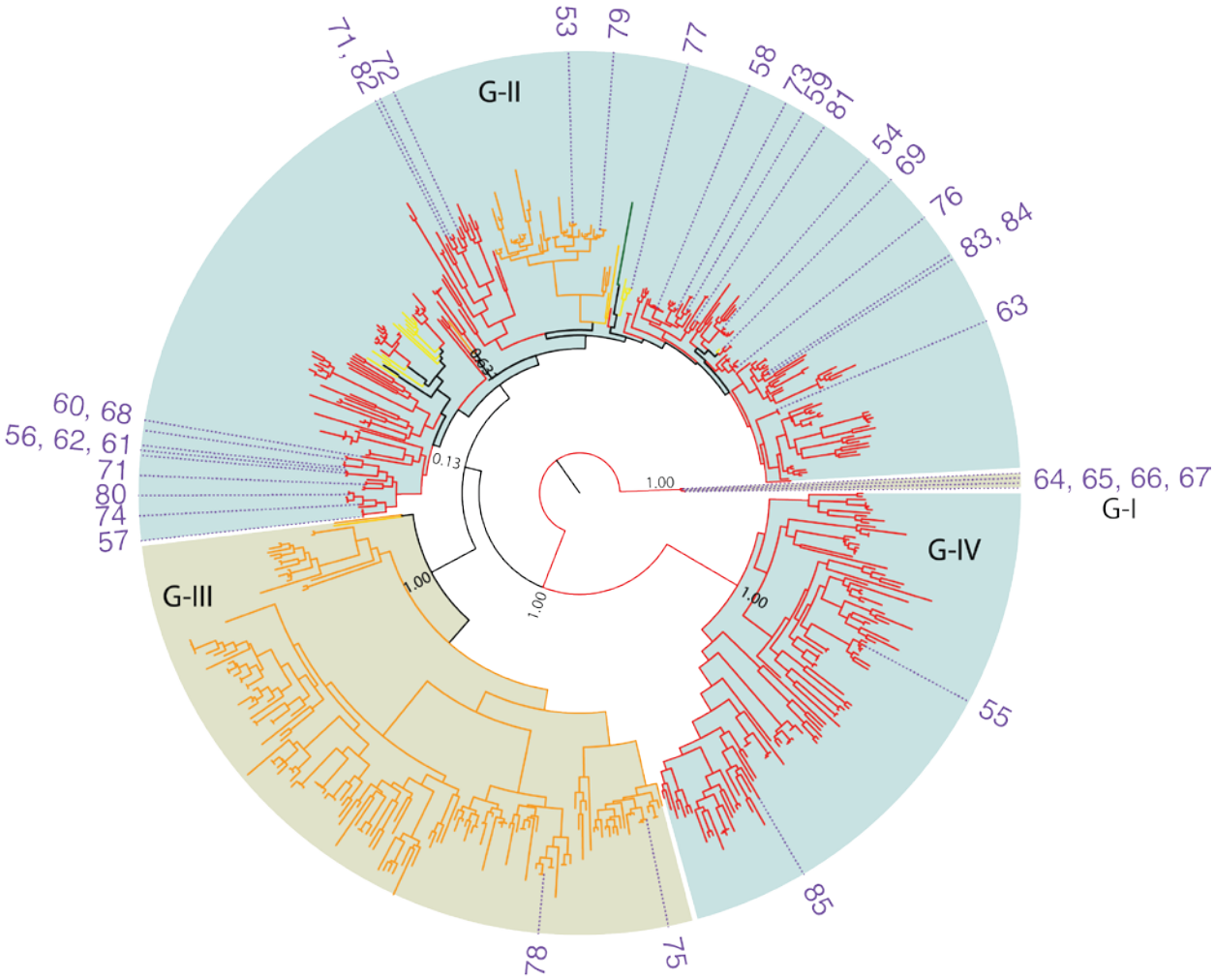
Supplementary Figure 5. Phylogenetic Expansion of BAHD Clade D
This clade is can be broken into sub clades D-I, D-II, and D-III. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



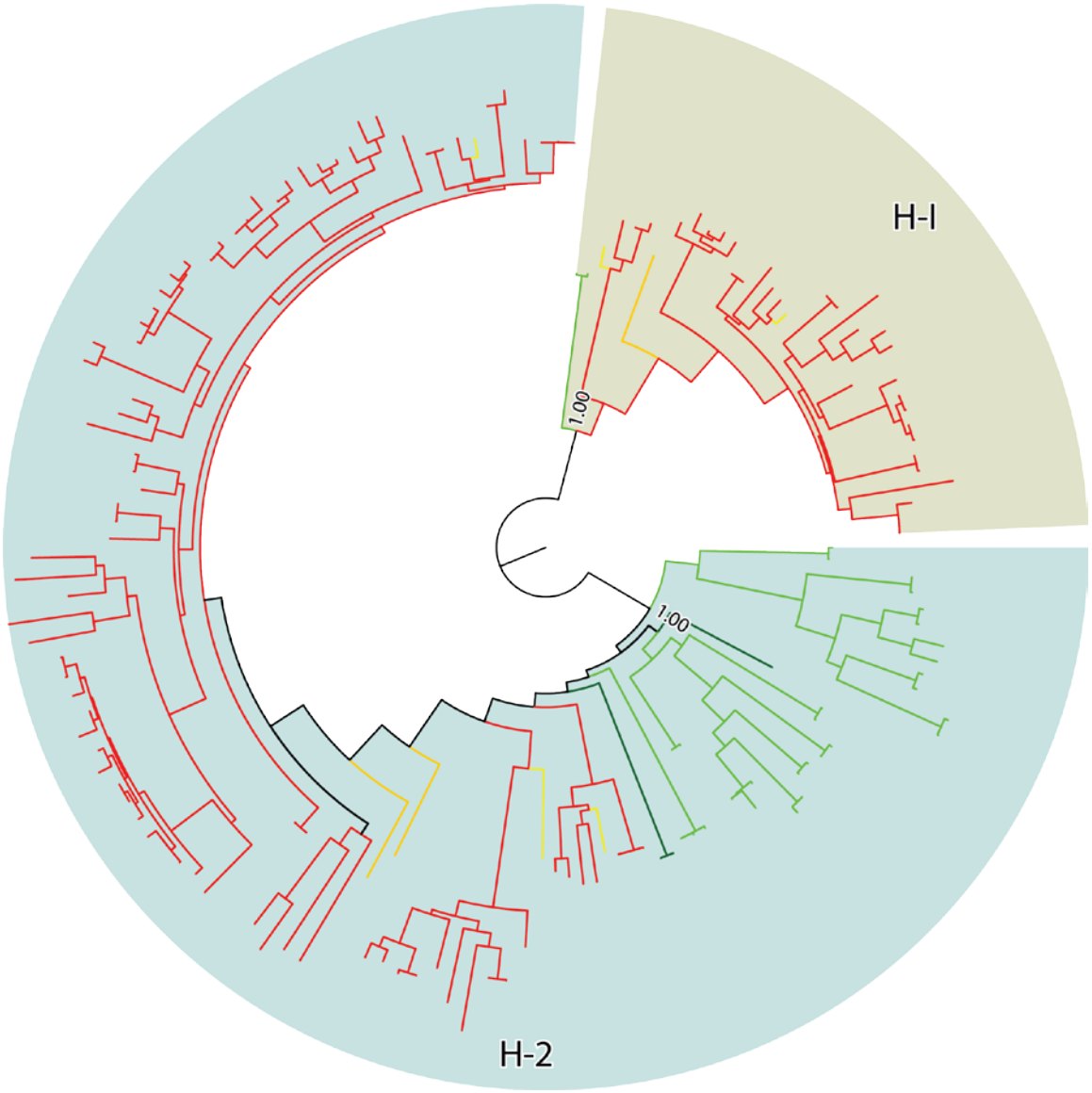
Supplementary Figure 6. Phylogenetic Expansion of BAHD Clade E
This clade is can be broken into sub clades E-I, E-II, and E-III. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



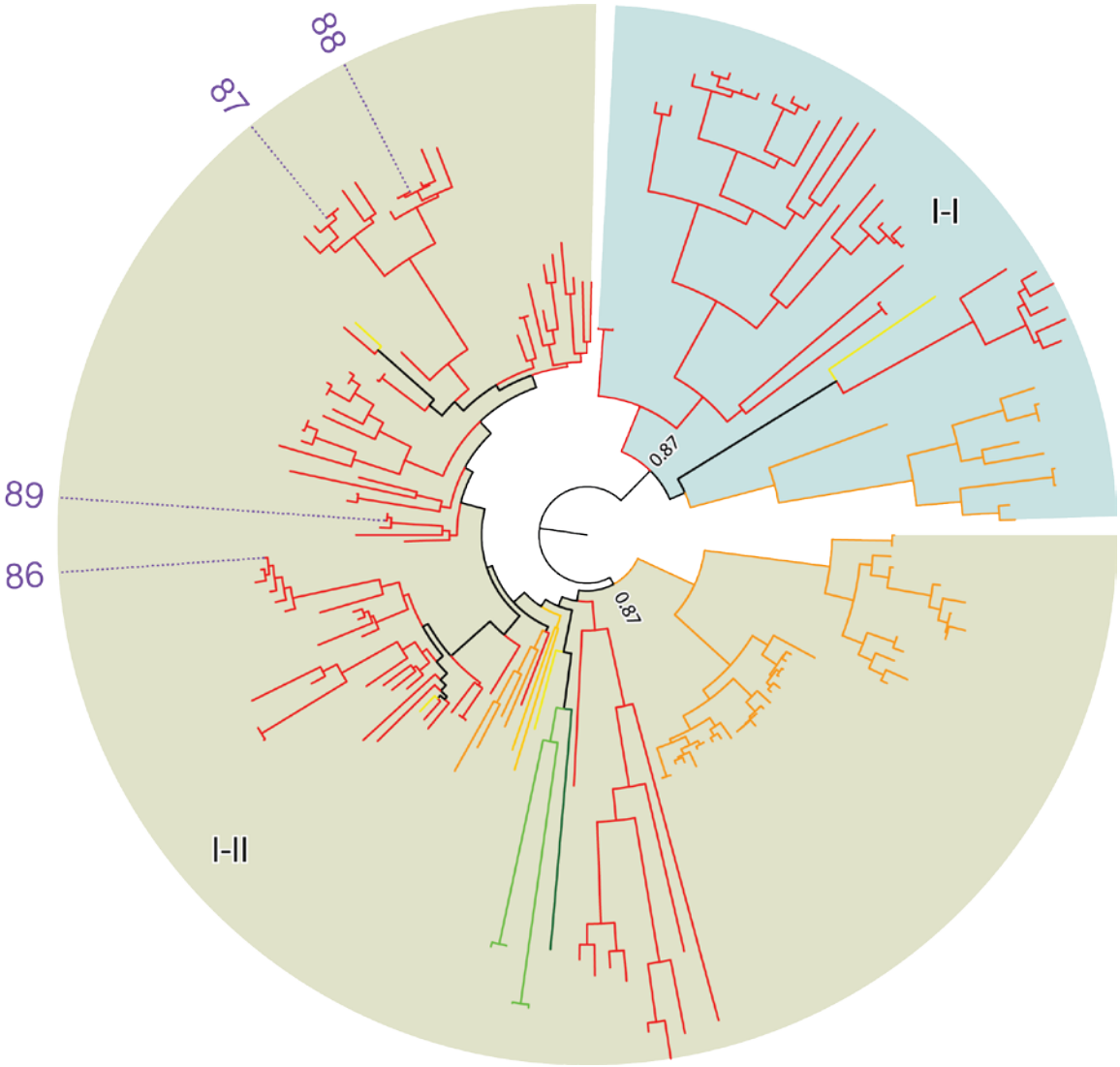
Supplementary Figure 7. Phylogenetic Expansion of BAHD Clade F
This clade is can be broken into sub clades F-I, and F-II. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



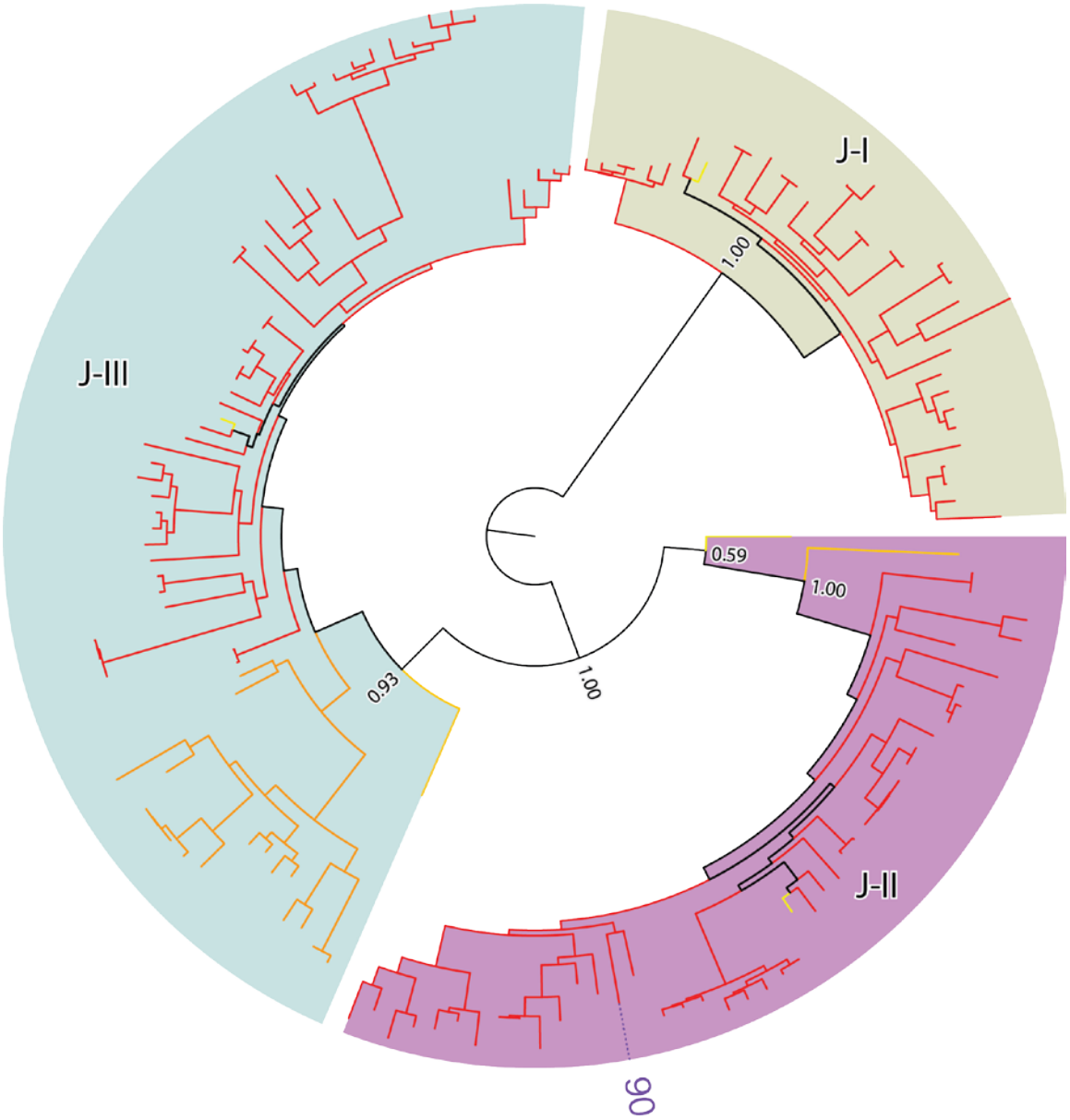
Supplementary Figure 8. Phylogenetic Expansion of BAHD Clade G
This clade is can be broken into sub clades G-I, G-II, G-III, and G-IV. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



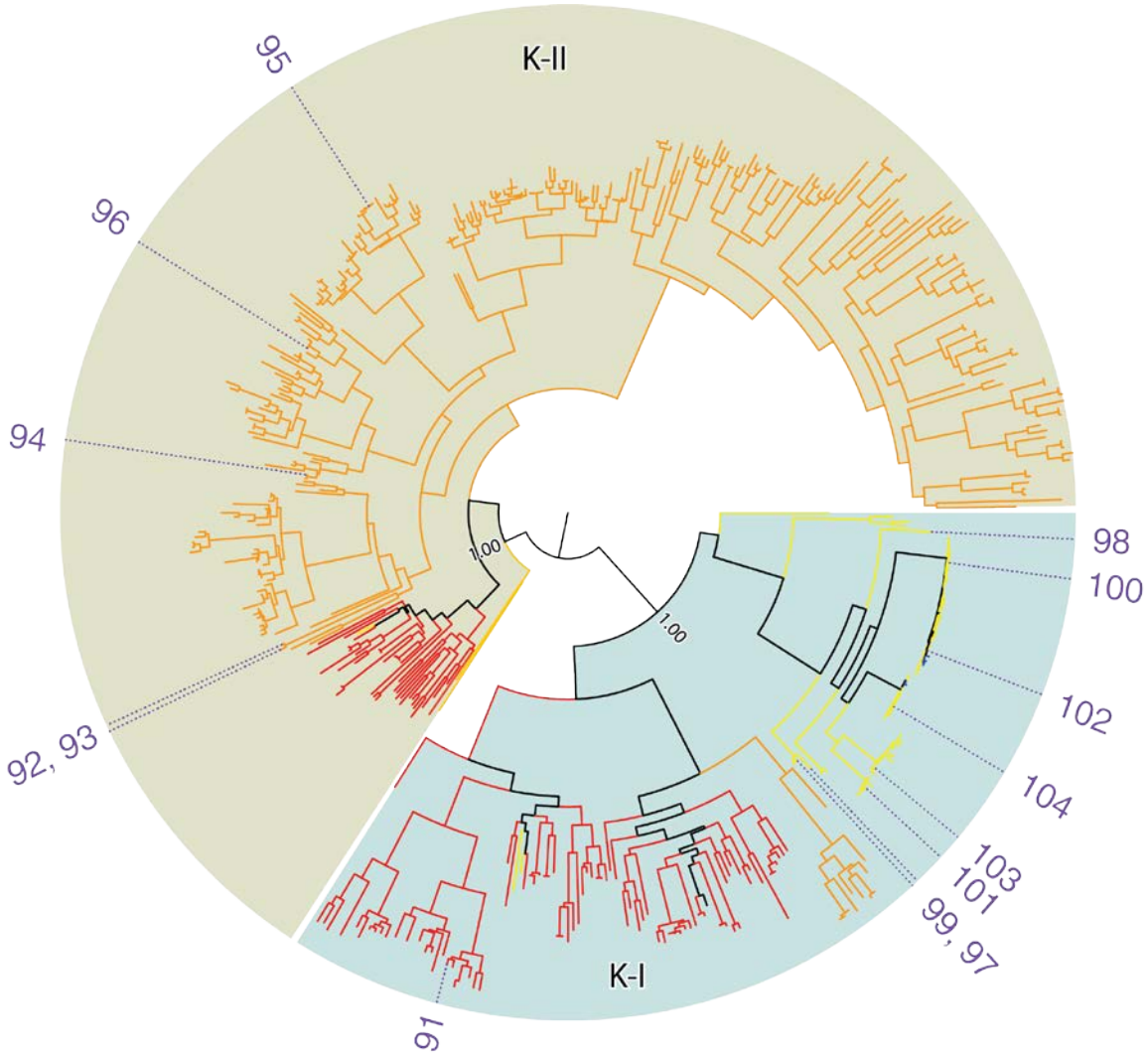
Supplementary Figure 9. Phylogenetic Expansion of BAHD Clade H
This clade is can be broken into sub clades H-I and H-II. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



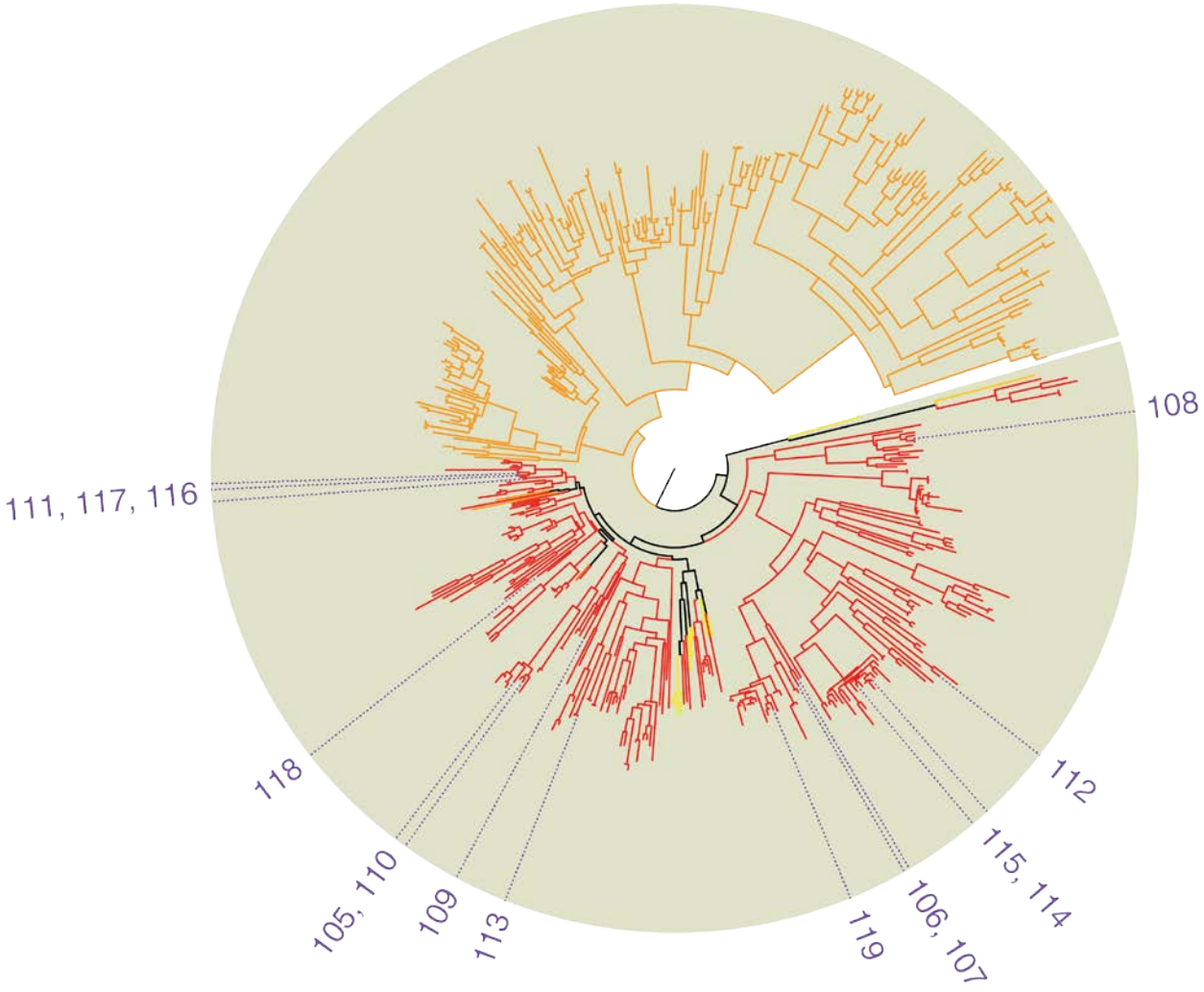
Supplementary Figure 10. Phylogenetic Expansion of BAHD Clade I
This clade is can be broken into sub clades I-I and I-II. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



Supplementary Figure 11. Phylogenetic Expansion of BAHD Clade J
Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.



Supplementary Figure 12. Phylogenetic Expansion of BAHD Clade K
This clade is can be broken into sub clades K-I and K-II. Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.

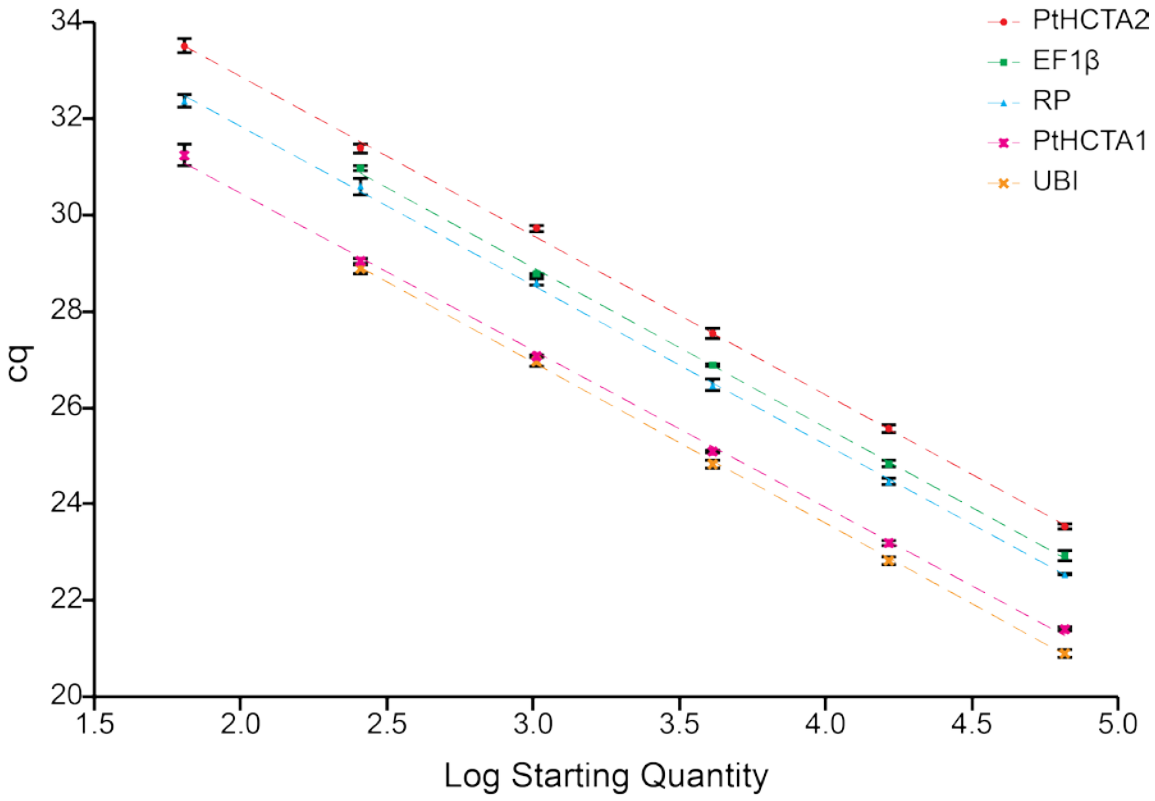


Supplementary Figure 13. Phylogenetic Expansion of BAHD Clade L
Enzyme numbering indicated in Supplementary Table 3; branch colour coding indicated in Table 2-1.

Supplementary Table 4. Primers Used for qPCR of HCTA Knock-Down Lines in *Populus* INRA 353-38

Primers which have previously been used for quantitation in *P. trichocarpa* and *Populus* INRA 717-1B4 were re-evaluated for use in *Populus* INRA 353-38.

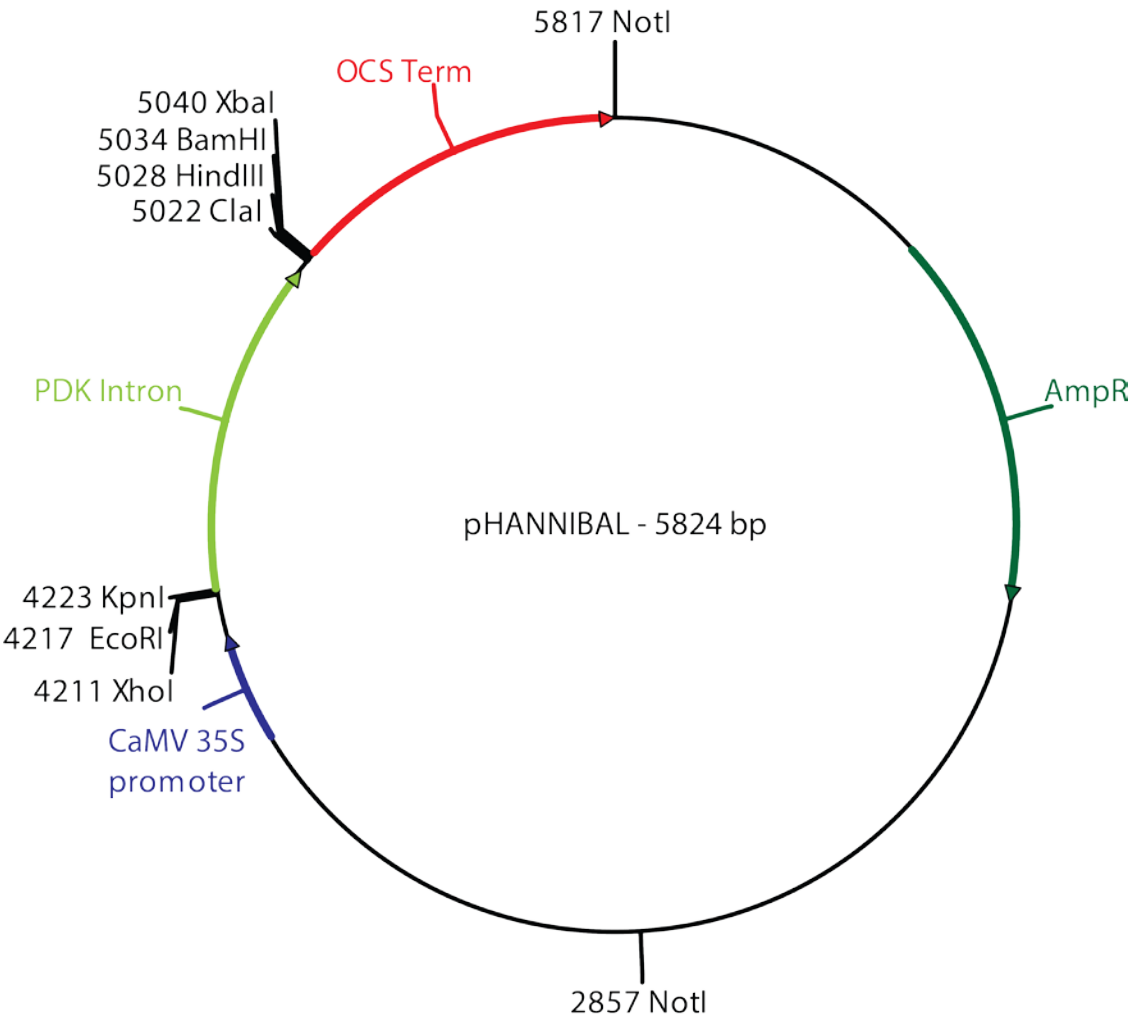
Name	Target	Sequence	Original Source
PtHCTA1_F	PtHCTA1 (Potri.001G042900)	TCCATCGACACCTGTTCACATGC	*P. trichocarpa* (Chapter 2)
PtHCTA1_R	PtHCTA1 (Potri.001G042900)	CACTCCCAGACCAAGTTCTCTTTG	*P. trichocarpa* (Chapter 2)
PtHCTA2_F	PtHCTA2 (Potri.003G183900)	TCAGGCTGGGAAAGTTCAGAGG	*P. trichocarpa* (Chapter 2)
PtHCTA2_R	PtHCTA2 (Potri.003G183900)	CTGCCTTGCATCAAACCATGTCC	*P. trichocarpa* (Chapter 2)
EF1β_F	elongation factor 1β (Potri.009G018600)	ACCTGGTCGTGATTTCCCTAATG	*Populus* INRA 717-1B4 (Tang, 2015)
EF1β_R	elongation factor 1β (Potri.009G018600)	GCCACAAATGCTTACACCAACA	*Populus* INRA 717-1B4 (Tang, 2015)
Ubi_R	Ubiquitin (Potri.014G115100)	ACCAAGCCCAAGAAGATCAAGCA	*Populus* INRA 717-1B4 (Tang, 2015)
Ubi_F	Ubiquitin (Potri.014G115100)	CCAGCACCGCACTCAGCA	*Populus* INRA 717-1B4 (Tang, 2015)
RP_R	ribosomal protein (Potri.001G342500)	TGTTGTGACCGCTGATTGTTTG	*P. trichocarpa* (Alber, 2016)
RP_F	ribosomal protein (Potri.001G342500)	CCACCTGTTCTTGCCTGTCTTA	*P. trichocarpa* (Alber, 2016)



Supplementary Figure 14. Primer Evaluation for Use in *Populus* INRA 353-38
gDNA was extracted from greenhouse-grown wild-type *Populus* INRA 353-38 xylem were pooled (n = 4). A four-fold serial dilution of was carried out on the gDNA. Primer pairs given in Supplemental Table 3-1 were used to amplify fragments of the transcripts indicated. The amplification threshold cycle (Cq) is plotted against starting template quantity for data points in the linear range with a standard error of mean (SEM) less than 0.5 Cq. PCR amplification efficiencies and the coefficient of determination (r^2) of a linear regression model were as follows: PtHCTA1 (Potri.001G042900; r^2 = 0.9989; efficiency = 101.39%), PtHCTA2 (Potri.003G183900; r^2 = 0.9992; efficiency = 100.87%), ubiquitin (UBI; Potri.014G115100; r^2 = 0.9997; efficiency = 99.86%), ribosomal peptide (RP; Potri.001G342500; r^2 = 0.9994; efficiency = 98.29%), elongation factor 1β (EF1β, Potri.009G018600; r^2 = 0.9994; efficiency = 99.83%). Error bars represent SEM (n=4).

Supplementary Table 5. HCT-RNAi Transgenics Lines Generated as Part of this Dissertation
The RNAi constructs were based on HCT candidate gene sequences from *Populus trichocarpa* Nisqually-1 and transformed into either *Populus* INRA 717-1B4 or *Populus* INRA 353-38. The total number of calli with transgene insertion confirmed by PCR are Indicated; however, transcriptional down regulation has not been confirmed.

Target gene pair	*Populus* INRA 717-1B4 transgenic lines	*Populus* INRA 353-38 transgenic lines
PtHCTA: HST	0	10
PtHCTB: Unknown HCT	12	16
PtHCTC: Unknown HCT	15	9
PtHCTD: Unknown HCT	6	11
PtHCTE: (SHT)	5	16



Supplementary Figure 15. pHANNIBAL Vector Used for Populus RNAi knock-down
pHANNIBAL vector was used to generate Populus INRA 717-1B4 and Populus INRA 353-38 RNAi knock-downs. Fragments were cloned into the first multiple cloning site (MCS) in the sense orientation using XhoI and KpnI, the reverse complement of the sequence was cloned into the second MCS using HindIII and ClaI, BamHI and ClaI, or BamHI and HindIII. The Cauliflower mosaic virus 35S RNA (CaMV 35S) promoter drives the generation of a transcript and transcription is terminated by octopine synthase (OCS Term). A pyruvate orthophosphate dikinase intron sequence from *Flaveria trinervia* helps to form the hairpin structure with the

sense orientation and anti-sense orientation RNA sequences. NotI cloning sites are used for transfer of this cassette into the pART27 *Agrobacterium tumefaciens* transformation vectors.

Supplementary Table 6. Primers Used to Generate HCT RNAi Knock-Down Cassettes

Primers include a spacer sequence (*italics*), appropriate restriction sites (**bold**) and a region of complementarity to the target gene (underline) sequence. Sense represents the sense arm of pHANNIBAL; anti-sense represents the anti-sense arm pHANNIBAL. The pHANNIBAL construct generates a transcript with a ~300 bp fragment (sense) and its reverse complement (anti-sense) flanking a PDK intron spacer.

Target	Direction	Forward Primer	Reverse Primer
PtHCTA	Sense	*GAGT***CTCGAG**GGTCAGATATGGCTCGTGGT	*GAGT***GGTACC**CTTCCTTTGCTTTGGCTTTG
PtHCTB	Sense	*GAGT***CTCGAG**GCAGTCAGGCATAGCACGTA	*GAGT***GGTACC**GCCCTCTCTTGAGAGCATTG
PtHCTC	Sense	*GAGT***CTCGAG**CAACACCCATTTCCAAAACC	*GAGT***GGTACC**TGTGAATTCTCTCCGCTGTG
PtHCTD	Sense	*GAGT***CTCGAG**CACGTGGAATCACAAACGAC	*GAGT***GGTACC**AAATAAGGTCGACCCCAACC
PtHCTE	Sense	*GAGT***CTCGAG**GTCACCTTTGCCTGATGGTT	*GAGT***GGTACC**CAGACCCGTCTCCATTAGGA
PtHCTA	Antisense	*GAGT***AAGCTT**GGTCAGATATGGCTCGTGGT	*GAGT***ATCGAT**CTTCCTTTGCTTTGGCTTTG
PtHCTB	Antisense	*GAGT***AAGCTT**GCAGTCAGGCATAGCACGTA	*GAGT***ATCGAT**GCCCTCTCTTGAGAGCATTG
PtHCTC	Antisense	*GAGT***AAGCTT**CAACACCCATTTCCAAAACC	*GAGT***ATCGAT**TGTGAATTCTCTCCGCTGTG
PtHCTD	Antisense	*GAGT***GGATCC**CACGTGGAATCACAAACGAC	*GAGT***ATCGAT**AAATAAGGTCGACCCCAACC
PtHCTE	Antisense	*GAGT***TCTAGA**GTCACCTTTGCCTGATGGTT	*GAGT***GGATCC**CAGACCCGTCTCCATTAGGA

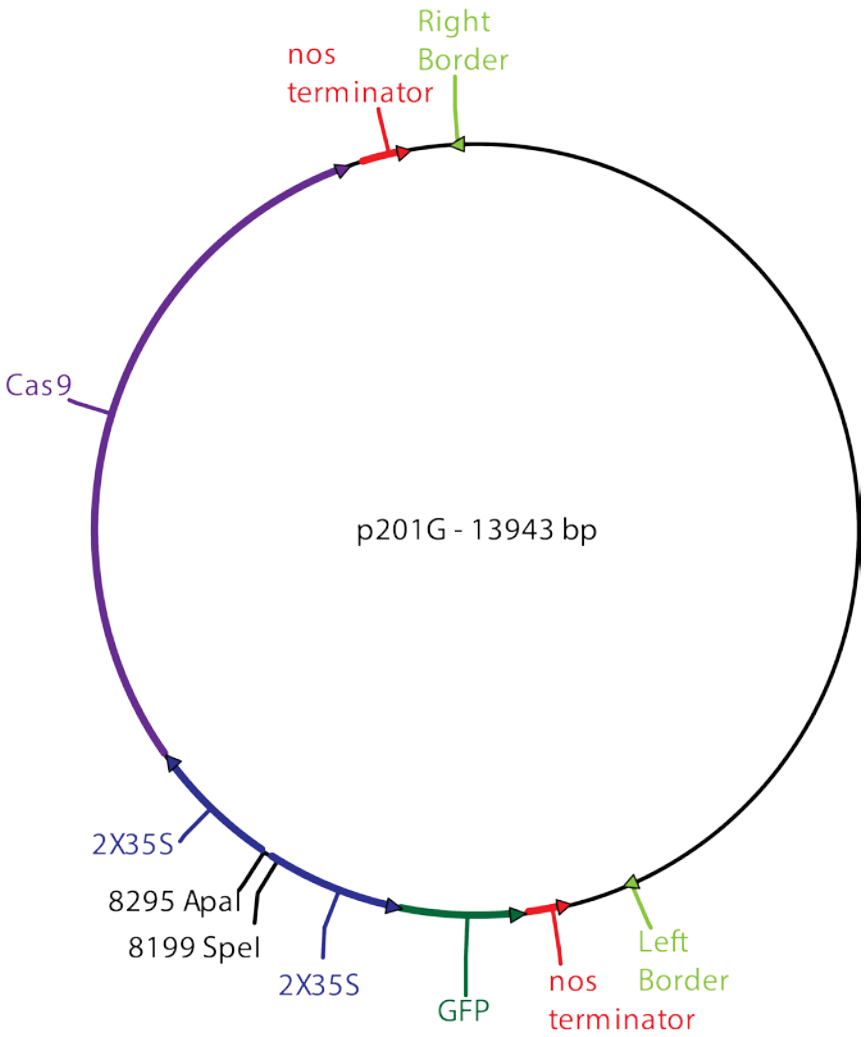
Supplementary Table 7. Lignin Composition of Ptthcta RNAi Knock-Down Lines
p-Hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin composition was determined based on thioacidolysis GCMS. Percentage relative to total are provided for H-lignin, S-lignin, and G-lignin as well as that ratio of S-lignin and H-lignin to G-lignin. PttHCTA knock-down lines show a clear change in the composition of lignin.

	%H	%G	%S	S:G	H:G
Ptthcta-1	27.5 ± 0.55	15.5 ± 0.45	57.0 ± 0.10	3.67 ± 0.11	1.8 ± 0.09
Ptthcta-2	29.1 ± 0.20	16.0 ± 0.18	54.9 ± 0.02	3.44 ± 0.04	1.8 ± 0.03
Ptthcta-9	11.5 ± 1.06	18.7 ± 0.51	69.8 ± 0.68	3.73 ± 0.14	0.6 ± 0.07
Ptthcta-4	2.8 ± 0.04	39.2 ± 0.00	58.0 ± 0.04	1.48 ± 0.00	0.1 ± 0.00
Ptthcta-8	0.4 ± 0.03	29.5 ± 1.18	70.1 ± 1.15	2.38 ± 0.13	0.0 ± 0.00
Empty-Vector	0.3 ± 0.03	27.3 ± 0.67	72.4 ± 0.68	2.65 ± 0.09	0.0 ± 0.00
Wild-Type	0.2 ± 0.06	28.8 ± 0.31	70.9 ± 0.31	2.46 ± 0.04	0.0 ± 0.00

```
 PtHCTC1  MQITIKESSM VVPTQDTPNH RLEVTNLDLF HAKYHVPLLF IYKPNGSSNF FEGKVLKEAL SKVLESFYPV AGRLARDAKG
PtaHCTC1  MQITIKESSM VLPTKDAPNH RQEVTNLDLF HAKYHVPLLF IYKPNGSSNF FEVKVLKEAL SKVLESFYPV AGRLARDANG
 PtHCTC2  MQITVKESSM VLPTQDTPDH RLEVTNLDLF HAKYHVPLLY LYKPNGSSNF FEVKVLKEAL SKVLESFYPV AGRLARDANG
PtaHCTC2  MQITVKESSM VLPTQDTPDH RLEVTNLDLF HAKYHVPLLY LYKPNGSSNF FEVKVLKEAL SKVLESFYPV AGRLARDANS

 PtHCTC1  RIEINCNGEG VLFVEAETDS AMGDFVGFKP SDELRQLIPT VDYSDISSYP LLVLQVTRFT CGGVCLGVGW QHTLADGTEC
PtaHCTC1  RIEINCDGEG VLFIEAETGS AMGDFCGFKP SDELRQLIPT VDYSDISSYP LLVI*VTRFT CGGVCLGVGW QHTLADGTEC
 PtHCTC2  RIEINCNGEG VLFIEAETDS AMGDFVDFKP SDELRQLIPT VDYSDISSYP LLVLQVTRFT CGGVCLGVGW HHTLADGTEC
PtaHCTC2  RIEINCDGEG VLFIEAETGS AMGDFVGFKP SDELRQLIPT VDYSDISSYP LLVLQVTRFA CGGVCLGVGW HHTLADGTEC

 PtHCTC1  LHFINTWSDI ARGLPVKTPP FIDRTILRGR VPPNPTFHHI EYDPYPTINT PFQNPIPESG SKDISVANLK ITSDLLNTLK
PtaHCTC1  LHFMNTWSDI ARGLPVKTPP FIERTILRGR VPPNPTFHHI EYDPFPTINT PFQNPIPESG SKDISVASLE ISSDLLDTLR
 PtHCTC2  LHFINTWSDI ARGLPVKTPP FIDRTILRGR VPPNPTFHHI EYDPFPTINT HFQNPIPESG SKDISVANLK IPSDLLNTLK
PtaHCTC2  LHFMNTWSDI ARGLPVKTPP FIDRTILRGR VPPNPTFHHI EYDPFPTINT PFQNPIPGSG SKDISVASLE IPSDLLDTLR

 PtHCTC1  AMAKNDIASK TEYSTYVILT AHIWRCACKA RGLSNDQATK LSIPTNGRDR FRPPLQPGYF GNVTFLATPI ALSGALLSEP
PtaHCTC1  AMAKSDIASN TEYSTYVILT AHIWRCACKA RGLSNDQATK LCMPTNGRNR FRPPIPPGYF GNVTFLATPI ALSGALLSEP
 PtHCTC2  AMAKNDIASK TEYSTYVILT AHIWRCACKA RGLSNDQATK INISTDGRNR FRPPIPPGYF GNVIFQATPI ALSGALLSEP
PtaHCTC2  AMAKNDIARK TEYSTYVILT AHIWRCACKA RGLSNDQATK LNISTDGRNR FRPPIPPGYF GNVIFQATLI ALSGDLLSEP

 PtHCTC1  LAHTAERIHK AIKRMDDEYL RSAVDYLEKV DDLTTVMRSS ETYRSPNLHI VNWVRLPFYD ADFGWGKPVY MRPASAFVGK
PtaHCTC1  LAHTAERIHK AIKRMDDEYL RSAVDYLEKV DDLTTVMRSP ETYRSPNLNI VSWVHLPFYD ADFGWGKPVY MRPASAFVGK
 PtHCTC2  LAHTAERIHK AIKRMDDEYL RSAVDYLERV DDFTTVMRSS ETYRSPNLTI ASWVRLPFYD ADFGWGKPVY MRPAFAFVGK
PtaHCTC2  LAHTAERIHK AIKRMDDEYL RSAVDYLEKV DDFTTVMRSP ETYRSPNLNI VSWMYLPFYD ADFGWGKPLY MRPAFAFEGK

 PtHCTC1  GYIQPSPTND GTLSLTIFLE TDHLQSFQKL FYEYHKRSCL *
PtaHCTC1  GYIQPSPTND GTLSLTIFLE TDHLQSFHKF FYEYSKRSCL *
 PtHCTC2  GYILPSPTDD GTLSLTICLE TDHLQSFQKL FYEYHKRSCL *
PtaHCTC2  GYILPGPAND GTLSLTICLE TDHLQSFHKF FYEYSKRSCL *
```

Supplementary Figure 16. Alignment of HCTC Paralogs

ClustalΩ was used to align PtaHCTC1 and PtHCTC1 (Potri.018G104800), and PtaHCTC2 and PtHCTC2 (Potri.018G104700) in order to highlight unique residues which are shown in red. The early stop codon in PtHCTC1 is PtaHCTC1 is outlined.



Supplementary Figure 17. p201G Vector Used for *Populus* CRISPR knock-out
p201G was used to generate *Populus* INRA 717-1B4 hairy root knock-outs. A fragment containing and MtU6 promoter, a CRISPR guide DNA, and CRISPR scaffold sequences was cloned between the SpeI and ApaI cut sites shows using Hot Fusion cloning. The right and left borders required for integration into the host genome are shown. The vector contains a GFP selection marker driven by two copies of cauliflower mosaic virus 35S RNA (CaMV 35S) promoters. As the sequence is introduced directionally, starting at the right border, the GFP sequence is adjacent to the left border ensuring expression only when the other fragments have been integrated. The expression of Cas9, the key protein in the CRISPR process, is also driven by two copies of the CaMV 35S promoter.

7. Bibliography

Abad, A.R., Cease, K.R., and Blanchette, R.A. 1988. A rapid technique using epoxy resin Quetol 651 to prepare woody plant tissues for ultrastructural study. Canadian Journal of Botany **66**(4): 677-682.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G.U.Y., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, Ø., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., and Taylor, M.F.J.R. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. Journal of Eukaryotic Microbiology **52**(5): 399-451.

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M.M.L., De Vos, R.C.H., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J., and O'Connell, A.P. 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. The Plant Cell **12**(5): 647-662.

Ahmad, M.Z., Li, P., Wang, J., Rehman, N.U., and Zhao, J. 2017. Isoflavone Malonyltransferases GmIMaT1 and GmIMaT3 Differently Modify Isoflavone Glucosides in Soybean (*Glycine max*) under Various Stresses. Frontiers in Plant Science **8**: 735.

Alber, A.V. 2016. Phenolic 3-hydroxylases in land plants: biochemical diversity and molecular evolution. Dissertation, Biology, University of Victoria, Victoria, Canada

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M.M. 2015. CRISPR/Cas9-mediated viral interference in plants. Genome Biology **16**: 238.

Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S., and Thompson, W.F. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nature Protocols **1**(5): 2320-2325.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology **215**(3): 403-410.

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

Anto, R.J., Sukumaran, K., Kuttan, G., Rao, M.N.A., Subbaraju, V., and Kuttan, R. 1995. Anticancer and antioxidant activity of synthetic chalcones and related compounds. Cancer Letters **97**(1): 33-37.

Aung, H.T., Nikai, T., Niwa, M., and Takaya, Y. 2010a. Rosmarinic acid in *Argusia argentea* inhibits snake venom-induced hemorrhage. Journal of Natural Medicines **64**(4): 482-486.

Aung, H.T., Nikai, T., Komori, Y., Nonogaki, T., Niwa, M., and Takaya, Y. 2010b. Biological and pathological studies of rosmarinic acid as an inhibitor of hemorrhagic *Trimeresurus* flavoviridis (habu) venom. Toxins **2**(10): 2478-2489.

Back, K., Jang, S.M., Lee, B.C., Schmidt, a., Strack, D., and Kim, K.M. 2001. Cloning and characterization of a hydroxycinnamoyl-CoA:tyramine N-(hydroxycinnamoyl)transferase induced in response to UV-C and wounding from *Capsicum annuum*. Plant and Cell Physiology **42**(5): 475-481.

Balbontín, C., Gaete-Eastman, C., Fuentes, L., Figueroa, C.R., Herrera, R., Manriquez, D., Latché, A., Pech, J.C., and Moya-León, M.A. 2010. VpAAT1, a gene encoding an alcohol acyltransferase, is involved in ester biosynthesis during ripening of mountain papaya fruit. Journal of Agricultural and Food Chemistry **58**(8): 5114-5121.

Bartley, L.E., Peck, M.L., Kim, S.-R.R., Ebert, B., Maniseri, C., Chiniquy, D., Sykes, R., Gao, L., Rautengarten, C., Vega-Sanchez, M.E., Benke, P.I., Canlas, P.E., Cao, P., Brewer, S., Lin, F., Smith, W.L., Zhang, X., Keasling, J.D., Jentoft, R.E., Foster, S.B., Zhou, J., Ziebell, A., An, G., Scheller, H.V., and Ronald, P.C. 2013. Overexpression of a BAHD acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. Plant Physiology **161**(4): 1615-1633.

Bayer, A., Ma, X., and Stöckigt, J. 2004. Acetyltransfer in natural product biosynthesis - functional cloning and molecular analysis of vinorine synthase. Bioorganic and Medicinal Chemistry **12**(10): 2787-2795.

Beckers, B., Op De Beeck, M., Weyens, N., Van Acker, R., Van Montagu, M., Boerjan, W., and Vangronsveld, J. 2016. Lignin engineering in field-grown poplar trees affects the endosphere bacterial microbiome. Proceedings of the National Academy of Sciences of the United States of America **113**(8): 2312-2317.

Beekwilder, J., Alvarez-Huerta, M., Neef, E., Verstappen, F.W.a., Bouwmeester, H.J., and Aharoni, A. 2004. Functional characterization of enzymes forming volatile esters from strawberry and banana. Plant Physiology **135**(4): 1865-1878.

Berger, A., Meinhard, J., and Petersen, M. 2006. Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. Planta **224**(6): 1503-1510.

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

Bhuiyan, N.H., Selvaraj, G., Wei, Y., and 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**(2): 509-521.

Blair, E.A. 1959. Process for producing aromatic carbonyl compounds and peroxide compounds. *Issued by* United States Patent and Trademark Office. *Assigned to* Welsbach Corp, United States of America. US2916499.

Boatright, J., Negre, F., Chen, X., Kish, C.M., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D., and Dudareva, N. 2004. Understanding *in vivo* benzenoid metabolism in petunia petal tissue. Plant Physiology **135**(4): 1993-2011.

Boatwright, J.L., and Pajerowska-Mukhtar, K. 2013. Salicylic acid: an old hormone up to new tricks. Molecular Plant Pathology **14**(6): 623-634.

Boerjan, W., Ralph, J., and Baucher, M. 2003. Lignin biosynthesis. Annual Review of Plant Biology **54**: 519-546.

Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**(15): 2114-2120.

Bradshaw, H.D., and Schemske, D.W. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature **426**(6963): 176-178.

Brown, D., Wightman, R., Zhang, Z., Gomez, L.D., Atanassov, I., Bukowski, J.-P., Tryfona, T., McQueen-Mason, S.J., Dupree, P., and Turner, S. 2011. *Arabidopsis* genes IRREGULAR XYLEM (IRX15) and IRX15L encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. The Plant Journal **66**(3): 401-413.

Burhenne, K., Kristensen, B.K., and Rasmussen, S.K. 2003. A new class of N-hydroxycinnamoyltransferases. Purification, cloning, and expression of a barley agmatine coumaroyltransferase (EC 2.3.1.64). Journal of Biological Chemistry **278**: 13919-13927.

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T., and Hou, W. 2015. CRISPR/Cas9-mediated genome editing in soybean hairy roots. PLOS One **10**(8): e0136064.

Camm, E.L., and Towers, G.N. 1973. Phenylalanine ammonia lyase. Phytochemistry **12**(5): 961-973.

Chau, M., Walker, K., Long, R., and Croteau, R. 2004. Regioselectivity of taxoid-O-acetyltransferases: heterologous expression and characterization of a new taxadien-5alpha-ol-O-acetyltransferase. Archives of Biochemistry and Biophysics **430**(2): 237-246.

Chedgy, R.J., Köllner, T.G., and Constabel, C.P. 2015. Functional characterization of two acyltransferases from *Populus trichocarpa* capable of synthesizing benzyl benzoate and salicyl benzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis. Phytochemistry **113**: 149-159.

Chen, H.-C., Li, Q., Shuford, C.M., Liu, J., Muddiman, D.C., Sederoff, R.R., and Chiang, V.L. 2011. Membrane protein complexes catalyze both 4- and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. Proceedings of the National Academy of Sciences of the United States of America **108**(52): 21253-21258.

Chowdhury, E., Choi, B.S., Park, S.U., Lim, H.S., and Bae, H. 2012. Transcriptional analysis of hydroxycinnamoyl transferase (HCT) in various tissues of *Hibiscus cannabinus* in response to abiotic stress conditions. Plant Omics **5**(3): 305-313.

Clé, C., Hill, L.M., Niggeweg, R., Martin, C.R., Guisez, Y., Prinsen, E., and Jansen, M.A.K. 2008. Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. Phytochemistry **69**(11): 2149-2156.

Clifford, M.N. 2000. Chlorogenic acids and other cinnamates - nature, occurrence, dietary burden, absorption and metabolism. Journal of the Science of Food and Agriculture **80**(7): 1033-1043.

Coleman, H.D., Park, J.-Y., Nair, R., Chapple, C., and Mansfield, S.D. 2008. RNAi-mediated suppression of *p*-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. Proceedings of the National Academy of Sciences **105**(11): 4501-4506.

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

Comino, C., Lanteri, S., Portis, E., Acquadro, A., Romani, A., Hehn, A., Larbat, R., and Bourgaud, F. 2007. Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. BMC Plant Biology **7**: 14.

Compton, D.L., Laszlo, J.A., and Berhow, M.A. 2000. Lipase-catalyzed synthesis of ferulate esters. Journal of the American Oil Chemists' Society **77**(5): 513-519.

Corea, O., Biswas, S., Hefer, C.A., Kolosova, N., Hannemann, J., Le, C.H., Mansfield, S.D., Douglas, C.J., Constabel, C.P., and Ehlting, J. 2017. Discovering co-expression networks in leaf and xylem of *Populus trichocarpa* by population-wide transcriptome analysis (unpublished).

Cumplido-Laso, G., Medina-Puche, L., Moyano, E., Hoffmann, T., Sinz, Q., Ring, L., Studart-Wittkowski, C., Caballero, J.L., Schwab, W., Muñoz-Blanco, J., and Blanco-Portales, R. 2012. The fruit ripening-related gene FaAAT2 encodes an acyl transferase involved in strawberry aroma biogenesis. Journal of Experimental Botany **63**(11): 4275-4290.

D'Auria, J.C. 2006. Acyltransferases in plants: a good time to be BAHD. Current Opinion in Plant Biology **9**(3): 331-340.

D'Auria, J.C., Chen, F., and 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**(1): 466-476.

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

D'Auria, J.C., Pichersky, E., Schaub, A., Hansel, A., and Gershenzon, J. 2007b. 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**(2): 194-207.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27**(8): 1164-1165.

Davis, K.R., and Hahlbrock, K. 1987. Induction of defense responses in cultured parsley cells by plant cell wall fragments. Plant Physiology **84**(4): 1286-1290.

Davison, B.H., Drescher, S.R., Tuskan, G.A., Davis, M.F., and Nghiem, N.P. 2006. Variation of S/G ratio and lignin content in a *Populus* family influences the release of xylose by dilute acid hydrolysis. Applied Biochemistry and Biotechnology **129-132**: 427-435.

Dexter, R., Qualley, A., Kish, C.M., Ma, C.J., Koeduka, T., Nagegowda, D.A., Dudareva, N., Pichersky, E., and Clark, D. 2007. Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. The Plant Journal **49**(2): 265-275.

Dexter, R.J., Verdonk, J.C., Underwood, B.a., Shibuya, K., Schmelz, E.a., and Clark, D.G. 2008. Tissue-specific PhBPBT expression is differentially regulated in response to endogenous ethylene. Journal of Experimental Botany **59**(3): 609-618.

Dhaubhadel, S., Farhangkhoee, M., and Chapman, R. 2008. Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. Journal of Experimental Botany **59**(4): 981-994.

Dicke, M., and Baldwin, I.T. 2010. The evolutionary context for herbivore-induced plant volatiles: beyond the 'cry for help'. Trends in Plant Science **15**(3): 167-175.

Ducki, S., Forrest, R., Hadfield, J.A., Kendall, A., Lawrence, N.J., McGown, A.T., and Rennison, D. 1998. Potent antimitotic and cell growth inhibitory properties of substituted chalcones. Bioorganic & Medicinal Chemistry Letters **8**(9): 1051-1056.

Dudareva, N., D'Auria, J.C., Nam, K.H., Raguso, R.A., and Pichersky, E. 1998. Acetyl-CoA:benzylalcohol acetyltransferase--an enzyme involved in floral scent production in *Clarkia breweri*. The Plant Journal **14**(3): 297-304.

Ehlting, J., Hamberger, B., Million-Rousseau, R., and Werck-Reichhart, D. 2006. Cytochromes P450 in phenolic metabolism. Phytochemistry Reviews **5**(2-3): 239-270.

El-Sharkawy, I., Manríquez, D., Flores, F.B., Regad, F., Bouzayen, M., Latché, A., and 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**(2): 345-362.

Elejalde-Palmett, C., de Bernonville, T.D., Glevarec, G., Pichon, O., Papon, N., Courdavault, V., St-Pierre, B., Giglioli-Guivarc'h, N., Lanoue, A., and Besseau, S. 2015. Characterization of a spermidine hydroxycinnamoyltransferase in *Malus domestica* highlights the evolutionary conservation of trihydroxycinnamoyl spermidines in pollen coat of core eudicotyledons. Journal of Experimental Botany **66**(22): 7271-7285.

English, S., Greenaway, W., and Whatley, F.R. 1991. Analysis of phenolics of *Populus trichocarpa* bud exudate by GC-MS. Phytochemistry **30**(2): 531-533.

English, S., Greenaway, W., and Whatley, F.R. 1992. Analysis of phenolics in the bud exudates of *Populus deltoides*, *P. fremontii*, *P. sargentii* and *P. Wislizenii* by GC-MS. Phytochemistry **31**(4): 1255-1260.

Escamilla-Trevino, L.L., Shen, H., Hernandez, T., Yin, Y., Xu, Y., and Dixon, R.A. 2014. Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (*Panicum virgatum* L.). Plant Molecular Biology **84**(4-5): 565-576.

Eudes, A., Pereira, J.H., Yogiswara, S., Wang, G., Teixeira Benites, V., Baidoo, E.E., Lee, T.S., Adams, P.D., Keasling, J.D., and Loque, D. 2016. Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase to reduce lignin. Plant and Cell Physiology **57**(3): 568-579.

Fang, J., and Ewald, D. 2004. Expression cloned cDNA for 10-deacetylbaccatin III-10-O-acetyltransferase in *Escherichia coli*: A comparative study of three fusion systems. Protein Expression and Purification **35**(1): 17-24.

Farmer, M.J., Czernic, P., Michael, A., and Negrel, J. 1999. Identification and characterization of cDNA clones encoding hydroxycinnamoyl-CoA:tyramine N-hydroxycinnamoyltransferase from tobacco. The FEBS Journal **263**(3): 686-694.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**(4): 783-791.

Felsenstein, J. 1989. PHYLIP - phylogeny inference package (version 3.2). Cladistics **5**: 164-166.

Franke, R., Hemm, M.R., Denault, J.W., Ruegger, M.O., Humphreys, J.M., and Chapple, C. 2002. Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of *Arabidopsis*. The Plant Journal **30**(1): 47-59.

Fu, C., Donovan, W.P., Shikapwashya-Hasser, O., Ye, X., and Cole, R.H. 2015a. Hot Fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. PLOS One **9**(12): e115318.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics **28**(23): 3150-3152.

Fu, X., Yin, Z.P., Chen, J.G., Shangguan, X.C., Wang, X., Zhang, Q.F., and Peng, D.Y. 2015b. Production of chlorogenic acid and its derivatives in hairy root cultures of *Stevia rebaudiana*. Journal of Agricultural and Food Chemistry **63**(1): 262-268.

Fujiwara, H., Tanaka, Y., Yonekura-Sakakibara, K., Fukuchi-Mizutani, M., Nakao, M., Fukui, Y., Yamaguchi, M., Ashikari, T., and Kusumi, T. 1998. cDNA cloning, gene expression and

subcellular localization of anthocyanin 5-aromatic acyltransferase from *Gentiana triflora*. The Plant Journal **16**: 421-431.

Furtado, A., Lupoi, J.S., Hoang, N.V., Healey, A., Singh, S., Simmons, B.A., and Henry, R.J. 2014. Modifying plants for biofuel and biomaterial production. Plant Biotechnology Journal **12**(9): 1246-1258.

Garvey, G.S., McCormick, S.P., and Rayment, I. 2008. Structural and functional characterization of the TRI101 trichothecene 3-O-acetyltransferase from *Fusarium sporotrichioides* and *Fusarium graminearum*: Kinetic insights to combating *Fusarium* head blight. Journal of Biological Chemistry **283**(3): 1660-1669.

Gawel, R., Schulkin, A., Smith, P.A., and Waters, E.J. 2014. Taste and textural characters of mixtures of caftaric acid and grape reaction product in model wine. Australian Journal of Grape and Wine Research **20**(1): 25-30.

Gerasimenko, I., Ma, X., Sheludko, Y., Mentele, R., Lottspeich, F., and Stöckigt, J. 2004. Purification and partial amino acid sequences of the enzyme vinorine synthase involved in a crucial step of ajmaline biosynthesis. Bioorganic and Medicinal Chemistry **12**(10): 2781-2786.

Giri, A., and Narasu, M.L. 2000. Transgenic hairy roots: recent trends and applications. Biotechnology Advances **18**(1): 1-22.

Gleave, A.P. 1992. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Molecular Biology **20**(6): 1203-1207.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40**(Database issue): D1178-1186.

Gou, J.-Y., Yu, X.-H., and Liu, C.-J. 2009. A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **106**(44): 18855-18860.

Gould, K.S., McKelvie, J., and Markham, K.R. 2002. Do anthocyanins function as antioxidants in leaves? Imaging of H2O2 in red and green leaves after mechanical injury. Plant, Cell & Environment **25**(10): 1261-1269.

Grace, S.C., Logan, B.A., and Adams, W.W. 1998. Seasonal differences in foliar content of chlorogenic acid, a phenylpropanoid antioxidant, in Mahonia repens. Plant, Cell & Environment **21**(5): 513-521.

Greenaway, W., and Whatley, F.R. 1990. Analysis of phenolics of bud exudate of *Populus angustifolia* by GC-MS. Phytochemistry **29**(8): 2551-2554.

Greenaway, W., and Whatley, F.R. 1991. Analysis of phenolics of bud exudate of *Populus ciliata* by GC-MS. Phytochemistry **30**(6): 1887-1889.

Greenaway, W., Scaysbrook, T., and Whatley, F.R. 1987. The analysis of bud exudate of *Populus x euramericana*, and of propolis, by gas chromatography-mass spectrometry. Proceedings of the Royal Society B **232**(1268): 249-272.

Greenaway, W., Scaysbrook, T., and Whatley, F.R. 1988. Phenolic analysis of bud exudate of *Populus lasiocarpa* by GC/MS. Phytochemistry **27**(11): 3513-3515.

Greenaway, W., English, S., and Whatley, F.R. 1990. Phytochemical affinity of *Populus angustifolia* with sect. *Aigeiros* poplars established by GC-MS of bud exudate. Biochemical Systematics and Ecology **18**(6): 439-445.

Greenaway, W., English, S., May, J., and Whatley, F.R. 1991a. Chemotaxonomy of section *Leuce* poplars by GC-MS of bud exudate. Biochemical Systematics and Ecology **19**(6): 507-518.

Greenaway, W., English, S., May, J., and Whatley, F.R. 1991b. Analysis of phenolics of bud exudate of *Populus sieboldii* by GC-MS. Phytochemistry **30**(9): 3005-3008.

Grienenberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C., Pollet, B., Heitz, T., and 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.

Grothe, T., Lenz, R., and Kutchan, T.M. 2001. Molecular characterization of the salutaridinol 7-O-acetyltransferase involved in morphine biosynthesis in opium poppy *Papaver somniferum*. Journal of Biological Chemistry **276**(33): 30717-30723.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology **52**(5): 696-704.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology **59**(3): 307-321.

Günther, C.S., Chervin, C., Marsh, K.B., Newcomb, R.D., and Souleyre, E.J.F. 2011. Characterisation of two alcohol acyltransferases from kiwifruit (*Actinidia* spp.) reveals distinct substrate preferences. Phytochemistry **72**: 700-710.

Guo, B., Kai, G., Gong, Y., Jin, H., Wang, Y., Miao, Z., Sun, X., and Tang, K. 2007. Molecular cloning and heterologous expression of a 10-deacetylbaccatin III-10-O-acetyl transferase cDNA from *Taxus x media*. Molecular Biology Reports **34**(2): 89-95.

Guo, J., Carrington, Y., Alber, A., and Ehlting, J. 2014. Molecular characterization of quinate and shikimate metabolism in *Populus trichocarpa*. Journal of Biological Chemistry **289**(34): 23846-23858.

Ha, C.M., Escamilla-Trevino, L., Yarce, J.C., Kim, H., Ralph, J., Chen, F., and Dixon, R.A. 2016. An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. The Plant Journal **86**(5): 363-375.

Han, K.-H., Gordon, M.P., and Strauss, S.H. 1997. High-frequency transformation of cottonwoods (genus *Populus*) by *Agrobacterium rhizogenes*. Canadian Journal of Forest Research **27**(4): 464-470.

Harborne, J., and Corner, J. 1961. Plant polyphenols. 4. Hydroxycinnamic acid-sugar derivatives. Biochemical Journal **81**(2): 242-250.

Haslam, T.M., Manas-Fernandez, a., Zhao, L., and Kunst, L. 2012. *Arabidopsis* ECERIFERUM2 is a component of the fatty acid elongation machinery required for fatty acid extension to exceptional lengths. Plant Physiology **160**(3): 1164-1174.

Hefer, C.A., Mizrachi, E., Myburg, A.A., Douglas, C.J., and Mansfield, S.D. 2015. Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis*. New Phytologist Journal **206**(4): 1391-1405.

Herman, W. 1949. Dihydrosafrol derivatives. *Issued by* United States Patent and Trademark Office. *Assigned to* U.S. Industrial Chemicals Inc.,, United States of America. US2485680.

Hether, N.H., Olsen, R.A., and Jackson, L.L. 1984. Chemical identification of iron reductants exuded by plant roots. Journal of Plant Nutrition **7**(1-5): 667-676.

Hirabayashi, T., Ochiai, H., Sakai, S., Nakajima, K., and Terasawa, K. 1995. Inhibitory effect of ferulic acid and isoferulic acid on murine interleukin-8 production in response to influenza virus infections *in vitro* and *in vivo*. Planta Medica **61**(03): 221-226.

Hoagland, D.R., and Arnon, D.I. 1950. The water-culture method for growing plants without soil.

Hoffmann, F. 1900. Acetyl salicylic acid. *Issued by* United States Patent and Trademark Office. *Assigned to* Farbenfabriken Of Elberfeld Company,, United States of America. US644077.

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., and Legrand, M. 2003. Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. Journal of Biological Chemistry **278**(1): 95-103.

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

Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., Pollet, B., and Legrand, M. 2005. Acyltransferase-catalysed *p*-coumarate ester formation is a committed step of lignin biosynthesis. Plant Biosystems **139**(1): 50-53.

Holderbaum, D.F., Kon, T., Kudo, T., and Guerra, M.P. 2010. Enzymatic browning, polyphenol oxidase activity, and polyphenols in four apple cultivars: dynamics during fruit development. HortScience **45**(8): 1150-1154.

Hooker, C.W., Lott, W.B., and Harrich, D. 2001. Inhibitors of human immunodeficiency virus type 1 reverse transcriptase target distinct phases of early reverse transcription. Journal of Virology **75**(7): 3095-3104.

Huang, Q.S., Zhu, Y.J., Li, H.L., Zhuang, J.X., Zhang, C.L., Zhou, J.J., Li, W.G., and Chen, Q.X. 2009. Inhibitory effects of methyl trans-cinnamate on mushroom tyrosinase and its antimicrobial activities. Journal of Agricultural and Food Chemistry **57**(6): 2565-2569.

Hufnagel, J.C., and Hofmann, T. 2008. Quantitative reconstruction of the nonvolatile sensometabolome of a red wine. Journal of Agricultural and Food Chemistry **56**(19): 9190-9199.

Hwang, S.Y., and Lindroth, R.L. 1997. Clonal variation in foliar chemistry of aspen: effects on gypsy moths and forest tent caterpillars. Oecologia **111**(1): 99-108.

Iaffaldano, B., Zhang, Y., and Cornish, K. 2016. CRISPR/Cas9 genome editing of rubber producing dandelion *Taraxacum kok-saghyz* using *Agrobacterium rhizogenes* without selection. Industrial Crops and Products **89**: 356-362.

Ikonen, A., Tahvanainen, J., and Roininen, H. 2001. Chlorogenic acid as an antiherbivore defence of willows against leaf beetles. Entomologia Experimentalis et Applicata **99**(1): 47-54.

Isidorov, V.A., and Vinogorova, V.T. 2003. GC-MS analysis of compounds extracted from buds of *Populus balsamifera* and *Populus nigra*. Zeitschrift für Naturforschung C **58**(5-6): 355-360.

IUPAC. 1976. IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) and IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Nomenclature of cyclitols. Recommendations, 1973. Biochemical Journal **153**(1): 23-31.

Izaguirre, M.M., Mazza, C.A., SvatoŠ, A., Baldwin, I.T., and BallarÉ, C.L. 2007. Solar ultraviolet-B radiation and insect herbivory trigger partially overlapping phenolic responses in *Nicotiana attenuata* and *Nicotiana longiflora*. Annals of Botany **99**(1): 103-109.

Jacobs, T.B., LaFayette, P.R., Schmitz, R.J., and Parrott, W.A. 2015. Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnology **15**(1): 16.

Katsumoto, Y., Fukuchi-Mizutani, M., Fukui, Y., Brugliera, F., Holton, T.a., Karan, M., Nakamura, N., Yonekura-Sakakibara, K., Togami, J., Pigeaire, A., Tao, G.Q., Nehra, N.S., Lu, C.Y., Dyson, B.K., Tsuda, S., Ashikari, T., Kusumi, T., Mason, J.G., and Tanaka, Y. 2007. Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin. Plant and Cell Physiology **48**(11): 1589-1600.

Kikuzaki, H., Hisamoto, M., Hirose, K., Akiyama, K., and Taniguchi, H. 2002. Antioxidant properties of ferulic acid and Its related compounds. Journal of Agricultural and Food Chemistry **50**(7): 2161-2168.

Kim, I.A., Kim, B.-G.G., Kim, M., and Ahn, J.-H.H. 2012. Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. Phytochemistry **76**: 25-31.

Kimura, M., Kaneko, I., Komiyama, M., Takatsuki, A., Koshino, H., Yoneyama, K., and Yamaguchi, I. 1998. Trichothecene 3-O-acetyltransferase protects both the producing organism and transformed yeast from related mycotoxins. Cloning and characterization of Tri101. Journal of Biological Chemistry **273**(3): 1654-1661.

Kind, T., and Fiehn, O. 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics **8**: 105.

Kirchner, T.W., Niehaus, M., Debener, T., Schenk, M.K., and Herde, M. 2017. Efficient generation of mutations mediated by CRISPR/Cas9 in the hairy root transformation system of *Brassica carinata*. PLOS One **12**(9): e0185429.

Kishimoto, N., Kakino, Y., Iwai, K., Mochida, K., and Fujita, T. 2005. *In vitro* antibacterial, antimutagenic and anti-influenza virus activity of caffeic acid phenethyl esters. Biocontrol Science **10**(4): 155-161.

Knobloch, K.H., and Hahlbrock, K. 1977. 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. partial purification, substrate specificity, and further properties. Archives of Biochemistry and Biophysics **184**(1): 237-248.

Kolosova, N., Miller, B., Ralph, S., Ellis, B.E., Douglas, C., Ritland, K., and Bohlmann, J. 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. Nature Reviews Genetics **2**: 353-359.

Kosma, D.K., Molina, I., Ohlrogge, J.B., Pollard, M., and Oa, R.W.W. 2012. Identification of an *Arabidopsis* fatty alcohol:caffeoyl-Coenzyme A acyltransferase required for the synthesis of alkyl hydroxycinnamates in root waxes. Plant Physiology **160**: 237-248.

Krebs, H.A., Wiggins, D., Stubbs, M., Sols, A., and Bedoya, F. 1983. Studies on the mechanism of the antifungal action of benzoate. Biochemical Journal **214**(3): 657-663.

Laflamme, P., St-Pierre, B., and V, D.L. 2001. Molecular and biochemical analysis of a Madagascar periwinkle root-specific minovincinine-19-hydroxy-O-acetyltransferase. Plant Physiology **125**: 189-198.

Lallemand, L.a., McCarthy, J.G., McSweeney, S., and McCarthy, A.a. 2012a. Purification, crystallization and preliminary X-ray diffraction analysis of a hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) from *Coffea canephora* involved in chlorogenic acid biosynthesis. Acta Crystallographica Section F **68**(Pt 7): 824-828.

Lallemand, L.A., Zubieta, C., Lee, S.G., Wang, Y., Acajjaoui, S., Timmins, J., McSweeney, S., Jez, J.M., McCarthy, J.G., and McCarthy, A.a. 2012b. A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. Plant Physiology **160**: 249-260.

Landmann, C., Hücherig, S., Fink, B., Hoffmann, T., Dittlein, D., Coiner, H.A., and Schwab, W. 2011. Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). Planta **234**(2): 305-320.

Lapierre, C., and Rolando, C. 1988. Thioacidolyses of pre-methylated lignin samples from pine compression and poplar woods. *In* Holzforschung. p. 1.

Lee, S., Kaminaga, Y., Cooper, B., Pichersky, E., Dudareva, N., and Chapple, C. 2012. Benzoylation and sinapoylation of glucosinolate R-groups in *Arabidopsis*. The Plant Journal **72**(3): 411-422.

Lehfeldt, C., Shirley, A.M., Meyer, K., Ruegger, M.O., Cusumano, J.C., Viitanen, P.V., Strack, D., and Chapple, C. 2000. Cloning of the SNG1 gene of *Arabidopsis* reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. The Plant Cell **12**(8): 1295-1306.

Leiss, K.A., Maltese, F., Choi, Y.H., Verpoorte, R., and Klinkhamer, P.G.L. 2009. Identification of chlorogenic acid as a resistance factor for thrips in *Chrysanthemum*. Plant Physiology **150**(3): 1567-1575.

Lenz, R., and Zenk, M.H. 1995. Acetyl coenzyme A:salutaridinol-7-O-acetyltransferase from *Papaver somniferum* plant cell cultures: The enzyme catalyzing the formation of thebaine in morphine biosynthesis. Journal of Biological Chemistry **270**(52): 31091-31096.

Lepelley, M., Cheminade, G., Tremillon, N., Simkin, A., Caillet, V., and McCarthy, J. 2007. Chlorogenic acid synthesis in coffee: an analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. Plant Science **172**(5): 978-996.

Leple, J.C., Dauwe, R., Morreel, K., Storme, V., Lapierre, C., Pollet, B., Naumann, A., Kang, K.Y., Kim, H., Ruel, K., Lefebvre, A., Joseleau, J.P., Grima-Pettenati, J., De Rycke, R., Andersson-Gunneras, S., Erban, A., Fehrle, I., Petit-Conil, M., Kopka, J., Polle, A., Messens, E., Sundberg, B., Mansfield, S.D., Ralph, J., Pilate, G., and Boerjan, W. 2007. Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. The Plant Cell **19**(11): 3669-3691.

Li, D., Xu, Y., Xu, G., Gu, L., Li, D., and Shu, H. 2006. Molecular cloning and expression of a gene encoding alcohol acyltransferase (MdAAT2) from apple (cv. Golden Delicious). Phytochemistry **67**(7): 658-667.

Li, X., Bonawitz, N.D., Weng, J.K., and Chapple, C. 2010. The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thaliana* is independent of flavonoids. The Plant Cell **22**(5): 1620-1632.

Liang, Z., Zhang, K., Chen, K.L., and Gao, C.X. 2014. Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas System. Journal of Genetics and Genomics **41**.

Lin, C.-Y., Wang, Jack P., Li, Q., Chen, H.-C., Liu, J., Loziuk, P., Song, J., Williams, C., Muddiman, David C., Sederoff, Ronald R., and Chiang, Vincent L. 2015. 4-Coumaroyl and caffeoyl shikimic acids inhibit 4-coumaric acid:coenzyme A ligases and modulate metabolic flux for 3-hydroxylation in monolignol biosynthesis of *Populus trichocarpa*. Molecular Plant **8**(1): 176-187.

Lin, Y.L., Chang, Y.Y., Kuo, Y.H., and Shiao, M.S. 2002. Anti-lipid-peroxidative principles from *Tournefortia sarmentosa*. Journal of Natural Products **65**(5): 745-747.

Luo, J., Fuell, C., Parr, A., Hill, L., Bailey, P., Elliott, K., Fairhurst, S.A., Martin, C., and Michael, A.J. 2009. A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in *Arabidopsis* seed. The Plant Cell **21**(1): 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, A.J., Saito, K., and 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**(4): 678-695.

Lyons, P.C., Wood, K.V., and Nicholson, R.L. 1990. Caffeoyl ester accumulation in corn leaves inoculated with fungal pathogens. Phytochemistry **29**(1): 97-101.

Ma, X., Koepke, J., Panjikar, S., Fritzsch, G., and Stöckigt, J. 2005. Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. Journal of Biological Chemistry **280**(14): 13576-13583.

Ma, X., Koepke, J., Bayer, A., Linhard, V., Fritzsch, G., Zhang, B., Michel, H., and Stöckigt, J. 2004. Vinorine synthase from *Rauvolfia*: the first example of crystallization and preliminary X-ray diffraction analysis of an enzyme of the BAHD superfamily. Biochimica et Biophysica Acta **1701**(1-2): 129-132.

Mader, M., Le Paslier, M.-C., Bounon, R., Bérard, A., Faivre-Rampant, P., Fladung, M., Leplé, J.-C., and Kersten, B. 2017. Whole-genome draft assembly of *Populus tremula* x *P. alba* clone INRA 717-1B4. Silvae Genetica **65**(2): 74.

Mahapatra, D.K., Bharti, S.K., and Asati, V. 2015. Chalcone scaffolds as anti-infective agents: structural and molecular target perspectives. European Journal of Medicinal Chemistry **101**: 496-524.

Manjasetty, B.a., Yu, X.H., Panjikar, S., Taguchi, G., Chance, M.R., and Liu, C.J. 2012. Structural basis for modification of flavonol and naphthol glucoconjugates by *Nicotiana tabacum* malonyltransferase (NtMaT1). Planta **236**(3): 781-793.

Marraffini, L.A., and Sontheimer, E.J. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature Reviews Genetics **11**(3): 181-190.

Masek, T., Vopalensky, V., Suchomelova, P., and Pospisek, M. 2005. Denaturing RNA electrophoresis in TAE agarose gels. Analytical Biochemistry **336**(1): 46-50.

Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayle, D., Bassard, J.-E., Pollet, B., Hehn, A., Heintz, D., and Ullmann, P. 2009. Evolution of a novel phenolic pathway for pollen development. Science **325**(5948): 1688-1692.

McCormick, S.P., Alexander, N.J., Trapp, S.E., and Hohn, T.M. 1999. Disruption of TRI101, the gene encoding trichothecene 3-O-acetyltransferase, from *Fusarium sporotrichioides*. Applied and Environmental Microbiology **65**(12): 5252-5256.

McKown, A.D., Guy, R.D., Klapste, J., Geraldes, A., Friedmann, M., Cronk, Q.C., El-Kassaby, Y.A., Mansfield, S.D., and Douglas, C.J. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. New Phytologist Journal **201**(4): 1263-1276.

Meilan, R., and Ma, C. 2007. Poplar (*Populus* spp.). *In* Agrobacterium Protocols. *Edited by* K. Wang. Springer-Verlag New York. pp. 143-151.

Michno, J.M., Wang, X., Liu, J., Curtin, S.J., Kono, T.J., and Stupar, R.M. 2015. CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. GM Crops & Food **6**(4): 243-252.

Milkowski, C., Baumert, A., Schmidt, D., Nehlin, L., and Strack, D. 2004. Molecular regulation of sinapate ester metabolism in *Brassica napus*: expression of genes, properties of the encoded proteins and correlation of enzyme activities with metabolite accumulation. The Plant Journal **38**(1): 80-92.

Mondolot, L., La Fisca, P., Buatois, B., Talansier, E., De Kochko, A., and Campa, C. 2006. Evolution in caffeoylquinic acid content and histolocalization during *Coffea canephora* leaf development. Annals of Botany **98**(1): 33-40.

Mueller, O., Lightfoot, S., and Schroeder, A. 2004. RNA Integrity Number (RIN) Standardization of RNA Quality Control. Agilent Technologies.

Mukherjee, D., Mukherjee, A., and Ghosh, T.C. 2016. Evolutionary rate heterogeneity of primary and secondary metabolic pathway genes in *Arabidopsis thaliana*. Genome Biology and Evolution **8**(1): 17-28.

Müller, T., and Vingron, M. 2000. Modeling amino acid replacement. Journal of Computational Biology **7**(6): 761-776.

Murashige, T., and Skoog, F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiologia Plantarum **15**.

Muroi, A., Ishihara, A., Tanaka, C., Ishizuka, A., Takabayashi, J., Miyoshi, H., and Nishioka, T. 2009. Accumulation of hydroxycinnamic acid amides induced by pathogen infection and identification of agmatine coumaroyltransferase in *Arabidopsis thaliana*. Planta **230**(3): 517-527.

Nam, K.H., Dudareva, N., and Pichersky, E. 1999. Characterization of benzylalcohol acetyltransferases in scented and non-scented *Clarkia* species. Plant and Cell Physiology **40**(9): 916-923.

Naoumkina, M., Farag, M.A., Sumner, L.W., Tang, Y., Liu, C.-J., and Dixon, R.A. 2007. Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proceedings of the National Academy of Sciences of the United States of America **104**(46): 17909-17915.

Narukawa, M., Kanbara, K., Tominaga, Y., Aitani, Y., Fukuda, K., Kodama, T., Murayama, N., Nara, Y., Arai, T., Konno, M., Kamisuki, S., Sugawara, F., Iwai, M., and Inoue, Y. 2009. Chlorogenic acid facilitates root hair formation in lettuce seedlings. Plant and Cell Physiology **50**(3): 504-514.

Natural Resources Canada. 2017. Climate impacts on productivity and health of aspen. Available from. [accessed December 13 2017].

Negrel, J., and Javelle, F. 1997. Purification, characterization and partial amino acid sequencing of hydroxycinnamoyl-CoA:tyramine N-(hydroxycinnamoyl)transferase from tobacco cell-suspension cultures. The FEBS Journal **247**: 1127-1135.

Negruk, V., Yang, P., Subramanian, M., McNevin, J.P., and Lemieux, B. 1996. Molecular cloning and characterization of the CER2 gene of *Arabidopsis thaliana*. The Plant Journal **9**(2): 137-145.

Nevarez, D.M., Mengistu, Y.a., Nawarathne, I.N., and Walker, K.D. 2009. An N-aroyltransferase of the bahd superfamily has broad aroyl CoA specificity in vitro with analogues of N-dearoylpaclitaxel. Journal of the American Chemical Society **131**(16): 5994-6002.

Niggeweg, R., Michael, A.J., and Martin, C. 2004. Engineering plants with increased levels of the antioxidant chlorogenic acid. Nature Biotechnology **22**(6): 746-754.

Noggle, J.F.T., Clark, C.R., and DeRuiter, J. 1991. Gas chromatographic and mass spectrometric analysis of samples from a clandestine laboratory involved in the synthesis of ecstacy from *Sassafras* oil. Journal of Chromatographic Science **29**(4): 168-173.

Nowakowska, Z. 2007. A review of anti-infective and anti-inflammatory chalcones. European Journal of Medicinal Chemistry **42**(2): 125-137.

O'Connell, A., Holt, K., Piquemal, J., Grima-Pettenati, J., Boudet, A., Pollet, B., Lapierre, C., Petit-Conil, M., Schuch, W., and Halpin, C. 2002. Improved paper pulp from plants with suppressed cinnamoyl-CoA reductase or cinnamyl alcohol dehydrogenase. Transgenic Research **11**(5): 495-503.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., and Pruitt, K.D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research **44**(D1): D733-745.

Okuda, T., Yoshida, T., and Hatano, T. 2000. Correlation of oxidative transformations of hydrolyzable tannins and plant evolution. Phytochemistry **55**(6): 513-529.

Olson, M.M., and Roseland, C.R. 1991. Induction of the coumarins scopoletin and ayapin in sunflower by insect-feeding stress and effects of coumarins on the feeding of sunflower beetle (Coleoptera: *Chrysomelidae*). Environmental Entomology **20**(4): 1166-1172.

Onkokesung, N., Gaquerel, E., Kotkar, H., Kaur, H., Baldwin, I.T., and Galis, I. 2012. MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A:polyamine transferases in *Nicotiana attenuata*. Plant Physiology **158**: 389-407.

Osakabe, N., Yasuda, A., Natsume, M., and Yoshikawa, T. 2004a. Rosmarinic acid inhibits epidermal inflammatory responses: anticarcinogenic effect of *Perilla frutescens* extract in the murine two-stage skin model. Carcinogenesis **25**(4): 549-557.

Osakabe, N., Takano, H., Sanbongi, C., Yasuda, A., Yanagisawa, R., Inoue, K., and Yoshikawa, T. 2004b. Anti-inflammatory and anti-allergic effect of rosmarinic acid (RA); inhibition of

seasonal allergic rhinoconjunctivitis (SAR) and its mechanism. BioFactors **21**(1-4): 127-131.

Pardo Torre, J.C., Schmidt, G.W., Paetz, C., Reichelt, M., Schneider, B., Gershenzon, J., and D'Auria, J.C. 2012. The biosynthesis of hydroxycinnamoyl quinate esters and their role in the storage of cocaine in *Erythroxylum coca*. Phytochemistry **91**: 177-186.

Pascal, S., Bernard, A., Sorel, M., Pervent, M., Vile, D., Haslam, R.P., Napier, J.a., Lessire, R., Domergue, F., and Joubès, J. 2013. The *Arabidopsis* cer26 mutant, like the cer2 mutant, is specifically affected in the very long chain fatty acid elongation process. The Plant Journal **73**(5): 733-746.

Payen, A. 1838. Mémoire sur la composition du tissu propre des plantes et du ligneux. Comptes Rendus Chimie **7**: 1052-1056.

Peng, X., Li, W., Wang, W., and Bai, G. 2010. Cloning and characterization of a cDNA coding a hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase involved in chlorogenic acid biosynthesis in *Lonicera japonica*. Planta Medica **76**(16): 1921-1926.

Petersen, M., Abdullah, Y., Benner, J., Eberle, D., Gehlen, K., Hücherig, S., Janiak, V., Kim, K.H., Sander, M., Weitzel, C., Wolters, S., and Hucherig, S. 2009. Evolution of rosmarinic acid biosynthesis. Phytochemistry **70**: 1663-1679.

Pfitzner, A., Polz, L., and Stöckigt, J. 1986. Properties of vinorine synthase - the *Rauwolfia* enzyme involved in the formation of the ajmaline skeleton. Zeitschrift für Naturforschung C **41**(1-2): 103-114.

Piquemal, J., Lapierre, C., Myton, K., O'connell, A., Schuch, W., Grima-pettenati, J., and Boudet, A.-M. 1998. Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. The Plant Journal **13**(1): 71-83.

Porth, I., Klapste, J., Skyba, O., Friedmann, M.C., Hannemann, J., Ehlting, J., El-Kassaby, Y.A., Mansfield, S.D., and Douglas, C.J. 2013. Network analysis reveals the relationship among wood properties, gene expression levels and genotypes of natural *Populus trichocarpa* accessions. The New Phytologist **200**(3): 727-742.

Power, R., Kurz, W.G., and De Luca, V. 1990. Purification and characterization of acetylcoenzyme A: deacetylvindoline 4-O-acetyltransferase from *Catharanthus roseus*. Archives of Biochemistry and Biophysics **279**: 370-376.

Price, M.N., Dehal, P.S., and Arkin, A.P. 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLOS One **5**(3): e9490.

Puchta, H. 2005. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. Journal of Experimental Botany **56**(409): 1-14.

Rani, S.H., Krishna, T.H.A., Saha, S., Negi, A.S., and Rajasekharan, R. 2010. Defective in cuticular ridges (DCR) of *Arabidopsis thaliana*, a gene associated with surface cutin formation, encodes a soluble diacylglycerol acyltransferase. Journal of Biological Chemistry **285**(49): 38337-38347.

Rautengarten, C., Ebert, B., Ouellet, M., Nafisi, M., Baidoo, E.E., Benke, P., Stranne, M., Mukhopadhyay, A., Keasling, J.D., Sakuragi, Y., and Scheller, H.V. 2012. *Arabidopsis* Deficient in Cutin ferulate encodes a transferase required for feruloylation of omega-hydroxy fatty acids in cutin polyester. Plant Physiology **158**: 654-665.

Renault, H., Alber, A., Horst, N.A., Basilio Lopes, A., Fich, E.A., Kriegshauser, L., Wiedemann, G., Ullmann, P., Herrgott, L., Erhardt, M., Pineau, E., Ehlting, J., Schmitt, M., Rose, J.K., Reski, R., and Werck-Reichhart, D. 2017. A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nature Communications **8**: 14713.

Ro, D.K., and Douglas, C.J. 2004. Reconstitution of the entry point of plant phenylpropanoid metabolism in yeast (*Saccharomyces cerevisiae*): implications for control of metabolic flux into the phenylpropanoid pathway. Journal of Biological Chemistry **279**(4): 2600-2607.

Robinson, A.R., and Mansfield, S.D. 2009. Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. The Plant Journal **58**(4): 706-714.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Current Biology **15**(14): 1325-1330.

Ron, M., Kajala, K., Pauluzzi, G., Wang, D., Reynoso, M.A., Zumstein, K., Garcha, J., Winte, S., Masson, H., Inagaki, S., Federici, F., Sinha, N., Deal, R.B., Bailey-Serres, J., and Brady, S.M. 2014. Hairy root transformation using *Agrobacterium rhizogenes* as a tool for exploring cell type-specific gene expression and function using tomato as a model. Plant Physiology **166**(2): 455-469.

Russell, D.W. 1971. The metabolism of aromatic compounds in higher plants X. Properties of the cinnamic acid 4-hydroxylase of pea seedlings and some aspects of its metabolic and developmental control. Journal of Biological Chemistry **246**(12): 3870-3878.

Salvador, V.H., Lima, R.B., dos Santos, W.D., Soares, A.R., Böhm, P.A.F., Marchiosi, R., Ferrarese, M.d.L.L., and Ferrarese-Filho, O. 2013. Cinnamic acid increases lignin production and inhibits soybean root growth. PLOS One **8**(7): e69105.

Samarut, E., Lissouba, A., and Drapeau, P. 2016. A simplified method for identifying early CRISPR-induced indels in zebrafish embryos using high resolution melting analysis. BMC Genomics **17**: 547.

Sanbongi, C., Takano, H., Osakabe, N., Sasa, N., Natsume, M., Yanagisawa, R., Inoue, K.I., Sadakane, K., Ichinose, T., and Yoshikawa, T. 2004. Rosmarinic acid in *Perilla* extract inhibits allergic inflammation induced by mite allergen, in a mouse model. Clinical & Experimental Allergy **34**(6): 971-977.

Sander, M., and Petersen, M. 2011. Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:Shikimate hydroxycinnamoyl-transferase, from *Coleus blumei* Benth. Planta **233**(6): 1157-1171.

Sanderson, M.J., Thorne, J.L., Wikstrom, N., and Bremer, K. 2004. Molecular evidence on plant divergence times. American Journal of Botany **91**(10): 1656-1665.

Scarpati, M.L., and Oriente, G. 1958. Chicoric acid (dicaffeyltartic acid): Its isolation from chicory (*Chicorium intybus*) and synthesis. Tetrahedron **4**(1-2): 43-48.

Schanz, S., Schröder, G., and Schröder, J. 1992. Stilbene synthase from Scots pine (*Pinus sylvestris*). FEBS Letters **313**(1): 71-74.

Schilmiller, A.L., Charbonneau, A.L., and Last, R.L. 2012. Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proceedings of the National Academy of Sciences of the United States of America **109**(40): 16377-16382.

Serra, O., Hohn, C., Franke, R., Prat, S., Molinas, M., and Figueras, M. 2010. A feruloyl transferase involved in the biosynthesis of suberin and suberin-associated wax is required for maturation and sealing properties of potato periderm. The Plant Journal **62**(2): 277-290.

Shadle, G., Chen, F., Srinivasa Reddy, M.S., Jackson, L., Nakashima, J., and Dixon, R.a. 2007. Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. Phytochemistry **68**(14): 1521-1529.

Shalit, M., Guterman, I., Volpin, H., Bar, E., Tamari, T., Menda, N., Adam, Z., Zamir, D., Vainstein, A., Weiss, D., Pichersky, E., and Lewinsohn, E. 2003. Volatile ester formation in roses. Identification of an acetyl-coenzyme A. Geraniol/citronellol acetyltransferase in developing rose petals. Plant Physiology **131**(4): 1868-1876.

Shan, W., Zhao, C., Fan, J., Cong, H., Liang, S., and Yu, X. 2012. Antisense suppression of alcohol acetyltransferase gene in ripening melon fruit alters volatile composition. Scientia Horticulturae **139**: 96-101.

Shay, P.-E. 2017. The effects of condensed tannins, nitrogen and climate on decay, nitrogen mineralisation and microbial communities in forest tree leaf litter. Doctor of Philosophy Ph.D., Department of Biology, Unviersity of Victoria, Victoria

Shi, J., Gao, H., Wang, H., Lafitte, H.R., Archibald, R.L., Yang, M., Hakimi, S.M., Mo, H., and Habben, J.E. 2017. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnology Journal **15**(2): 207-216.

Shimizu, B. 2014. 2-Oxoglutarate-dependent dioxygenases in the biosynthesis of simple coumarins. Frontiers in Plant Science **5**: 549.

Shirley, A.M., and Chapple, C. 2003. Biochemical characterization of sinapoylglucose:choline sinapoyltransferase, a serine carboxypeptidase-like protein that functions as an acyltransferase in plant secondary metabolism. Journal of Biological Chemistry **278**(22): 19870-19877.

Shirley, A.M., McMichael, C.M., and Chapple, C. 2001. The sng2 mutant of *Arabidopsis* is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. The Plant Journal **28**(1): 83-94.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology **7**: 539.

Skyba, O., Douglas, C.J., and Mansfield, S.D. 2013. Syringyl-rich lignin renders poplars more resistant to degradation by wood decay fungi. Applied and Environmental Microbiology **79**(8): 2560-2571.

Sondheimer, E. 1958. On the distribution of caffeic acid and the chlorogenic acid isomers in plants. Archives of Biochemistry and Biophysics **74**(1): 131-138.

Sonnante, G., D'Amore, R., Blanco, E., Pierri, C.L., De Palma, M., Luo, J., Tucci, M., and Martin, C. 2010. Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiology **153**(3): 1224-1238.

Souleyre, E.J., Greenwood, D.R., Friel, E.N., Karunairetnam, S., and Newcomb, R.D. 2005. An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. The FEBS Journal **272**: 3132-3144.

Stewart, J.J., Akiyama, T., Chapple, C., Ralph, J., and Mansfield, S.D. 2009. The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. Plant Physiology **150**(2): 621-635.

Studer, M.H., DeMartini, J.D., Davis, M.F., Sykes, R.W., Davison, B., Keller, M., Tuskan, G.A., and Wyman, C.E. 2011. Lignin content in natural *Populus* variants affects sugar release. Proceedings of the National Academy of Sciences of the United States of America **108**(15): 6300-6305.

Sullivan, M.L. 2008. Multiple hydroxycinnamoyl transferases from red clover differ in sequence, expression pattern, and enzymatic activity. *In* 41st North American Alfalfa Improvement Conference. US Dairy Forage Research Center, Irving, Texas. pp. 53706-53706.

Sullivan, M.L. 2009. A novel red clover hydroxycinnamoyl transferase has enzymatic activities consistent with a role in phaselic acid biosynthesis. Plant Physiology **150**(4): 1866-1879.

Sullivan, M.L., and Zarnowski, R. 2011. Red clover HCT2, a hydroxycinnamoyl-coenzyme A:malate hydroxycinnamoyl transferase, plays a crucial role in biosynthesis of phaselic acid and other hydroxycinnamoyl-malate esters *in vivo*. Plant Physiology **155**(3): 1060-1067.

Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., and Xi, Y. 2015. Targeted mutagenesis in soybean using the CRISPR-Cas9 system. Scientific Reports **5**: 10342.

Suzuki, H., Nakayama, T., and 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**(6): 1764-1771.

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

Suzuki, H., Murakoshi, I., Saito, K., and Lupinus, I.N. 1994. A novel O-tigloyltransferase for alkaloid biosynthesis in plants. Purification, characterization, and distribution in *Lupinus* plants. Journal of Biological Chemistry **269**: 15853-15860.

Suzuki, H., Nakayama, T., Yamaguchi, M.A., and Nishino, T. 2004a. cDNA cloning and characterization of two *Dendranthema x morifolium* anthocyanin malonyltransferases with different functional activities. Plant Science **166**(1): 89-96.

Suzuki, H., Nakayama, T., Nagae, S., Yamaguchi, M.A., Iwashita, T., Fukui, Y., and Nishino, T. 2004b. cDNA cloning and functional characterization of flavonol 3-O-glucoside-6″-O-malonyltransferases from flowers of *Verbena hybrida* and *Lamium purpureum*. Journal of Molecular Catalysis B **28**: 87-93.

Suzuki, H., Sawada, S.y., Watanabe, K., Nagae, S., Yamaguchi, M.-A., Nakayama, T., and Nishino, T. 2004c. Identification and characterization of a novel anthocyanin malonyltransferase from scarlet sage (*Salvia splendens*) flowers: an enzyme that is phylogenetically separated from other anthocyanin acyltransferases. The Plant Journal **38**(6): 994-1003.

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Yamaguchi, M.-A., Tanaka, Y., Kusumi, T., and Nishino, T. 2002. cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A:anthocyanidin 3-o-glucoside-6"-o-malonyltransferase from *Dahlia* flowers. Plant Physiology **130**(4): 2142-2151.

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Nakao, M., Tanaka, Y., Yamaguchill, M.A., Kusumi, T., and Nishino, T. 2001. Malonyl-CoA:anthocyanin 5-O-glucoside-6"'-O-malonyltransferase from scarlet sage (*Salvia splendens*) flowers. Enzyme purification, gene cloning, expression, and characterization. Journal of Biological Chemistry **276**(52): 49013-49019.

Swarup, V., Ghosh, J., Ghosh, S., Saxena, A., and Basu, A. 2007. Antiviral and anti-inflammatory effects of rosmarinic acid in an experimental murine model of Japanese encephalitis. Antimicrobial Agents and Chemotherapy **51**(9): 3367-3370.

Tacke, E., Korfhage, C., Michel, D., Maddaloni, M., Motto, M., Lanzini, S., Salamini, F., and Doring, H.P. 1995. Transposon tagging of the maize Glossy2 locus with the transposable element En/Spm. The Plant Journal **8**(6): 907-917.

Taguchi, G., Shitchi, Y., Shirasawa, S., Yamamoto, H., and Hayashida, N. 2005. Molecular cloning, characterization, and downregulation of an acyltransferase that catalyzes the malonylation of flavonoid and naphthol glucosides in tobacco cells. The Plant Journal **42**(4): 481-491.

Taguchi, G., Ubukata, T., Nozue, H., Kobayashi, Y., Takahi, M., Yamamoto, H., and Hayashida, N. 2010. Malonylation is a key reaction in the metabolism of xenobiotic phenolic glucosides in *Arabidopsis* and tobacco. The Plant Journal **63**(6): 1031-1041.

Tang, F., Yang, S., Liu, J., and Zhu, H. 2016. Rj4, a gene controlling nodulation specificity in soybeans, encodes a thaumatin-like protein but not the one previously reported. Plant Physiology **170**(1): 26-32.

Tang, H. 2015. Characterization of a putative flavonoid 3', 5' Hydroxylase (PtF3'5'H1) in *Populus*. Biology, Universtiy of Victoria, Victoria

The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Research **45**(D1): D158-D169.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal **37**(6): 914-939.

Tokai, T., Fujimura, M., Inoue, H., Aoki, T., Ohta, K., Shibata, T., Yamaguchi, I., and Kimura, M. 2005. Concordant evolution of trichothecene 3-O-acetyltransferase and an rDNA species phylogeny of trichothecene-producing and non-producing fusaria and other ascomycetous fungi. Microbiology **151**(Pt 2): 509-519.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols **7**(3): 562-578.

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

Tsai, C.J., Harding, S.a., Tschaplinski, T.J., Lindroth, R.L., and Yuan, Y. 2006. Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. New Phytologist Journal **172**(1): 47-62.

Tsuchiyama, M., Sakamoto, T., Tanimori, S., Murata, S., and Kawasaki, H. 2007. Enzymatic synthesis of hydroxycinnamic acid glycerol esters using type A feruloyl esterase from *Aspergillus niger*. Bioscience, Biotechnology, and Biochemistry **71**(10): 2606-2609.

Tuominen, L.K., Johnson, V.E., and Tsai, C.-J. 2011. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. BMC Genomics **12**(1): 236-236.

Tuskan, G.a., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, a., Schein, J., Sterck, L., Aerts, a., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, a., Brunner, a., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, a., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, a., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, a., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, a., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, a., Sterky, F., Terry, a., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y., and Rokhsar, D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**(5793): 1596-1604.

Tzagoloff, A. 1963. Metabolism of sinapine in mustard plants. I. Degradation of sinapine into sinapic acid & choline. Plant Physiology **38**(2): 202-206.

Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., Nishino, T., Kusunoki, M., and Nakayama, T. 2007. Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. Journal of Biological Chemistry **282**(21): 15812-15822.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. 2012. Primer3--new capabilities and interfaces. Nucleic Acids Research **40**(15): e115.

Valinas, M.A., Lanteri, M.L., ten Have, A., and Andreu, A.B. 2015. Chlorogenic acid biosynthesis appears linked with suberin production in potato tuber (*Solanum tuberosum*). Journal of Agricultural and Food Chemistry **63**(19): 4902-4913.

Van Doorsselaere, J., Baucher, M., Chognot, E., Chabbert, B., Tollier, M.-T., Petit-Conil, M., Leplé, J.-C., Pilate, G., Cornu, D., Monties, B., Van Montagu, M., Inzé, D., Boerjan, W., and Jouanin, L. 1995. A novel lignin in poplar trees with a reduced caffeic acid/5-hydroxyferulic acid O-methyltransferase activity. The Plant Journal **8**(6): 855-864.

Vanholme, B., Cesarino, I., Goeminne, G., Kim, H., Marroni, F., Van Acker, R., Vanholme, R., Morreel, K., Ivens, B., Pinosio, S., Morgante, M., Ralph, J., Bastien, C., and Boerjan, W. 2013a. Breeding with rare defective alleles (BRDA): A natural *Populus nigra* HCT mutant with modified lignin as a case study. New Phytologist Journal **198**: 765-776.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. 2010. Lignin Biosynthesis and Structure. Plant Physiology **153**(3): 895-905.

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., Kim, H., Cross, J., Morreel, K., Araujo, P., Welsh, L., Haustraete, J., McClellan, C., Vanholme, B., Ralph, J., Simpson, G.G., Halpin, C., and Boerjan, W. 2013b. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. Science **341**(6150): 1103-1106.

Vanneste, K., Maere, S., and Van de Peer, Y. 2014. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philosophical Transactions of the Royal Society of London B **369**(1648).

Vanzo, A., Cecotti, R., Vrhovsek, U., Torres, A.M., Mattivi, F., and Passamonti, S. 2007. The fate of trans-caftaric acid administered into the rat stomach. Journal of Agricultural and Food Chemistry **55**(4): 1604-1611.

Varbanova, M., Porter, K., Lu, F., Ralph, J., Hammerschmidt, R., Jones, a.D., and Day, B. 2011. Molecular and biochemical basis for stress-induced accumulation of free and bound *p*-coumaraldehyde in cucumber. Plant Physiology **157**(3): 1056-1066.

Vargas, L., Cesarino, I., Vanholme, R., Voorend, W., de Lyra Soriano Saleme, M., Morreel, K., and Boerjan, W. 2016. Improving total saccharification yield of *Arabidopsis* plants by vessel-specific complementation of caffeoyl shikimate esterase (CSE) mutants. Biotechnology for Biofuels **9**: 139.

Vlot, A.C., Dempsey, D.A., and Klessig, D.F. 2009. Salicylic acid, a multifaceted hormone to combat disease. Annual Review of Phytopathology **47**: 177-206.

Vogt, T. 2010. Phenylpropanoid biosynthesis. Molecular Plant **3**(1): 2-20.

Volkmann, D., and Baluska, F. 2006. Gravity: one of the driving forces for evolution. Protoplasma **229**(2-4): 143-148.

Von Roepenack-Lahaye, E., Newman, M.-A.A., Schornack, S., Hammond-Kosack, K.E., Lahaye, T., Jones, J.D.G., Daniels, M.J., and Dow, J.M. 2003. *p*-Coumaroylnoradrenaline, a novel plant metabolite implicated in tomato defense against pathogens. Journal of Biological Chemistry **278**: 43373-43383.

Wagner, A., Ralph, J., Akiyama, T., Flint, H., Phillips, L., Torr, K., Nanayakkara, B., and Te Kiri, L. 2007. Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase in *Pinus radiata*. Proceedings of the National Academy of Sciences of the United States of America **104**(28): 11856-11861.

Walker, K., and Croteau, R. 2000a. Taxol biosynthesis: molecular cloning of a benzoyl-CoA:taxane 2alpha-O-benzoyltransferase cDNA from *Taxus* and functional expression in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **97**(25): 13591-13596.

Walker, K., and Croteau, R. 2000b. Molecular cloning of a 10-deacetylbaccatin III-10-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **97**(2): 583-587.

Walker, K., Schoendorf, a., and Croteau, R. 2000. Molecular cloning of a taxa-4(20),11(12)-dien-5alpha-ol-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. Archives of Biochemistry and Biophysics **374**(2): 371-380.

Walker, K., Fujisaki, S., Long, R., and Croteau, R. 2002. Molecular cloning and heterologous expression of the C-13 phenylpropanoid side chain-CoA acyltransferase that functions in Taxol biosynthesis. Proceedings of the National Academy of Sciences of the United States of America **99**(20): 12715-12720.

Waltz, E. 2015. Nonbrowning GM apple cleared for market. Nature Biotechnology **33**(4): 326-327.

Wang, J., and Luca, V.D. 2005. The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. The Plant Journal **44**(4): 606-619.

Wang, J.P., Naik, P.P., Chen, H.-C., Shi, R., Lin, C.-Y., Liu, J., Shuford, C.M., Li, Q., Sun, Y.-H., Tunlaya-Anukit, S., Williams, C.M., Muddiman, D.C., Ducoste, J.J., Sederoff, R.R., and Chiang, V.L. 2014. Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in *Populus trichocarpa*. The Plant Cell **26**(3): 894-914.

Wang, L., Wang, L., Tan, Q., Fan, Q., Zhu, H., Hong, Z., Zhang, Z., and Duanmu, D. 2016. Efficient inactivation of symbiotic nitrogen fixation related genes in *Lotus japonicus* using CRISPR-Cas9. Frontiers in Plant Science **7**: 1333.

Wang, Q., Yang, S., Liu, J., Terecskei, K., Ábrahám, E., Gombár, A., Domonkos, Á., Szűcs, A., Körmöczi, P., and Wang, T. 2017. Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. Proceedings of the National Academy of Sciences of the United States of America.

Wang, W., Tang, K., Yang, H.-R., Wen, P.-F., Zhang, P., Wang, H.-L., and Huang, W.-D. 2010. Distribution of resveratrol and stilbene synthase in young grape plants (*Vitis vinifera* L. cv. Cabernet Sauvignon) and the effect of UV-C on its accumulation. Plant Physiology and Biochemistry **48**(2): 142-152.

Waters, E.R. 2003. Molecular adaptation and the origin of land plants. Molecular Phylogenetics and Evolution **29**(3): 456-463.

Weier, D., Mittasch, J., Strack, D., and Milkowski, C. 2008. The genes BnSCT1 and BnSCT2 from *Brassica napus* encoding the final enzyme of sinapine biosynthesis: molecular characterization and suppression. Planta **227**(2): 375-385.

Weitzel, C., and Petersen, M. 2011. Cloning and characterisation of rosmarinic acid synthase from *Melissa officinalis* L. Phytochemistry **72**(7): 572-578.

Wesley, S.V., Helliwell, C., Wang, M.-B., and Waterhouse, P. 2004. Posttranscriptional gene silencing in plants. *In* RNA Interference, Editing, and Modification: Methods and Protocols. *Edited by* J.M. Gott, Totowa, NJ. pp. 117-129.

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., Robinson, S.P., Gleave, A.P., Greenand, A.G., and Waterhouse, P.M. 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. The Plant Journal **27**(6): 581-590.

Whatley, F.R., Greenaway, W., and May, J. 1989. *Populus candicans* and the Balm of Gilead. Zeitschrift für Naturforschung C **44**(5-6): 353-356.

Whitlon, D.S., Sadowski, J.A., and Suttie, J.W. 1978. Mechanism of coumarin action: significance of vitamin K epoxide reductase inhibition. Biochemistry **17**(8): 1371-1377.

Wink, M. 2010. Biochemistry, physiology and ecological functions of secondary metabolites. *In* Annual Plant Reviews. Wiley-Blackwell. pp. 1-19.

Winkel-Shirley, B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology **126**(2): 485-493.

Withers, S., Lu, F., Kim, H., Zhu, Y., Ralph, J., and Wilkerson, C.G. 2012. Identification of grass-specific enzyme that acylates monolignols with *p*-coumarate. Journal of Biological Chemistry **287**(11): 8347-8355.

World Health Organization. 2015. 19th WHO model list of essential medicines. *Edited by* Expert Committee on the Selection and Use of Essential Medicines, Geneva.

Xie, C.-Y., Ying, C.C., Yanchuk, A.D., and Holowachuk, D.L. 2009. Ecotypic mode of regional differentiation caused by restricted gene migration: a case in black cottonwood (*Populus trichocarpa*) along the Pacific Northwest coast. Canadian Journal of Forest Research **39**(3): 519-526.

Xue, L.-J., Alabady, M.S., Mohebbi, M., and Tsai, C.-J. 2015. Exploiting genome variation to improve next-generation sequencing data analysis and genome editing efficiency in *Populus tremula* × *alba* 717-1B4. Tree Genetics & Genomes **11**(4): 82.

Yahyaoui, F.E., Wongs-Aree, C., Latche, A., Hackett, R., Grierson, D., and Pech, J.C. 2002. Molecular and biochemical characteristics of a gene encoding an alcohol acyl-transferase involved in the generation of aroma volatile esters during melon ripening. The FEBS Journal **269**(9): 2359-2366.

Yang, Q., Reinhard, K., Schiltz, E., and Matern, U. 1997. Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. Plant Molecular Biology **35**: 777-789.

Yang, Q., Trinh, H.X., Imai, S., Ishihara, A., Zhang, L., Nakayashiki, H., Tosa, Y., and Mayama, S. 2004. Analysis of the involvement of hydroxyanthranilate hydroxycinnamoyltransferase and caffeoyl-CoA 3-O-methyltransferase in phytoalexin biosynthesis in oat. Molecular Plant-Microbe Interactions Journal **17**(1): 81-89.

Yonekura-Sakakibara, K., Tanaka, Y., Fukuchi-Mizutani, M., Fujiwara, H., Fukui, Y., Ashikari, T., Murakami, Y., Yamaguchi, M., and Kusumi, T. 2000. Molecular and biochemical characterization of a novel hydroxycinnamoyl-CoA: anthocyanin 3-O-glucoside-6"-O-acyltransferase from *Perilla frutescens*. Plant and Cell Physiology **41**(4): 495-502.

Yoshida, K., Ma, D., and Constabel, C.P. 2015. The MYB182 protein down-regulates proanthocyanidin and anthocyanin biosynthesis in poplar by repressing both structural and regulatory flavonoid genes. Plant Physiology **167**(3): 693-710.

Yoshihara, N., Imayama, T., Matsuo, Y., Fukuchi-Mizutani, M., Tanaka, Y., Ino, I., and Yabuya, T. 2006. Characterization of cDNA clones encoding anthocyanin 3-*p*-coumaroyltransferase from *Iris hollandica*. Plant Science **171**(5): 632-639.

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

Zhao, J., Huhman, D., Shadle, G., He, X.Z., Sumner, L.W., Tang, Y., and Dixon, R.A. 2011. MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. The Plant Cell **23**: 1536-1555.

Zhong, R., Richardson, E.A., and Ye, Z.-H. 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. The Plant Cell **19**(9): 2776-2792.