

# Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*

Chung-Jui Tsai<sup>1</sup>, Scott A. Harding<sup>1</sup>, Timothy J. Tschaplinski<sup>2</sup>, Richard L. Lindroth<sup>3</sup> and Yinan Yuan<sup>1</sup>

<sup>1</sup>Biotechnology Research Center, School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI 49931, USA;

<sup>2</sup>Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA; <sup>3</sup>Department of Entomology, University of Wisconsin, Madison, WI 53706, USA

## Summary

Author for correspondence:

Chung-Jui Tsai

Tel. +1 906 4872914

Fax: +1 906 4872915

Email: [chtsai@mtu.edu](mailto:chtsai@mtu.edu)

Received: 4 April 2006

Accepted: 28 April 2006

- Salicin-based phenolic glycosides, hydroxycinnamate derivatives and flavonoid-derived condensed tannins comprise up to one-third of *Populus* leaf dry mass. Genes regulating the abundance and chemical diversity of these substances have not been comprehensively analysed in tree species exhibiting this metabolically demanding level of phenolic metabolism.

- Here, shikimate-phenylpropanoid pathway genes thought to give rise to these phenolic products were annotated from the *Populus* genome, their expression assessed by semiquantitative or quantitative reverse transcription polymerase chain reaction (PCR), and metabolic evidence for function presented.

- Unlike *Arabidopsis*, *Populus* leaves accumulate an array of hydroxycinnamoyl-quininate esters, which is consistent with broadened function of the expanded hydroxycinnamoyl-CoA transferase gene family. Greater flavonoid pathway diversity is also represented, and flavonoid gene families are larger. Consistent with expanded pathway function, most of these genes were upregulated during wound-stimulated condensed tannin synthesis in leaves.

- The suite of *Populus* genes regulating phenylpropanoid product accumulation should have important application in managing phenolic carbon pools in relation to climate change and global carbon cycling.

**Key words:** phenylpropanoid metabolism, condensed tannins, phenolic glycosides, flavonoid biosynthesis, hydroxycinnamate derivatives, *Populus* genome.

*New Phytologist* (2006) **172**: 47–62

© The Authors (2006). Journal compilation © *New Phytologist* (2006)

doi: 10.1111/j.1469-8137.2006.01798.x

## Introduction

*Populus* species (cottonwoods, aspens and poplars) comprise the largest fraction of intensively managed hardwood forest acreage in North America (Coyle *et al.*, 2005). *Populus* species also shape the ecology and productivity of unmanaged habitats (Whitham *et al.*, 1996). An important determinant of *Populus* productivity is its secondary metabolism. Shikimate/phenylpropanoid-derived phenolic glycosides (PGs), hydroxycinnamate derivatives (HCDs) and condensed tannins (CTs) exhibit a wide range of concentrations in *Populus* tissues, with PGs and CTs alone comprising up to 35% leaf dry weight (Lindroth & Hwang, 1996). By their abundance and chemical diversity, PGs, HCDs

and CTs impact not only tree growth, but also the persistence of shed organs and thus the long-term retention of fixed carbon in forest ecosystems. The release of the *Populus* genome provides an opportunity to analyse the control of phenolic partitioning in a model system characterized by its highly plastic metabolic commitment to such sinks.

Numerous structural forms of salicylate-derived PGs have been identified, with some (e.g. salicin and salicortin) occurring widely throughout the genus and others (e.g. grandidentatin and HCH-salicortin) showing a narrower taxonomic distribution (R.L. Lindroth, unpublished). Phenolic glycosides act as deterrents to a variety of insect herbivores, and provide protection against pathogens and abiotic stress, such as UV radiation

(Lindroth & Hwang, 1996). Different structural forms of PGs vary in biological activity. For example, those containing a hydroxy-cyclohexen-*on*-oyl moiety are more toxic to insects than are related compounds lacking the functional group (Lindroth *et al.*, 1988).

The HCDs represent a major class of low molecular weight phenylpropanoid derivatives, constituting 2–8% of leaf dry weight in *Populus* and closely related taxa, including *Salix* (willow) and *Betula* (birch), with similar phenolic profiles (Ikonen *et al.*, 2001; Peltonen *et al.*, 2005; T. J. Tschaplinski, unpublished). The quantitatively significant HCDs, largely chlorogenic acid isomers and coumaryl-shikimate/quinic esters, have roles in the formation of protein adducts for defense, as radical scavengers for the dissipation of excess photoenergy, and as intermediates in lignin biosynthesis (Grace & Logan 2000; Boerjan *et al.*, 2003). Species-specific accumulation of HCDs among *Populus* has led to their use in chemical taxonomy (Greenaway *et al.*, 1991a, 1992). In monocot species, HCDs are thought to play important roles in cell wall cross-linking (Iiyama *et al.*, 1994). Some HCDs exhibit antimicrobial, antifungal and antioxidant activities (Ravn *et al.*, 1989), although their functions in *Populus* defense have not been investigated.

Condensed tannins are polymers formed from proanthocyanidins of varying chemical structure determined by flavonoid pathway steps (Bavage *et al.*, 1997; Marles *et al.*, 2003). *Populus* CTs exhibit a broad range of compositions and biological activities that are not observed in *Arabidopsis*, and that probably reflect differences in the complexity or regulation of flavonoid pathways between these two species (Ayres *et al.*, 1997; Winkel-Shirley, 2001). They affect forage and litter digestibility, as well as mineralization of soil nitrogen and carbon in *Populus* habitats (Schimel *et al.*, 1996; Schweitzer *et al.*, 2004). Biological activity depends largely on the ratio of prodelfinidin (PD) and procyanidin (PC) subunits, degree of polymerization and configuration of the polymers (Zucker, 1983; Ayres *et al.*, 1997; Behrens *et al.*, 2003).

The shikimate-phenylpropanoid pathways controlling PG, HCD and CT synthesis also generate intermediates with roles in developmental signaling and phytohormone homeostasis (Tamagnone *et al.*, 1998; Taylor & Grotewold, 2005). For example, flavonoids are modulators of polar auxin transport, central to vascular development and plant form (Peer *et al.*, 2004), and HCDs are known growth regulators (Tamagnone *et al.*, 1998). Here, we present an annotation of more than 100 *Populus* shikimate-phenylpropanoid pathway genes along with expression data for over 50 of these. A comparative analysis is made with *Arabidopsis*, which exhibits a quantitatively more conservative commitment to phenolic metabolism suited to its herbaceous annual growth habit. *Populus* therefore presents a highly relevant system for modeling carbon partitioning in large-biomass woody perennials that colonize and shape their habitat for long-term success.

## Materials and Methods

### Populus gene annotation and phylogenetic analysis

*Populus* genes described in this study were identified using a combination of approaches and manually annotated. Candidate genes were first identified based on computational gene-calling models implemented in the *Populus* genome portal (v1.1) hosted at the Joint Genome Institute (JGI), and/or by BLASTP search of the JGI *Populus* protein database using reference sequences from *Arabidopsis* or other herbaceous model species (Table S1). The resulting *Populus* sequences and gene models were inspected by conducting multiple protein sequence alignment with similar sequences from other species available in GenBank, and by preliminary phylogenetic analysis for identification of relevant members in a given gene family. In some cases, gene models were further edited to correct reading frame predictions (e.g. translation start or stop sites, or exon–intron junction sites). Nucleotide sequence alignment with available *Populus* cDNAs or expressed sequence tags (ESTs) were also used whenever possible. For the purpose of this study, only gene models predicted to contain full-length proteins were included (Table S1).

Phylogenetic analysis of each protein family was conducted using *Populus*, *Arabidopsis* and rice (TIGR Rice Annotation Release 4) sequences and, whenever possible, homologs from other species that had been biochemically or genetically characterized to guide inference of phylogenetic relationships. Full-length amino acid sequences were aligned by CLUSTALW version 1.82 (Chenna *et al.*, 2003) and imported into the Molecular Evolutionary Genetics Analysis (MEGA) package version 3.0 (Kumar *et al.*, 2004). Phylogenetic analyses were conducted using the neighbor-joining (NJ) method implemented in MEGA, with the pairwise deletion option for handling alignment gaps, and with the Poisson correction model for distance computation. Bootstrap tests were conducted using 1000 replicates.

### Reverse transcription polymerase chain reaction (RT-PCR) analysis

Total RNA was extracted from various tissues of *Populus fremontii* × *angustifolia* using the cetyltrimethylammonium bromide (CTAB) method (Chang *et al.*, 1993). Semi-quantitative RT-PCR analysis of genes involved in PG and HCD biosynthesis was conducted according to Ranjan *et al.* (2004), using newly emerged (young) leaves, expanding leaves, primary- and secondary-growth stems, root tips and xylem. The PCR amplification varied between 25 and 33 cycles of denaturation at 94°C for 15 s, annealing at 56–69°C for 20 s, and extension at 72°C for 1–2 min, using gene-specific primers (Table S2). Analysis of flavonoid pathway genes was similarly performed using young leaves and root tips.

Relative transcript abundance of flavonoid pathway genes in expanding sink leaves (leaf plastochron index LPI 0–1) of

control and wounded plants was analysed by quantitative RT-PCR. Wounding was accomplished by pinching source leaves, LPI 6-11, with a pair of pliers. Five days after wounding, leaves were harvested into liquid nitrogen for PG and CT analysis, according to Harding *et al.* (2005), and RNA extraction. Real-time RT-PCR amplification was performed with cDNA equivalent to 2.5 ng of DNase-treated total RNA, using the Brilliant SYBR Green QPCR Core Reagent Kit and the Mx3000P Real-Time PCR system (Stratagene, La Jolla, CA, USA). Each reaction was performed in duplicate and with three biological replicates along with no-template controls. The PCR parameters were 10 min at 95°C, and 40 cycles of 30 s at 95°C, 1 min at 55°C and 1 min at 72°C. Specificity of the amplification was assessed by dissociation curve analysis at the end of each run using the MxPro software (Stratagene). The threshold level was manually adjusted and applied consistently across all amplifications to obtain the threshold cycle ( $C_T$ ) values. Relative target transcript levels normalized to the geometric mean of 4 housekeeping genes (actin, ubiquitin, elongation factor 1-beta, and alpha-tubulin) were determined using the following equation:

Relative transcript abundance of gene of interest =  $2^{-\Delta C_T}$

$$(\Delta C_T = C_{T_{\text{gene}}} - C_{T_{\text{geomean}}}).$$

### Analysis of hydroxycinnamoyl-quininate esters in various *Populus* spp.

Leaf samples were fast frozen on dry ice, lyophilized, and then ground in a micro-Wiley mill (1 mm mesh size). Approximately 50 mg of leaf tissue (dry weight) were twice extracted with 5 ml 80% ethanol overnight and then combined before drying a 3-ml aliquot in a nitrogen stream. For generation of trimethylsilyl (TMS) derivatives (Gebre & Tschaplinski, 2002), dried extracts were dissolved in 500  $\mu$ l silylation-grade acetonitrile followed by 500  $\mu$ l *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS), and heated for 1 h at 70°C, with samples injected after 24 h into an HP 5890 Series II gas chromatograph coupled to an HP 5972 mass spectrometer (Hewlett-Packard, Avondale PA, USA), fitted with an Rtx-5MS (crosslinked 5% PH ME Siloxane) 30 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m film thickness capillary column (Restek). Gas chromatography–mass spectrometry (GC-MS) was operated in electron impact (70 eV) mode, with 1.5 full spectra (50–550 Da) scans per second. Gas (helium) flow was set at 0.6 ml min<sup>-1</sup> with the injection port configured in the splitless mode. The injection port and detector temperatures were set to 250°C and 300°C, respectively. The initial oven temperature was held at 100°C for 4 min and was programmed to increase at 8°C min<sup>-1</sup> to 300°C where it was held for a further 21 min, before cycling to the initial conditions. Peaks were identified based on a user-created database of TMS metabolites.

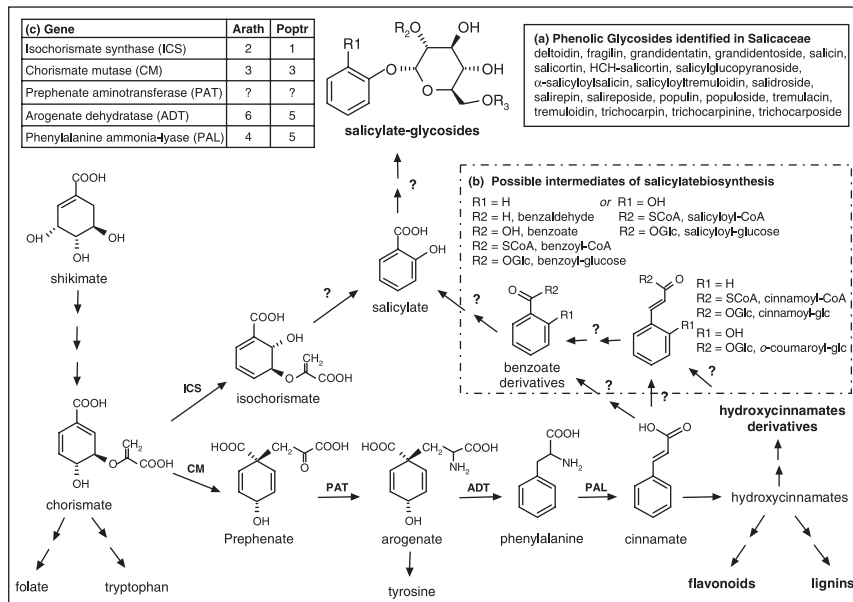
## Results

### Salicylate-glycoside pathway

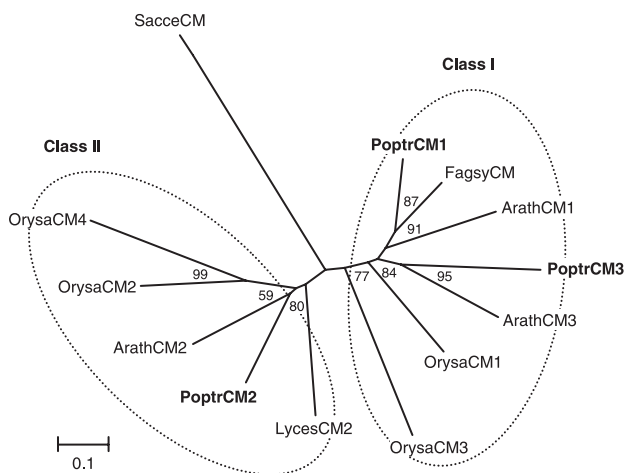
The major phenolic glycosides of *Populus* contain esterified salicin, likely derived from precursors of salicylic acid (SA) biosynthesis (Zenk, 1967; Pierpoint, 1994). Salicylic acid is synthesized from chorismate via two alternative, partially elucidated pathways (Fig. 1), one, a phenylalanine ammonia-lyase (PAL)-dependent phenylpropanoid route (reviewed in Métraux, 2002), the other, an isochorismate synthase (ICS)-dependent route (Wildermuth *et al.*, 2001). Early steps in the PAL-dependent pathway are thought to be mediated by chorismate mutase (CM), an uncharacterized prephenate aminotransferase (PAT) and arogenate dehydratase (ADT). There are no reports on roles for any of these steps in salicin or PG synthesis. This may be due at least partly to the low abundance of SA and SA-derived molecules in *Arabidopsis* and tobacco model systems

The *CM* gene family structure and predicted protein subcellular localization (two plastidic and one cytosolic) are similar in *Populus* and *Arabidopsis* (Fig. 2). Strong expression of *PoptrCM1* in developing xylem (Fig. 1d) and upregulation of its homolog *ArathCM1* during *Arabidopsis* stem development (Ehling *et al.*, 2005), suggests functional conservation as well. Interestingly, *PoptrCM2* was not detected in xylem and thus may have a more limited or specialized role in the synthesis of nonstructural phenolics in other tissues. The cytosolic *PoptrCM3* was absent in leaf tissue where PGs accumulate (Fig. 1d). Arogenate dehydratase catalyses the final step in phenylalanine biosynthesis, and is represented by six putative *ADT* genes in the *Arabidopsis* genome and five in *Populus* (Fig. 3). Only three of the *Arabidopsis* genes (*ArathADT3*, *ArathADT5* and *ArathADT6*, class I) were coregulated with other shikimate and phenylpropanoid pathway genes during stem development (Ehling *et al.*, 2005). All five putative *ADT* genes in *Populus* were well-expressed, though at different levels, in vascular as well as nonvascular tissues (Fig. 1d). This is consistent with high phenolic contents, not only in stems, but also in leaves and root tips of *Populus*. The presence of two distinct classes of putative ADTs in plants (Fig. 3) along with the differences in expression we observed (Fig. 1d), suggest organ-specific shikimate and phenylpropanoid functions that remain to be elucidated.

Phenylalanine ammonia-lyase catalyses the conversion of phenylalanine to cinnamate, and thereby initiates phenylpropanoid metabolism. *Arabidopsis* contains four *PALs*, two of which (*ArathPAL1* and *ArathPAL2*) have been experimentally associated with lignin and flavonoid biosynthesis (Rohde *et al.*, 2004). *Populus* contains five *PALs* in two distinct phylogenetic groups (data not shown), arising from one ancient (eurosoid) and two recent (salicoid) genome-wide gene duplication events (G.A. Tuskan, pers. comm.). Differential expression of the two *PAL* gene lineages in condensed tannin-rich (e.g.



**Fig. 1** Biosynthetic pathways leading to salicylate and salicin-based phenolic glycosides. Two alternative pathways have been proposed for salicylic acid (SA) biosynthesis, one requiring isochorismate synthase (ICS), and the other requiring phenylalanine ammonia-lyase (PAL). (a) List of phenolic glycosides identified in Salicaceae. (b) Possible SA precursors (benzoate or cinnamate derivatives) from the PAL-dependent pathway are shown. (c) Comparison of gene copy numbers in *Arabidopsis thaliana* (Arath) and *Populus trichocarpa* (Poptr). *PAT* genes have yet to be functionally characterized in plants. (d) Gene expression analysis by semi-quantitative reverse transcription polymerase chain reaction (RT-PCR) using various *Populus* tissues. YL, young leaves; R, root tips; EL, expanding leaves; PS, primary-growth stems; SS, secondary-growth stems; X, xylem. Numbers in parentheses indicate PCR cycles. Elongation factor 1-beta (*EF1 $\beta$* ) is included as a housekeeping gene control.



**Fig. 2** Neighbor-joining tree of *Populus*, *Arabidopsis* and rice chorismate mutase (CM) proteins. Bootstrap values (> 50%) are indicated at the nodes. Additional sequences include *Fagus sylvatica* (ABA54871), *Lycopersicon esculentum* (L47356) and *Saccharomyces cerevisiae* (P32178). Rice gene models were from the TIGR Rice Annotation Release 4 database.

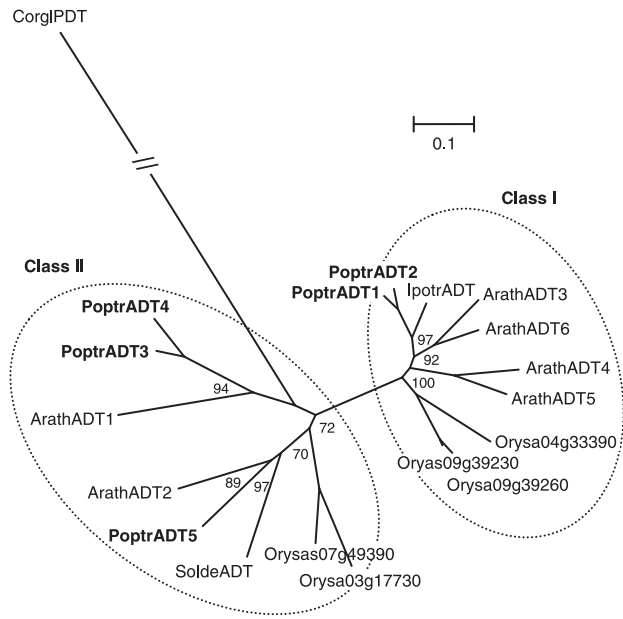
*PoptrPAL1*) and lignifying (e.g. *PoptrPAL2*) cells of aspen has been reported (Kao *et al.*, 2002). Similar expression patterns were observed among paralogs of each phylogenetic lineage (i.e. *PAL1* and *PAL3*; *PAL2* and *PAL4/5*; Fig. 1d), implying functional conservation. *PoptrPAL4* and *PoptrPAL5* are tandem

repeats with very high sequence homology (98.5% amino acid identity).

The PAL-independent SA biosynthetic pathway was reported following pathogen-induced *ArathICS1* expression in *Arabidopsis* (Fig. 1; Wildermuth *et al.*, 2001). This pathway is plastid-localized and may be analogous to the biosynthesis of salicylate-derived siderophores in microbes (Walsh *et al.*, 1990). An additional enzyme, isochorismate pyruvate lyase, converts isochorismate to salicylate in bacteria, but no homolog has been identified in plants (Fig. 1). Two *ICS* genes are present in *Arabidopsis*, but *ArathICS2*, encoding a cytosolic isoform, has not been characterized. *Populus* contains a single *PoptrICS* gene, with a plastid-targeting sequence, that was expressed primarily in green tissues (Fig. 1d) where PGs are known to accumulate. The relationship of *ICS* to SA and PG homeostasis in *Populus* remains unexplored.

### Hydroxycinnamates and hydroxycinnamate-derivatives

Biosynthesis and homeostatic control of PGs along with the other abundant phenolic derivatives of *Populus* (e.g. CTs and lignins) all interface with hydroxycinnamate (HC) metabolism (Fig. 1). Hydroxycinnamates originate from cinnamate through a series of hydroxylation and *O*-methylation reactions at the aromatic ring (Fig. 4). Their utilization in *Populus* is partly controlled by phylogenetically and biochemically distinct

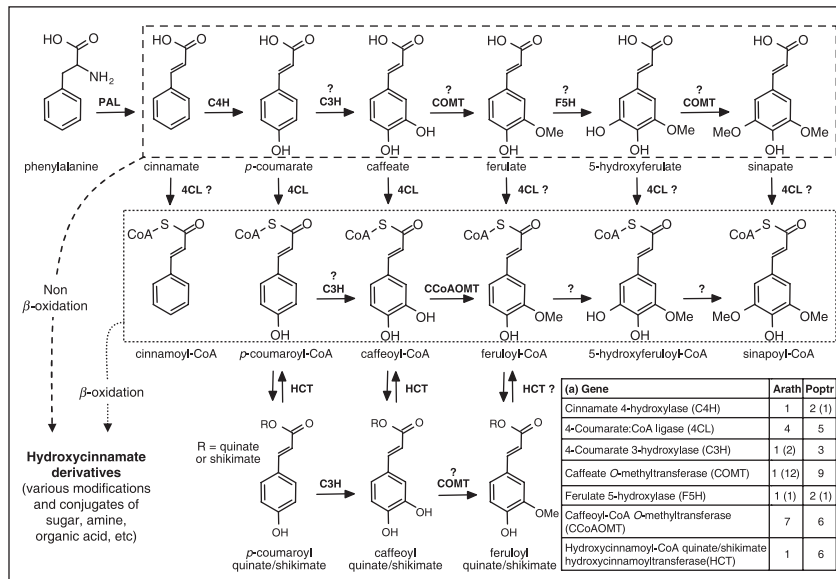


**Fig. 3** Neighbor-joining tree of putative *Populus*, *Arabidopsis* and rice arogenate dehydratases (ADTs). Bootstrap values (> 50%) are shown at major nodes. Additional sequences include *Corynebacterium glutamicum* (P10341), *Ipomoea trifida* (AAS79603), and *Solanum demissum* (AAT39307).

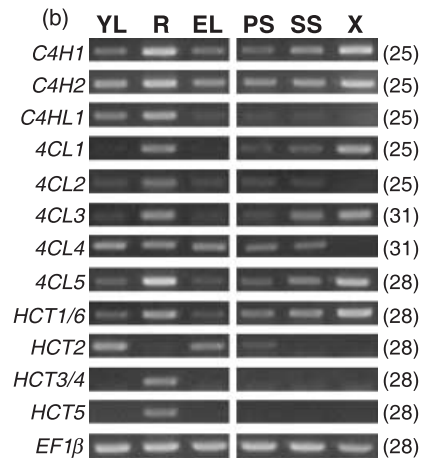
isoforms of 4-coumarate:CoA ligase (4CL) which are differentially expressed in lignifying (e.g. Poptr4CL1) and green/soft (Poptr4CL2) tissues (Harding *et al.*, 2002; Fig. 4b). Genome-wide duplication has given rise to two pairs of 4CL paralogs

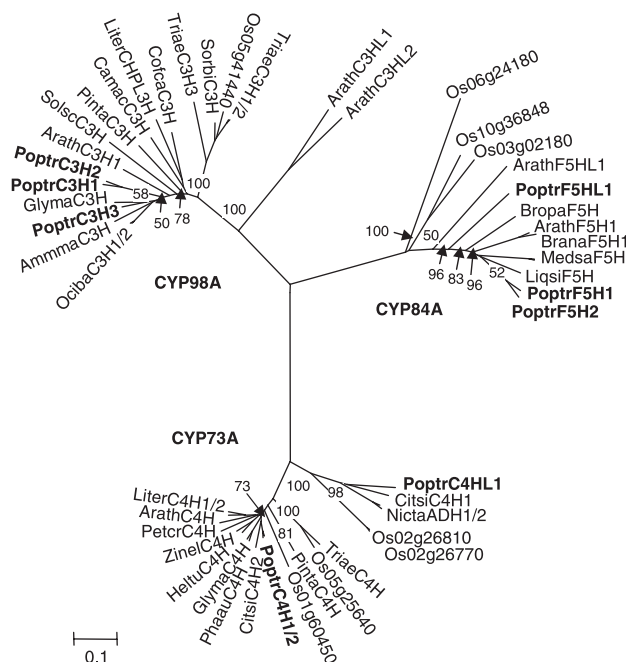
(Poptr4CL1/5 and Poptr4CL3/4). While Poptr4CL1 and Poptr4CL5 were similarly expressed, Poptr4CL3 and Poptr4CL4 exhibited distinct expression patterns of unexplored functional significance to HC utilization (Fig. 4b). Hydroxycinnamate synthesis starts with cinnamate 4-hydroxylase (C4H), a P450 monooxygenase of the CYP73A family represented by two highly similar isoforms, PoptrC4H1-2, in *Populus* (Fig. 5). Their transcripts were most abundant in lignifying xylem and root tips (Fig. 4b). A closely related C4H-like isoform (PoptrC4HL1) clusters with a wound-inducible citrus CitsiC4H2 and two tobacco elicitor-inducible CYP73A dihydroxylases involved in capsidiol biosynthesis (Fig. 5). PoptrC4HL1 transcript was detected in root tips and young leaves (Fig. 4b).

Specific isoforms of some of the downstream-acting HC/HCD-modifying hydroxylases and methyltransferases, (i.e. coumarate 3-hydroxylase (C3H), caffeate *O*-methyltransferase (COMT), caffeoyl *O*-methyltransferase (CCoAOMT) and ferulate 5-hydroxylase (F5H)), are now also known to utilize lignin synthesis-associated substrates, including HC-shikimate pathway esters and HC-aldehydes (for a recent review, see Boerjan *et al.*, 2003). The C5 hydroxylation is mediated by F5H, a CYP84A monooxygenase represented by two highly homologous, lignin-specific F5H isoforms (PoptrF5H1/2, on the basis of their EST digital expression profiles, Sterky *et al.*, 2004), and one F5H-like isoform (PoptrF5HL1) in *Populus* (Fig. 5). The C3 hydroxylation is catalysed by C3H of the CYP98A family, represented by a bona fide ArathC3H and two divergent, C3H-like isoforms in *Arabidopsis*, and by three members in *Populus* that cluster close to ArathC3H (Fig. 5).



**Fig. 4** Biosynthesis of hydroxycinnamates (HCs) and hydroxycinnamate derivatives (HCDs). The HCDs may be synthesized from HCs without  $\beta$ -oxidation, or via  $\beta$ -oxidation of HC-CoAs. Enzymatic steps that have not been unequivocally demonstrated are marked with '?'. (a) Comparison of gene copy numbers in *Arabidopsis* and *Populus*. Numbers in parentheses indicate '-like' genes. (b) Gene expression analysis by semiquantitative reverse transcription polymerase chain reaction (RT-PCR) using various *Populus* tissues as described in Fig. 1(d). RT-PCR analysis for the COMT and CCoAOMT families was not performed owing to the large number of genes.

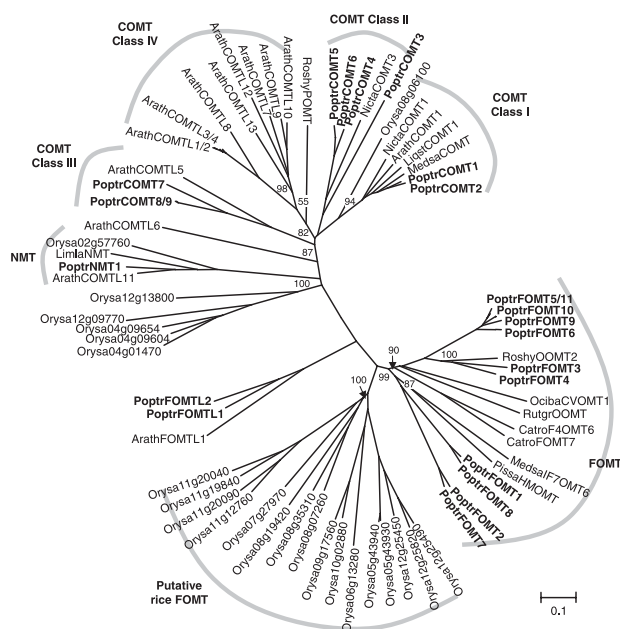




**Fig. 5** Neighbor-joining tree of P450 families involved in phenylpropanoid metabolism. Bootstrap values (> 50%) are shown at major nodes. Additional sequences analyzed are *Ammi majus* (AAT06912), *Brassica napus* (AAG14961), *Broussonetia papyrifera* (AAW50818), *Camptotheca acuminata* (AAS57921), *Citrus sinensis* (AAF66065-6), *Coffea canephora* (ABB83676), *Glycine max* (O48922, Q42797), *Helianthus tuberosus* (Q04468), *Liquidambar styraciflua* (AAD48912), *Lithospermum erythrorhizon* (BAB71716-7), *Medicago sativa* (ABB02162), *Ocimum basilicum* (AAL99200-1), *Petroselinum crispum* (Q43033), *Phaseolus aureus* (P37115), *Pinus taeda* (AAL47685, AAD23378), *Sesamum indicum* (AAL47545), *Solenostemon scutellarioides* (CAD20576), *Sorghum bicolor* (O48956), *Triticum aestivum* (CAE47489-71, AF123610), and *Zinnia elegans* (Q43240). A *Lithospermum erythrorhizon* protein (BAC44836), 4-coumaroyl-4'-hydroxyphenyllactic acid 3-hydroxylase (CHPL3H), clusters closer to *bona fide* C3Hs than to the two ArathC3HLs, suggesting functional divergence of the latter. Similarly, PoptrC4HL1, CitsiC4H1 and two rice homologs cluster with two tobacco elicitor-inducible 5-epi-aristolochene-1,3-dihydroxylases (ADH, AAK62344-5), away from the *bona fide* C4Hs.

The three *Populus C3H* genes appear to be differentially expressed, with *PoptrC3H3* and *PoptrC3H1/2* ESTs primarily detected in lignifying and nonvascular tissues, respectively (Sterky *et al.*, 2004).

The *O*-methylation of HCs and HC-CoA esters is catalysed by SAM-dependent COMTs and CCoAOMTs, respectively. Nine COMTs forming three phylogenetic classes (Fig. 6) have been identified in the *Populus* genome. They appear to arise from two ancient, genome-wide duplication events (*COMT1-3/5* and *COMT3/5-7/9*), with recent gene duplications giving rise to additional isoforms within each class (*COMT1-2*, *COMT3-5* and *COMT7-9*). Two tandem repeats (*COMT3-4* and *COMT5-6*) are also noted. Class I members *PoptrCOMT1* and *PoptrCOMT2* share 95% amino acid similarity and have been associated





**Table 1** Hydroxycinnamoyl–quininate conjugates in various pure and hybrid *Populus* species<sup>1</sup>

Section	<i>Aigeiros</i>				ISH <sup>2</sup>		<i>Tacamahaca</i>		<i>Populus</i>		ISH <sup>2</sup>
	A <sup>3</sup>	F	D	DN	TD	TDxTD	T	B	G	TR	TA
1-O-coumaroyl-quininate			X tr	X	X tr	X tr			X (1)	X (1)	X
3-O-coumaroyl-quininate			X tr	X	X tr	X tr				X (2)	X (3)
4-O-coumaroyl-quininate	X (1)	X (2)			X tr	X	X	X tr		X (3)	X (5)
5-O-coumaroyl-quininate	X tr	X			X tr	X tr	X tr		X	X	X (6)
<i>cis</i> -4-O-coumaroyl-quininate	X (2)	X (1)								X tr	X
1-O-caffeoyl-quininate		X tr	X		X tr	X tr	X tr	X tr	X		X
3-O-caffeoyl-quininate	X tr	X	X (1)	X (1)	X (1)	X (1)	X (1)	X	X tr	X	X (1)
4-O-caffeoyl-quininate	X tr	X (3)	X		X	X	X (2)	X (1)	X tr	X tr	X (2)
5-O-caffeoyl-quininate	X tr	X	X tr		X	X	X (3)	X	X	X tr	X (4)
<i>cis</i> -4-O-caffeoyl-quininate											X
1-O-feruloyl-quininate		X	X tr						X tr		X tr
3-O-feruloyl-quininate			X								X
4-O-feruloyl-quininate		X									X
5-O-feruloyl-quininate	X tr	X									X

<sup>1</sup>X denotes that the metabolite is present in the leaves of a given species or hybrid clone. A particular conjugate may be present in only trace levels (tr). The ranks of particularly abundant conjugates in a particular are shown in parentheses. *Populus* species: A, *P. angustifolia*; F, *P. fremontii*; D, *P. deltoides*; DN, *P. deltoides* × *nigra*; TD, *P. trichocarpa* × *deltoides* (F<sub>1</sub> hybrid); TDxTD, *P. trichocarpa* × *deltoides* (F<sub>2</sub> hybrid); T, *P. trichocarpa*; TR, *P. tremuloides*; B, *P. balsamifera*; G, *P. grandidentata*; TA, *P. tremula* × *alba*.

<sup>2</sup>ISH, Interspecific hybrids.

<sup>3</sup>*Populus angustifolia* was conventionally classified in Section *Tacamahaca*, but exhibits chemotaxonomic characteristics typical of Section *Aigeiros* (Greenaway & Whatley, 1990).

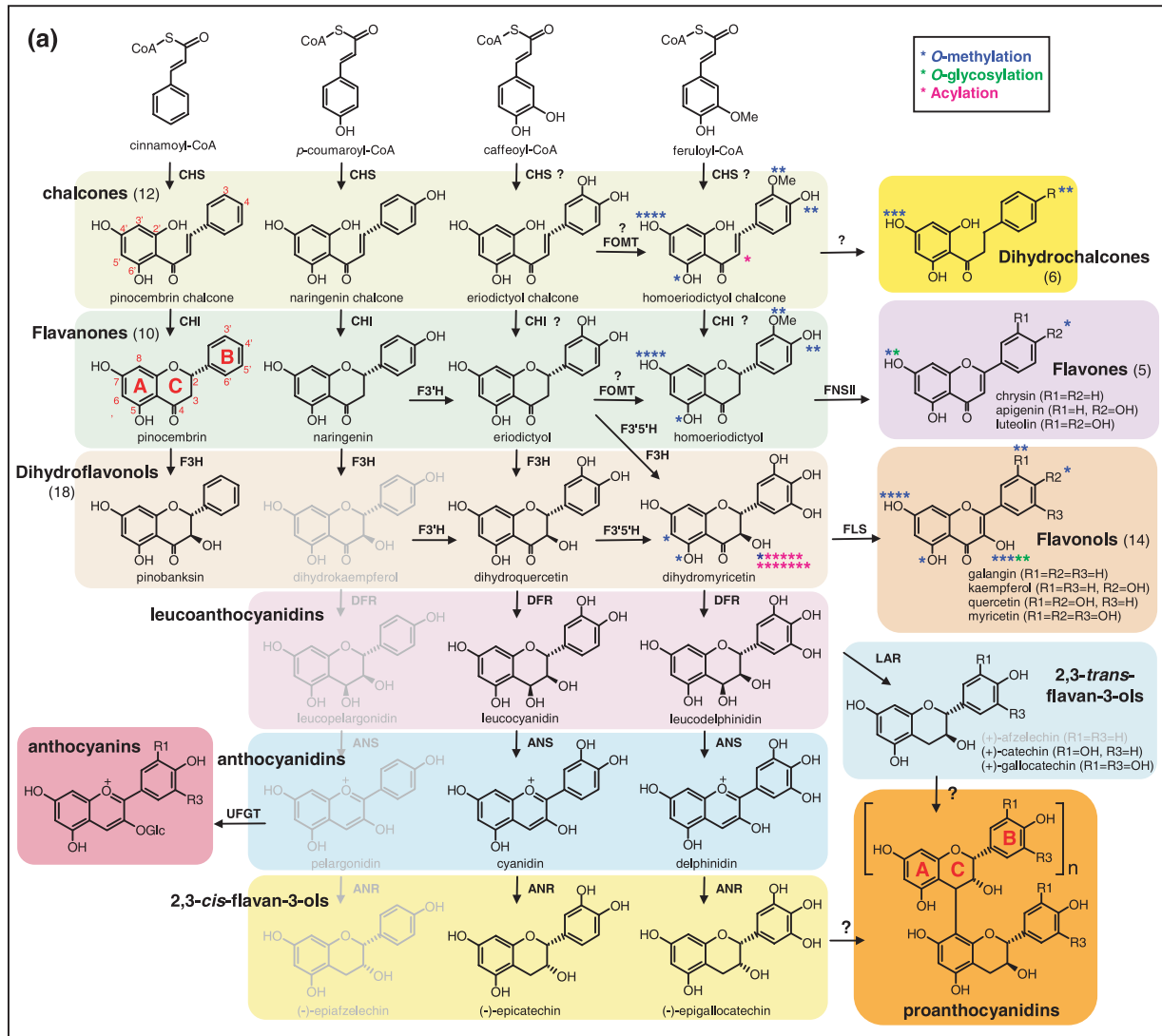
(Fig. 9), rarely accumulate to detectable levels in angiosperms (Harborne, 1977) and are absent from Table S3. Like HCDs, occurrence and concentration of flavonoids vary among *Populus* species and clones (Greenaway *et al.*, 1991b, 1992). In addition, CT concentrations vary among vegetative tissues. For example, in the *P. fremontii* × *angustifolia* clone used in this study, CT concentrations are highest in young leaves and root tips (*c.* 100 µg CT mg<sup>-1</sup> DW) and much lower in green stems and developing xylem (< 4 µg CT mg<sup>-1</sup> DW; S. A. Harding, unpublished).

Flavonoid biosynthesis is initiated by chalcone synthase (CHS), a plant-specific polyketide synthase, that occurs as a large gene family in *Populus*, consisting of at least six *CHS* and seven *CHS*-like (*CHSL*) genes, in addition to three or four putative pseudogenes. Five of the six *CHS* genes are situated as tandem repeats in linkage groups I (*CHS2-3*) and III (*CHS4-6*), resulting from recent, genome-wide duplication events (G.A. Tuskan, pers. comm.). All six *Populus* CHS proteins cluster near bona fide angiosperm CHSs that have been biochemically or genetically characterized (Fig. 10). Transcripts of *PoptrCHS1* and *PoptrCHS3* were more abundant in root tips than in young leaves, whereas *PoptrCHS2*, *PoptrCHS5* and *PoptrCHS6* appeared to encode root isoforms, as their expression in leaves was below detection (Fig. 9c). *PoptrCHS4* is the only isoform not expressed in root tips. Expression of *PoptrCHSL2-5* was also detected in leaves and root tips (data not shown).

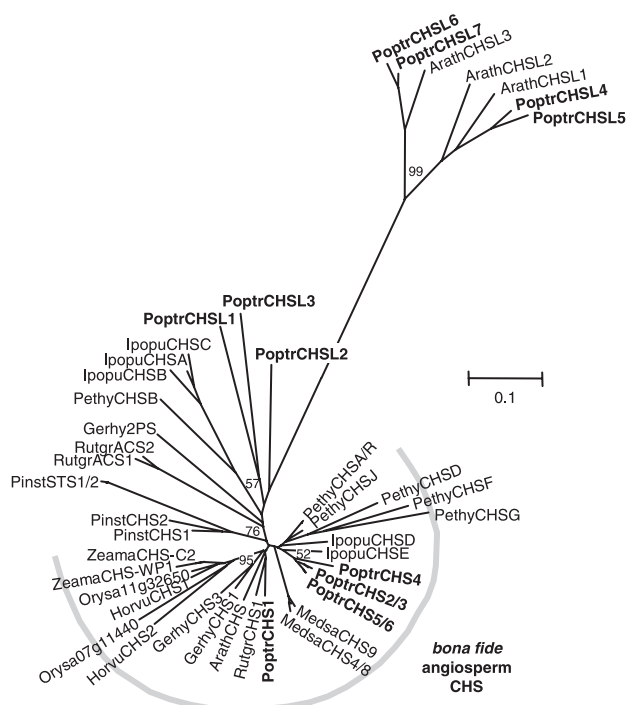
Chalcone isomerase (CHI) is encoded by a single gene in both *Arabidopsis* and *Populus*. *ArathCHI1* and *PoptrCHI1* cluster with other Class I CHIs associated with flavonoid biosynthesis, away from Class II isoflavonoid-related CHIs from legumes (Fig. 11). In addition, *Arabidopsis* and *Populus* each contain one of the class III and class IV CHI-like (CHIL) proteins of unknown function recently reported in soybean (Ralston *et al.*, 2005). The single-copy *PoptrCHI* gene was more highly expressed in root tips than in leaves (Fig. 9c).

Flavonoid chemical diversity results in part from the pattern of B-ring hydroxylation, regulated by two P450 enzymes, flavonoid 3'-hydroxylase (F3'H) of the CYP75B family and flavonoid 3',5'-hydroxylase (F3'5'H) of the CYP75A family (Fig. 12). One *F3'H* and two duplicated *F3'5'H* genes are present in the *Populus* genome. *PoptrF3'H* expression was much stronger in root tips than in young leaves (Fig. 9c). *PoptrF3'5'H1/2* transcript abundance was similar in young leaves and root tips. Another P450 enzyme of the CYP93B family, flavone synthase II (FNSII) required for flavone synthesis is encoded by at least two genes in *Populus* (Fig. 9b). Neither gene was amplified from leaf or root tip cDNA, but *FNSIII* was readily detected in apical buds of *P. tremuloides* (data not shown), consistent with the detection of flavones in *Populus* bud exudates (Greenaway *et al.*, 1991b, 1992). GenBank EST analysis also identified five *FNSIII* homologs from terminal vegetative buds (CV245564), imbibed seed (BU863232, DN498422) and shoot tips (CN522512),





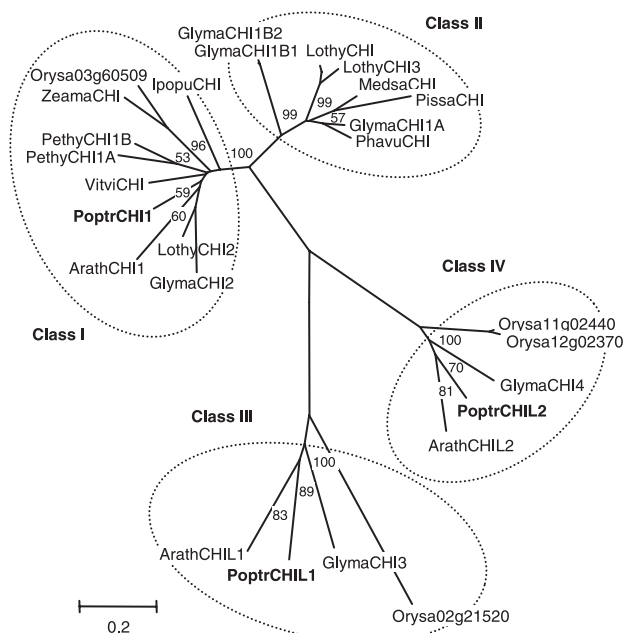
**Fig. 9** Possible network of flavonoid biosynthesis genes in *Populus*. (a) Different flavonoid classes are shaded in color. The number in parentheses following each flavonoid class indicates the number of compounds identified by GC-MS (Table S3). Gray indicates metabolites/enzymatic steps that are unsupported in *Populus*. Colored stars indicate the position and identity (legend top right) of additional substitutions listed in Table S3. The number of stars corresponds to the number of reported occurrences of a substitution. (b) Comparison of gene copy numbers in *Arabidopsis* and *Populus*. Numbers in parentheses indicate 'like' genes. ArathCOMT1 has been shown to exhibit flavonol 3'-OMT activity, and hence is also listed as FOMT. (c) Gene expression analysis by semiquantitative reverse transcription polymerase chain reaction (RT-PCR) using young leaves (YL) and root tips (R). Numbers in parentheses indicate PCR cycles.



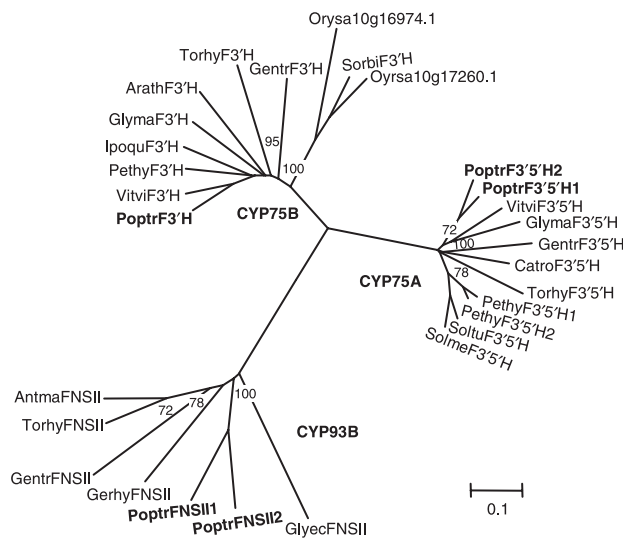
**Fig. 10** Neighbor-joining tree of chalcone synthases (CHSs) and other representative polyketide synthases. Bootstrap values (> 50%) are indicated for principal nodes. Sequences shown are *Medicago sativa* (P30075-7), *Gerbera hybrida* (S56699, S55464-5), *Hordeum vulgare* (P26018, Q96562), *Ipomoea purpurea* (O22045, O22047, P48399, T10951, T11000), *Petunia hybrida* (SYPJCA-B, SYPJCD, SYPJCF-G, SYPJCI, SYPJCN), *Pinus strobus* (CAA06077, CAA05214, P48407-8), *Ruta graveolens* (Q9F5B9, Q9F5C0, S60241) and *Zea mays* (P24824-5).

suggesting that flavone synthesis is under strict developmental regulation.

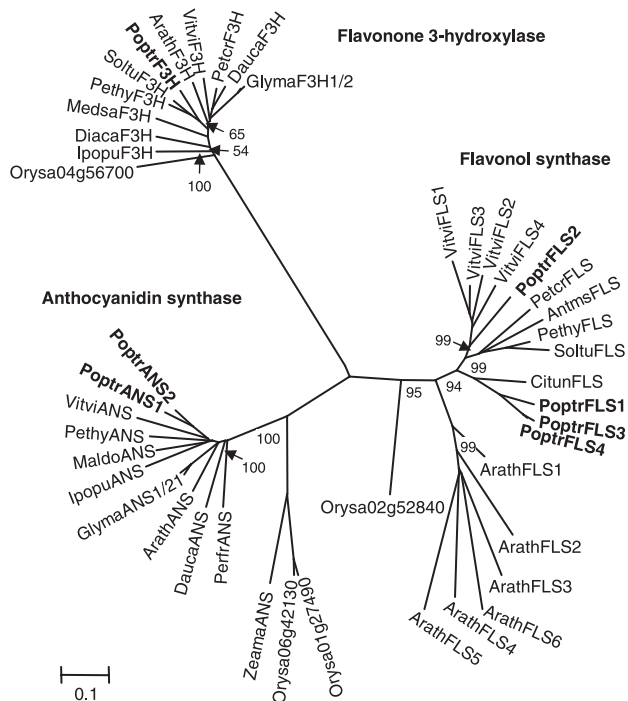
Subsequent conversion of flavanones to dihydroflavonols, flavonols, leucoanthocyanidins, anthocyanidins and flavan-3-ols follows a series of oxidation-reduction reactions on the heterocyclic C-ring, catalysed by flavanone 3-hydroxylase (F3H), flavonol synthase (FLS), dihydroflavonol reductase (DFR), anthocyanidin synthase (ANS), leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR), also known as BANYULS or BAN (Fig. 9). Flavanone 3-hydroxylase, FLS and ANS are oxoglutarate-dependent dioxygenases, encoded by one, four and two genes, respectively, in *Populus* (Fig. 13). The putatively single-copy *PoptrF3H* was more highly expressed in root tips than young leaves (Fig. 9c). *PoptrANS1* was also preferentially expressed in root tips, while *PoptrANS2* transcript level was high in both young leaves and root tips. *FLS* genes exhibited three distinct expression patterns: *PoptrFLS1* appeared to be leaf-specific, while the more divergent *PoptrFLS2* was poorly expressed in leaves. Transcript levels of *PoptrFLS3* and *PoptrFLS4* were similar in leaves and roots (Fig. 9c). Interestingly, phylogenetic analysis clusters *Arabidopsis* FLS isoforms separately from *Populus* and other previously



**Fig. 11** Neighbor-joining tree of chalcone isomerases (CHIs). Bootstrap values (> 50%) are indicated for principal nodes. Sequences shown are *Glycine max* (P11651, AAT94359-62, AAT94364), *Ipomoea purpurea* (O22604), *Lotus corniculatus* var. *japonicus* (BAC54038, BAC53983-53984, CAD69022), *Medicago sativa* (P28012), *Petunia hybrida* (AAF60296, P11651), *Phaseolus vulgaris* (P14298), *Pisum sativum* (P41089), *Vitis vinifera* (P51117), *Zea mays* (Q08704). CHIL, CHI-like.



**Fig. 12** Neighbor-joining tree of P450 families involved in flavonoid biosynthesis. Bootstrap values (> 50%) are shown at the nodes. Additional sequences include *Antirrhinum majus* (BAA84071), *Catharanthus roseus* (CAA09850), *Gentiana triflora* (BAD91808-9, Q96581), *Gerbera hybrida* (AAD39549), *Glycine max* (AAM51564, BAD97828), *Ipomoea quamoclit* (AAS46257), *Glycyrrhiza echinata* (P93149), *Petunia hybrida* (Q95BQ9, P48418-9), *Solanum melongena* (P37120), *Solanum tuberosum* (AAV85471), *Sorghum bicolor* (AAV74195), *Torenia hybrida* (BAA84072, BAB20076, BAB87838) and *Vitis vinifera* (CAI54277, BAE47005). F3'H, flavanone 3-hydroxylase; F3'5'H, flavonoid 3',5'-hydroxylase; FNS, flavone synthase.

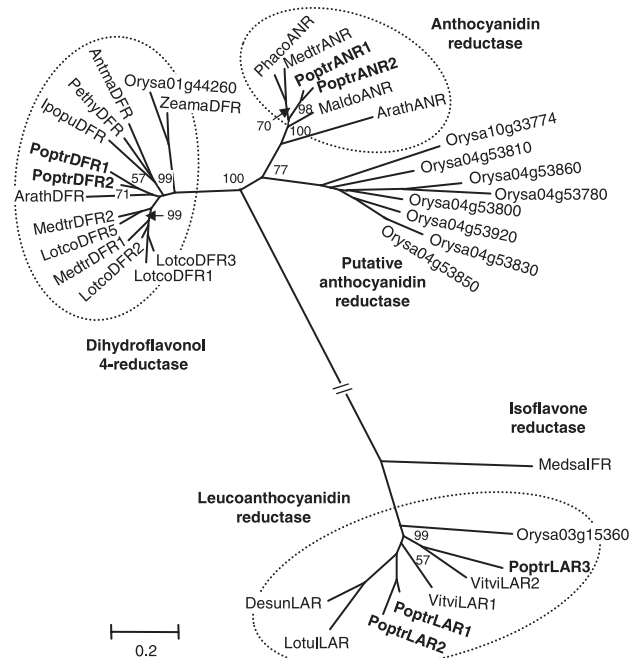


**Fig. 13** Neighbor-joining tree of flavanone 3-hydroxylase (F3H), anthocyanidin synthase (ANS) and flavonol synthase (FLS) proteins. Bootstrap values (> 50%) are indicated at the nodes. Sequences analyzed are *Antirrhinum majus* (ABB53382), *Citrus unshiu* (Q9ZWQ9), *Daucus carota* (AAD56577, AAD56580), *Dianthus caryophyllus* (Q05964), *Glycine max* (AAY82085, AAR26525-6), *Ipomoea purpurea* (AAB41102, AAB84049), *Malus x domestica* (P51091), *Medicago sativa* (CAA55628), *Perilla frutescens* (O04274), *Petroselinum crispum* (AAP57394, AAP57395), *Petunia hybrida* (AAC49929, P51092, Q07512), *Solanum tuberosum* (AAM48289, Q41452), *Vitis vinifera* (BAE75806-9, P41090, P51093), *Zea mays* (P41213).

characterized *Petunia* and *Citrus* FLS isoforms (Fig. 13), pointing to possible functional divergence.

Dihydroflavonol reductase, ANR and LAR are NADPH-dependent reductases, each encoded by two or three genes in *Populus* (Fig. 14). The two *DFR* genes are derived from a recent gene duplication but exhibited different expression patterns (Fig. 9c). *PoptrDFR2* transcripts were detected in both roots and leaves, but *PtDFR1* expression was very weak in leaves (Fig. 9c). By contrast, the duplicated *ANR* genes exhibited similar expression patterns, both with a higher transcript abundance in roots. The three *Populus* *LAR* genes occur in two distinct phylogenetic lineages, but showed little difference in their tissue distribution.

Numerous structural variants, including *O*-methylated, *O*-glycosylated and acylated flavonoids have been identified (Fig. 9a). These modifications vary across species and involve enzymes of the glycosyltransferase, *O*-methyltransferase and acyltransferase superfamilies. The most common is *O*-methylation, which occurs in all classes of flavonoids (Bohm, 1998). Analysis of the *Populus* genome identified 11 putative

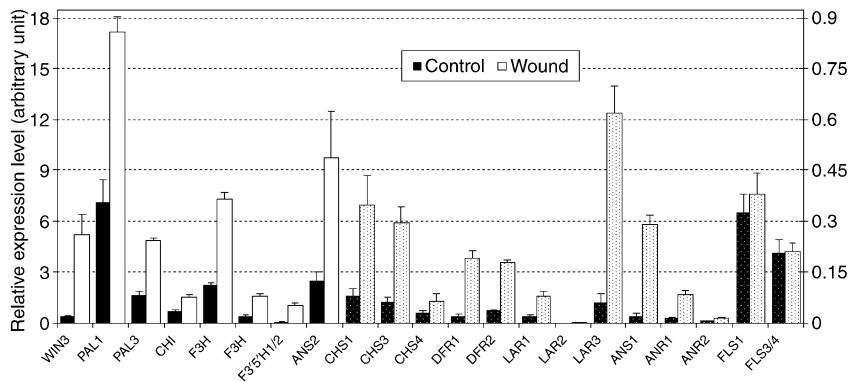


**Fig. 14** Neighbor-joining tree of dihydroflavonol reductase (DFR), anthocyanidin reductase (ANR) and leucoanthocyanidin reductase (LAR) proteins. Bootstrap values (> 50%) are indicated at the nodes. Sequences analysed are *Desmodium uncinatum* (Q84V83), *Ipomoea purpurea* (BAA74699), *Lotus corniculatus* var. *japonicus* (BAE19948-9, BAE19950, BAE19953), *Lotus uliginosus* (AAU45392), *Malus x domestica* (AAZ17408), *Medicago sativa* (P52575), *Medicago truncatula* (AAN77735, AAR27014-5), *Petunia hybrida* (P14721), *Phaseolus coccineus* (CAD91909), *Vitis vinifera* (CAI26308, CAI26310), *Zea mays* (P51108).

flavonoid *O*-methyltransferases (FOMTs, Fig. 6) that are phylogenetically related to an alfalfa isoflavonoid OMT (MedsalF7OMT, He *et al.*, 1998), and a rose orcinol OMT (RoshyOOMT2, Lavid *et al.*, 2002) involved in floral scent volatile emission. By contrast, the only *Arabidopsis* OMT reported to methylate flavonoids (ArathCOMT1; Muzac *et al.*, 2000) clusters with lignin-related COMTs (Fig. 6) and is expressed most abundantly in lignifying tissues (Raes *et al.*, 2003; Ehltling *et al.*, 2005). Two distantly related *Populus* and one *Arabidopsis* proteins were also identified and designated FOMT-like (FOMTL). Like other *FOMT* genes that have been characterized, the 11 putative *PoptrFOMT* and the two *PoptrFOMTL* genes contain a single intron, whereas *COMT1* *COMTL*/*NMT* genes contain 4 introns.

### Wound-induced foliar CT response and flavonoid gene expression

To confirm association of the annotated *Populus* flavonoid genes with CT biosynthesis, leaf wounding was applied to a hybrid cottonwood (*P. fremontii* × *angustifolia*) genotype exhibiting systemically inducible CT, but not PG, accumulation. Five days



**Fig. 15** Quantitative reverse transcription polymerase chain reaction analysis of flavonoid pathway gene transcript abundance in rapidly expanding sink leaves of control (closed bars) and wounded (open bars) cottonwood plants. Refer to the right-hand side axis for dotted bars. Expression levels were normalized to the geometric mean of four housekeeping genes. Data represent the mean  $\pm$  SE of three biological replicates.

after source leaf wounding, CT concentrations in unwounded sink leaves increased nearly 2.5-fold, from  $114.1 \pm 15.9 \mu\text{g CT mg}^{-1}$  leaf DW in unwounded control plants to  $273.8 \pm 20.8 \mu\text{g CT mg}^{-1}$  leaf DW in wounded plants. Accompanying the CT increase were a 14-fold upregulation of a wound-inducible *WIN3* gene belonging to the Kunitz-type trypsin inhibitor family (Bradshaw *et al.*, 1989), and a two- to three-fold upregulation of the CT-specific *PoptrPAL1* (Kao *et al.*, 2002) and its paralog *PoptrPAL3* (Fig. 15). Transcript abundance of those leaf-expressed flavonoid genes essential for CT precursor proanthocyanidin synthesis was also wound-stimulated by 2- to 17-fold (Fig. 15). Flavonol synthases do not catalyse flavonoid interconversions for CT biosynthesis (Fig. 9), and in this experiment, *PoptrFLS1* and *PoptrFLS3/4* were the only leaf-expressed flavonoid genes not upregulated by wounding. *PoptrPAL1/3*, *PoptrANS2*, *PoptrF3H* and *PoptrCHI1* were among the most highly expressed flavonoid genes in unwounded leaves and remained the most abundant flavonoid transcript species in wounded leaves (Fig. 15). Differential wound-inducibility among gene family members was also noted, such as *PoptrLAR3* (tenfold) vs *PoptrLAR1/2* (fourfold); and *PoptrANR1* (sixfold) vs *PoptrANR2* (twofold), pointing to differential regulation.

## Discussion

*Populus* species and hybrids support localized zones of distinct ecological character through differences in their phytochemistry (Whitham *et al.*, 2003; Schweitzer *et al.*, 2004; Bangert *et al.*, 2006). To a large extent, this depends on traits controlling the quantity and composition of shikimate-phenylpropanoid products accumulating in various vegetative organs. Genetic analysis of the transmission of these traits within cross-hybridizing populations of *Populus* and *Salix* has defined phytochemically distinct lineages (Hardig *et al.*, 2000; Whitham *et al.*, 2003), which should be of value for investigating secondary metabolism and its relevance to evolutionary selection, phytoremediation, carbon sequestration, and energy and biomass production. To this end, a comprehensive list of structural gene sequences corresponding to parts of the shikimate and

aromatic amino acid biosynthetic pathways, as well as to known steps of HCD and flavonoid biosynthesis and modification, was compiled from the *Populus* genome.

In general, gene families involved in upstream steps controlling the utilization of chorismate for phenylalanine (and potentially, PG) biosynthesis were found to be conserved between *Populus* and *Arabidopsis*. However, the downstream HCD and flavonoid metabolism gene families are much larger in *Populus*, and more reflective of the common tendency toward expanded gene families in the secondary metabolism of higher plant systems (Durbin *et al.*, 2000; Rausher *et al.*, 1999; Gebhardt *et al.*, 2005). In *Populus*, the quantitative commitment to phenolic metabolism is high both in soft tissues (PG, HCD and CT) and in stems (HCD and lignin), and may be an important driving force for gene family expansion and specialization among organs. In this report, gene expression (Figs 1d, 4b, 9c and 15) and metabolite data (Tables 1 and S3) were provided to infer functional significance for the expanded *Populus* HCD and flavonoid gene families.

*Arabidopsis* lacks two attributes – reproduction by cross-pollination, and perennial, woody growth habit – that strongly correlate with flavonoid diversification and abundance in higher plants (Harborne, 1977; Stafford, 1991). While the metabolic expense of flavonoids appears to be low in *Arabidopsis*, it is high in *Populus*, where some of the costliest and most elaborate flavonoid-derived structures – the various CTs – accumulate throughout the plant (Lindroth & Hwang, 1996; Kandil *et al.*, 2004). Consistent with this, all enzymes of the flavonoid biosynthetic pathway, except FLS, are encoded by single-copy genes in *Arabidopsis*. By contrast, all but CHI and F3'H, and possibly F3H (see Table S1), are encoded by gene families in *Populus*. A striking example is the greatly expanded *CHS* gene family in *Populus* vs *Arabidopsis*. *CHS* mutations and gene duplications have been associated with altered substrate specificity (Christensen *et al.*, 1998), albino flower color phenotypes (Durbin *et al.*, 2000) and evolution of closely related polyketide synthases (Fig. 10), including stilbene synthase (STS; Raiber *et al.*, 1995), acridone synthase (ACS; Springob *et al.*, 2000), and 2-pyrone synthase (2PS; Helariutta *et al.*, 1996). Chalcone synthase catalyses the conversion of

4-coumaryl-CoA, and in some species, cinnamoyl-CoA, to naringenin chalcone and pinocembrin chalcone, respectively (reviewed in Bohm, 1998). Based on the frequent occurrence, and in many instances the predominance of pinobanksin and its derivatives in *Populus* bud exudates (Table S3), *Populus* CHS isoforms appear to utilize both cinnamyl-CoA and *p*-coumaryl-CoA to initiate flavonoid synthesis. Separate CHS isoforms have been reported in barley where alternative substrates, caffeoyl-CoA and feruloyl-CoA, are utilized for conversion to eriodictyol and homoeriodictyol chalcone, respectively (Christensen *et al.*, 1998). Pinocembrin, eriodictyol and homoeriodictyol chalcones and their corresponding flavanones are also detected in *Populus* (Fig. 9, Greenaway *et al.*, 1991c). It remains to be determined whether the large *Populus* CHS family somehow controls flavonoid diversity in *Populus*, but the expression of root- and leaf-specific isoforms, as well as their differential upregulation by wounding, are consistent with expanded CHS function.

Dihydroflavonol reductase catalyses the conversion of dihydroflavonols (i.e. dihydrokaempferol, dihydroquercetin and dihydromyricetin) to colorless leucoanthocyanidins which, through the activity of ANS, give rise to pelargonidin, cyanidin and delphinidin pigments, respectively (Fig. 9a). Like *CHS*, *DFR* is present as a multigene family in many higher plant species but not in *Arabidopsis*. The specificity of *DFR* toward its various dihydroflavonol substrates is known to influence flower color (Helariutta *et al.*, 1993; Johnson *et al.*, 2001), seed coat pigmentation (Marles *et al.*, 2003) and CT composition in vegetative tissues (Bavage *et al.*, 1997; Xie *et al.*, 2004; Shimada *et al.*, 2005). For example, petunia *DFR* cannot efficiently reduce dihydrokaempferol to the corresponding leucopelargonidin, but a single amino acid substitution at position 134 led to a mutant *DFR* that preferentially utilized dihydrokaempferol (Johnson *et al.*, 2001). Neither *Populus* *DFR* isoform is predicted to be active toward dihydrokaempferol (Peters & Constabel, 2002), nor is dihydrokaempferol among the 16 dihydroflavonols commonly detected in *Populus* (Table S3). In vegetative tissues, this is consistent with an absence of kaempferol-derived pelargonidin type subunits in *Populus* CTs (Ayres *et al.*, 1997). The apparent absence of dihydrokaempferol may be attributed to F3'H/F3'5'H that converts dihydrokaempferol into dihydroquercetin and dihydromyricetin, or it can be attributed to *FLS* catalytic activity that converts dihydroflavonols into flavonols, which do not enter into CT biosynthesis. *Populus* *FLS* apparently converts dihydrokaempferol very efficiently, since various methylated and glycosylated forms of the *FLS* product, kaempferol, are routinely detected in *Populus* (Fig. 9, Table S3). In carnation (*Dianthus caryophyllous*) and rose (*Rosa hybrida*), differential timing of *DFR* and *FLS* expression provides transcriptional control to mitigate the competition for dihydroflavonol substrates during floral pigmentation (Stich *et al.*, 1992; Noda *et al.*, 2004). Control mechanisms in phenol-rich vegetative tissues of *Populus* may be organ specific, as

exemplified in Fig. 9c, and subject to a wide range of environmental cues.

*Populus* CTs exhibit a range of subunit compositions in the ratio of procyanidin:prodelphinidin (PC:PD) type subunits, and in the ratio of heterocyclic C-rings with *cis* vs *trans* stereochemistry at C3 (Ayres *et al.*, 1997). Such chemical complexity is not seen in *Arabidopsis* and certain herbaceous species, and this can be attributed to flavonoid pathway gene diversity. *Arabidopsis*, snapdragon and maize are incapable of producing delphinidin pigments and PD-type CT subunits because of the absence of F3'5'H (Holton & Cornish, 1995; Abrahams *et al.*, 2003). Further, *Arabidopsis* produces only epicatechin-type CT monomers because of the absence of LAR (Tanner *et al.*, 2003). Within *Populus* species, the variation in CT subunit composition and stereochemistry likely affords differential biological activities that are not well understood. Leaf wounding, for example, stimulated expression of flavonoid pathway genes or gene family members by varying degrees (e.g. *LAR*, *ANS* and *ANR*). Effects of altered abundance or ratio of isoforms on flavonoid/CT composition remain to be investigated, but it is thought that protein–protein interactions are essential for the control of flavonoid product diversity (reviewed in Winkel, 2004). Continuing investigation into regulation of flavonoid gene expression, CT composition and biological activities in *Populus* species should shed light on the biosynthesis and role of CTs in the adaptive strategies of these species.

In addition to CTs, large reserves of PGs and HCDs accumulate in bark and foliage of *Populus* and willow species (Lindroth & Hwang, 1996). While production of CTs is highly plastic and varies in response to resource availability and damage (Fig. 15), production of PGs is largely genetically canalized, and potentially costly to plant growth, especially of high-PG genotypes (Osier & Lindroth, 2001, 2004, 2006). As indicated in Fig. 15, gene function pertinent to CT homeostasis can be elucidated using inductive treatments such as wounding. Gene function pertinent to PG homeostasis, however, will likely benefit from comparative transcriptomic analysis of distinct genotypes. Now that the *Populus* genome has been sequenced, better understanding of the gene networks controlling PG, HCD and CT dynamics of consequence to *Populus* tree growth, habitat productivity, species interactions and global carbon cycling seems much nearer. Expanding resources of phytochemically characterized, genetically related lineages should fuel detailed investigation of these pathways and their roles in unique aspects of tree biology.

## Acknowledgements

This work is supported by the US National Science Foundation Plant Genome Program DBI-0421756 (C.-J.T., S.A.H.) and the Ecology Program DEB-0074427 and DEB-0078280 (R.L.L.), and by the US Department of Energy, Office of Science, Biological and Environmental Research Program DE-FG02–

05ER64112 (S.A.H., C.-J.T., R.L.L.), and the Biological and Environmental Research, the Ecosystem Research and the Molecular Biology Programs at Oak Ridge National Laboratory (T.J.T.). Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. The authors thank Björn Hamberger and Carl Douglas of the University of British Columbia for sharing some of the shikimate/phenylpropanoid pathway gene annotation information.

## References

- Abrahams S, Lee E, Walker AR, Tanner GJ, Larkin PJ, Ashton AR. 2003. The *Arabidopsis* TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *Plant Journal* 35: 624–636.
- Ayres MP, Clausen TP, MacLean SF Jr, Redman AM, Reichardt PB. 1997. Diversity of structure and antiherbivore activity in condensed tannins. *Ecology* 78: 1696–1712.
- Bangert RK, Turek RJ, Rehil B, Allan GJ, Wimp GM, Schweitzer JA, Allan GJ, Bailey JK, Martinsen GD, Keim P, Lindroth RL, Whitham TG. 2006. A genetic similarity rule determines arthropod community structure. *Molecular Ecology* 15: 1379–1392.
- Bavage AD, Davies IG, Robbins MP, Morris P. 1997. Expression of an *Antirrhinum* dihydroflavonol reductase gene results in changes in condensed tannin structure and accumulation in root cultures of *Lotus corniculatus* (bird's foot trefoil). *Plant Molecular Biology* 35: 443–458.
- Behrens A, Maie N, Knicker H, Kogel-Knabner I. 2003. MALDI-TOF mass spectrometry and PSD fragmentation as means for the analysis of condensed tannins in plant leaves and needles. *Phytochemistry* 62: 1159–1170.
- Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* 54: 519–546.
- Bohm BA. 1998. *Introduction to flavonoids*, Vol. 2. Amsterdam, Netherlands: Hardwood Academic Publishers.
- Bradshaw HD, Hollick JB, Parsons TJ, Clarke HRG, Gordon MP. 1989. Systemically wound-responsive genes in poplar trees encode proteins similar to sweet potato sporamins and legume Kunitz trypsin inhibitors. *Plant Molecular Biology* 14: 51–59.
- Bugos RC, Chiang VLC, Campbell WH. 1991. cDNA cloning, sequence analysis and seasonal expression of lignin-bispecific caffeic acid 5-hydroxyferulic acid *O*-methyltransferase of aspen. *Plant Molecular Biology* 17: 1203–1215.
- Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* 11: 113–116.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the CLUSTAL series of programs *Nucleic Acids Research* 31: 3497–3500.
- Christensen AB, Gregersen PL, Schroder J, Collinge DB. 1998. A chalcone synthase with an unusual substrate preference is expressed in barley leaves in response to UV light and pathogen attack. *Plant Molecular Biology* 37: 849–857.
- Coyle DR, Nebeker TE, Hart ER, Mattson WJ Jr. 2005. Biology and management of insect pests in North American intensively-managed hardwood forest systems. *Annual Review of Entomology* 50: 1–29.
- Durbin ML, McCaig B, Clegg MT. 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Molecular Biology* 42: 79–92.
- Ehlting J, Mattheus N, Aeschliman DS, Li EY, Hamberger B, Cullis IF, Zhuang J, Kaneda M, Mansfield SD, Samuels L, Ritland K, Ellis BE, Bohlman J, Douglas CJ. 2005. Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant Journal* 42: 618–640.
- Gebhardt Y, Witte S, Forkmann G, Lukacin R, Matern U, Martens S. 2005. Molecular evolution of flavonoid dioxygenases in the family Apiaceae. *Phytochemistry* 66: 1273–1284.
- Gebre GM, Tschaplinski TJ. 2002. Solute accumulation of chestnut oak and dogwood leaves in response to throughfall manipulation of an upland oak forest. *Tree Physiology* 22: 251–260.
- Grace SC, Logan BA. 2000. Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. *Philosophical Transactions of the Royal Society of London Series B* 355: 1499–1510.
- Greenaway W, Whatley FR. 1990. Analysis of phenolics of bud exudate of *Populus angustifolia* by GC-MS. *Phytochemistry* 29: 2551–2554.
- Greenaway W, English S, May J, Whatley FR. 1991a. Chemotaxonomy of section *Leuce* poplars by GC-MS of bud exudates. *Biochemical Systematics and Ecology* 19: 507–518.
- Greenaway W, English S, Whatley FR. 1991b. Interrelationships of poplars in a hybrid swarm as studied by gas chromatography–mass spectrometry. *Canadian Journal of Botany* 69: 203–208.
- Greenaway W, English S, May J, Whatley FR. 1991c. Analysis of phenolics of bud exudate of *Populus sieboldii* by GC-MS. *Phytochemistry* 30: 3005–3008.
- Greenaway W, English S, Whatley FR. 1992. Relationships of *Populus* × *acuminata* and *Populus* × *generosa* with their parental species examined by gas chromatography–mass spectrometry of bud exudates. *Canadian Journal of Botany* 70: 212–221.
- Harborne JB. 1977. Flavonoids and the evolution of the angiosperms. *Biochemical Systematics and Ecology* 5: 7–22.
- Hardig TM, Brunsfeld SJ, Fritz RS, Morgan M, Orrians C. 2000. Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. *Molecular Ecology* 9: 9–24.
- Harding SA, Leshkevich J, Chiang VL, Tsai C-J. 2002. Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiology* 128: 428–438.
- Harding SA, Jiang H, Jeong ML, Casado FL, Lin H-W, Tsai C-J. 2005. Functional genomics analysis of foliar condensed tannin and phenolic glycoside regulation in natural cottonwood hybrids. *Tree Physiology* 25: 1475–1486.
- He XZ, Reddy JT, Dixon RA. 1998. Stress responses in alfalfa (*Medicago sativa* L.). XXII. cDNA cloning and characterization of an elicitor-inducible isoflavone 7-*O*-methyltransferase. *Plant Molecular Biology* 36: 43–54.
- Helariutta Y, Elomaa P, Kotilainen M, Seppanen P, Teeri TH. 1993. Cloning of cDNA coding for dihydroflavonol-4-reductase (DFR) and characterization of DFR expression in the corollas of *Gerbera hybrida* var. *regina* (Compositae). *Plant Molecular Biology* 22: 183–193.
- Helariutta Y, Kotilainen M, Elomaa P, Kalkkinen N, Bremer K, Teeri TH, Albert VA. 1996. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. *Proceedings of the National Academy of Science, USA* 93: 9033–9038.
- Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M. 2003. Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry* 278: 95–103.
- Holton TA, Cornish EC. 1995. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7: 1071–1083.
- Ibdah M, Zhang XH, Schmidt J, Vogt T. 2003. A novel Mg<sup>2+</sup>-dependent *O*-methyltransferase in the phenylpropanoid metabolism of *Mesembryanthemum crystallinum*. *Journal of Biological Chemistry* 278: 43961–43972.
- Iiyama K, Lam TBT, Stone BA. 1994. Covalent cross-links in the cell wall. *Plant Physiology* 104: 315–320.

- Ikonen A, Tahvanainen J, Roininen H. 2001. Chlorogenic acid as an antiherbivore defence of willows against leaf beetles. *Entomologia Experimentalis et Applicata* 99: 47–54.
- Johnson ET, Ryu S, Yi H, Shin B, Cheong H, Choi G. 2001. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *Plant Journal* 25: 325–333.
- Kandil FE, Grace MH, Seigler DS, Cheeseman JM. 2004. Polyphenolics in *Rhizophora mangle* L. leaves and their changes during leaf development and senescence. *Trees* 18: 518–528.
- Kao Y-Y, Harding SA, Tsai C-T. 2002. Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. *Plant Physiology* 130: 796–807.
- Ye ZH, Kneusel RE, Matern U, Varner JE. 1994. An alternative methylation pathway in lignin biosynthesis in zinnia. *Plant Cell* 6: 1427–1439.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150–163.
- Lavid N, Wang J, Shalit M, Guterman I, Bar E, Beuerle T, Menda N, Shafir S, Zamir D, Adam Z, Vainstein A, Weiss D, Pichersky E, Lewinsohn E. 2002. O-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. *Plant Physiology* 129: 1899–1907.
- Lindroth RL, Hwang S-Y. 1996. Diversity, redundancy and multiplicity in chemical defense systems of aspen. In: Romeo JT, Saunders JA, Barbosa P, eds. *Phytochemical diversity and redundancy in ecological interactions*. New York, NY, USA: Plenum Press, 25–54.
- Lindroth RL, Scriber JM, Hsia MTS. 1988. Chemical ecology of the tiger swallowtail: mediation of host use by phenolic glycosides. *Ecology* 69: 814–822.
- Marles MAS, Ray H, Gruber MY. 2003. New perspectives on proanthocyanidin biochemistry and molecular regulation. *Phytochemistry* 64: 367–383.
- Métraux JP. 2002. Recent breakthroughs in the study of salicylic acid biosynthesis. *Trends in Plant Science* 7: 332–334.
- Meyermans H, Morreel K, Lapierre C, Pollet B, De Bruyn A, Busson R, Herdewijn P, Devreese B, van Beeumen J, Marita JM, Ralph J, Chen CY, Burggraef B, van Montagu M, Messens E, Boerjan W. 2000. Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. *Journal of Biological Chemistry* 275: 36899–36909.
- Muzac I, Wang J, Anzellotti D, Zhang H, Ibrahim RK. 2000. Functional expression of an *Arabidopsis* cDNA clone encoding a flavonol 3'-O-methyltransferase and characterization of the gene product. *Archives of Biochemistry and Biophysics* 375: 385–388.
- Niggeweg R, Michael AJ, Martin C. 2004. Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology* 22: 746–754.
- Noda N, Kanno Y, Kato N, Kazuma K, Suzuki M. 2004. Regulation of gene expression involved in flavonol and anthocyanin biosynthesis during petal development in lisanthus (*Eustoma grandiflorum*). *Physiologia Plantarum* 122: 305–313.
- Osier TL, Lindroth RL. 2001. Effects of genotype, nutrient availability and defoliation on aspen phytochemistry and insect performance. *Journal of Chemical Ecology* 27: 1289–1313.
- Osier TL, Lindroth RL. 2004. Long-term effects of defoliation on quaking aspen in relation to genotype and nutrient availability: plant growth, phytochemistry and insect performance. *Oecologia* 139: 55–65.
- Osier TL, Lindroth RL. 2006. Genotype and environment determine allocation to and costs of resistance in quaking aspen. *Oecologia*. doi: 10.1007/s00442-006-0373-8.
- Peer WA, Bandyopadhyay A, Blakeslee JJ, Makam SI, Chen RJ, Masson PH, Murphy AS. 2004. Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with altered auxin transport in *Arabidopsis thaliana*. *Plant Cell* 16: 1898–1911.
- Pellegrini L, Geoffroy P, Fritig B, Legrand M. 1993. Molecular cloning and expression of a new class of ortho-diphenol-O-methyltransferases induced in tobacco (*Nicotiana tabacum* L.) leaves by infection or elicitor treatment. *Plant Physiology* 103: 509–517.
- Peltonen PA, Vapaavuori E, Julkunen-Tiitto R. 2005. Accumulation of phenolic compounds in birch leaves is changed by elevated carbon dioxide and ozone. *Global Change Biology* 11: 1305–1324.
- Peters DJ, Constabel CP. 2002. Molecular analysis of herbivore-induced condensed tannin synthesis: cloning and expression of dihydroflavonol reductase from trembling aspen (*Populus tremuloides*). *Plant Journal* 32: 701–712.
- Pierpoint WS. 1994. Salicylic acid and its derivatives in plants: medicines, metabolites and messenger molecules. *Advances in Botanical Research* 20: 163–235.
- Raes J, Rohde A, Christensen JH, van de Peer Y, Boerjan W. 2003. Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiology* 133: 1051–1071.
- Raiber S, Schroder G, Schroder J. 1995. Molecular and enzymatic characterization of 2 stilbene synthases from eastern white-pine (*Pinus strobus*) – a single Arg/His difference determines the activity and the pH-dependence of the enzymes. *FEBS Letters* 361: 299–302.
- Ralston L, Subramanian S, Matsuno M, Yu O. 2005. Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. *Plant Physiology* 137: 1375–1388.
- Raman SB, Rathinasabapathi B. 2003. Beta-alanine N-methyltransferase of *Limonium latifolium*. cDNA cloning and functional expression of a novel N-methyltransferase implicated in the synthesis of the osmoprotectant beta-alanine betaine. *Plant Physiology* 132: 1642–1651.
- Ranjan P, Kao Y-Y, Jiang H, Joshi CP, Harding SA, Tsai C-J. 2004. Suppression subtractive hybridization-mediated transcriptome analysis from multiple tissues of aspen (*Populus tremuloides*) altered in phenylpropanoid metabolism. *Planta* 219: 694–704.
- Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Molecular Biology and Evolution* 16: 266–274.
- Ravn H, Andary C, Kovacs G, Mølgaard P. 1989. Caffeic acid esters as *in vitro* inhibitors of plant pathogenic bacteria and fungi. *Biochemical Systematics and Ecology* 17: 175–184.
- Rohde A, Morreel K, Ralph J, Goeminne G, Hostyn V, De Rycke R, Kushnir S, van Doorselaere J, Joseleau JP, Vuylsteke M, van Driessche G, van Beeumen J, Messens E, Boerjan W. 2004. Molecular phenotyping of the pal1 and pal2 mutants of *Arabidopsis thaliana* reveals far-reaching consequences on and carbohydrate metabolism. *Plant Cell* 16: 2749–2771.
- Schimel JP, van Cleve K, Cates RG, Clausen TP, Reichardt PB. 1996. Effects of balsam poplar (*Populus balsamifera*) tannins and low molecular weight phenolics on microbial activity in taiga floodplain soil: implications for changes in N cycling during succession. *Canadian Journal of Botany* 74: 84–90.
- Schweitzer JA, Bailey JK, Rehill BJ, Martinsen GD, Hart SC, Lindroth RL, Keim P, Whitham TG. 2004. Genetically based trait in a dominant tree affects ecosystem processes. *Ecology Letter* 7: 127–134.
- Shimada N, Sasaki R, Sato S, Kaneko T, Tabata S, Aoki T, Ayabe S. 2005. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the *Lotus japonicus* genome. *Journal of Experimental Botany* 56: 2573–2585.
- Springob K, Lukacin R, Ernwein C, Groning I, Matern U. 2000. Specificities of functionally expressed chalcone and acridone synthases from *Ruta graveolens*. *European Journal of Biochemistry* 267: 6552–6559.
- Stafford HA. 1991. Flavonoid evolution: an enzymatic approach. *Plant Physiology* 96: 680–685.

- Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandré K, Strauss SH, Sundberg B, Gustafsson P, Uhlen M, Bhalerao RP, Nilsson O, Sandberg G, Karlsson J, Lundeberg J, Jansson S. 2004. A *Populus* EST resource for plant functional genomics. *Proceedings of the National Academy of Science, USA* 101: 13951–13956.
- Stich K, Eidenberger T, Wurst F, Forkmann G. 1992. Flavonol synthase activity and the regulation of flavonol and anthocyanin biosynthesis during flower development in *Dianthus caryophyllus* L. (Carnation). *Zeitschrift Fur Naturforschung C* 47: 553–560.
- Tamagnone L, Merida A, Stacey N, Plaskitt K, Parr A, Chang CF, Lynn D, Dow JM, Roberts K, Martin C. 1998. Inhibition of phenolic acid metabolism results in precocious cell death and altered cell morphology in leaves of transgenic tobacco plants. *Plant Cell* 10: 1801–1816.
- Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR. 2003. Proanthocyanidin biosynthesis in plants – purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *Journal of Biological Chemistry* 278: 31647–31656.
- Taylor LP, Grotewold E. 2005. Flavonoids as developmental regulators. *Current Opinion in Plant Biology* 8: 317–323.
- Tsai C-J, Popko JL, Mielke MR, Hu W-J, Podila GK, Chiang VL. 1998. Suppression of *O*-methyltransferase gene by homologous sense transgene in quaking aspen causes red–brown wood phenotypes. *Plant Physiology* 17: 101–112.
- Walsh CT, Liu J, Rusnak F, Sakaitani M. 1990. Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chemical Reviews* 90: 1105–1129.
- Whitham TG, Floate KD, Martinsen GD, Driebe EM, Keim P. 1996. Ecological and evolutionary implications of hybridization: *Populus*–herbivore interactions. In: Stettler RF, Bradshaw HD Jr, Heilman PE, Hinckley TM, eds. *Biology of Populus and its Implications for Management and Conservation*. Ottawa: NRC. Research Press, 247–276.
- Whitham TG, Young W, Martinsen GD, Gehring CA, Schweitzer JA, Shuster SM, Wimp GM, Fischer DG, Bailey JK, Lindroth RL, Woolbright S, Kuske CR. 2003. Community and ecosystem genetics: a consequence of the extended phenotype. *Ecology* 84: 559–573.
- Wildermuth MC, Dewdney J, Wu G, Ausubel FM. 2001. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* 414: 562–565.
- Winkel BSJ. 2004. Metabolic channeling in plants. *Annual Review of Plant Biology* 55: 85–107.
- Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* 126: 485–493.
- Wu S, Watanabe N, Mita S, Dohra H, Ueda Y, Shibuya M, Ebizuka Y. 2004. The key role of phloroglucinol *O*-methyltransferase in the biosynthesis of *Rosa chinensis* volatile 1,3,5-trimethoxybenzene. *Plant Physiology* 135: 95–102.
- Xie D-Y, Jackson LA, Cooper JD, Ferreira D, Paiva NL. 2004. Molecular and biochemical analysis of two cDNA clones encoding dihydroflavonol-4-reductase from *Medicago truncatula*. *Plant Physiology* 134: 979–994.
- Zenk MH. 1967. Pathways of salicyl alcohol and salicin in *Salix purpurea* L. *Phytochemistry* 6: 245–252.
- Zucker WV. 1983. Tannins: does structure determine function? An ecological perspective. *American Naturalist* 121: 335–365.

## Supplementary Material

The following supplementary material is available for this article online:

**Table S1.** List of Arabidopsis and Populus genes and locus information used in this study

**Table S2.** List of primers used for RT-PCR

**Table S3.** List of flavonoid compounds detected in Populus spp.

This material is available as part of the online article from <http://www.blackwell-synergy.com>.



## About New Phytologist

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at [www.newphytologist.org](http://www.newphytologist.org).
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *OnlineEarly* – the 2004 average submission to decision time was just 30 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £109 in Europe/\$202 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office ([newphytol@lancaster.ac.uk](mailto:newphytol@lancaster.ac.uk); tel +44 1524 594691) or, for a local contact in North America, the US Office ([newphytol@ornl.gov](mailto:newphytol@ornl.gov); tel +1 865 576 5261).