Characterization of *O*-Methyltransferase (OMT)-like cDNA clones from *Arabidopsis thaliana* and *Serratula tinctoria*, and Phylogenomic Analysis of Plant OMTs

Kevin Chee Loong Lam

A Thesis

in

The Department

of

Biology

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Biology) at

Concordia University

Montréal, Québec, Canada

April 2005

Canada

ABSTRACT

Characterization of *O*-Methyltransferase (OMT)-like cDNA clones from *Arabidopsis thaliana* and *Serratula tinctoria*, and Phylogenomic Analysis of Plant OMTs.

Kevin Chee Loong Lam

O-Methylation of plant secondary metabolites, which plays important roles in the biosynthesis of various natural products and function of plants, is catalyzed by a superfamily of *S*-adenosyl-L-methionine-dependent *O*-methyltransferases (OMTs). The objectives of the research were to isolate and characterize OMT-like cDNA clones from *Arabidopsis thaliana* and *Serratula tinctoria*, and carry out a phylogenomic analysis of well characterized OMTs to develop a framework phylogenetic tree for predicting putative function and substrate specificity of newly identified gene sequences.

Two OMT-like cDNA clones of *A. thaliana* obtained from TAIR (The *Arabidopsis* Information Resources) were heterologously expressed for molecular and biochemical characterization. The recombinant proteins did not show enzymatic activity with the substrates tested suggesting that these two putative OMTs genes could be pseudogenes.

S. tinctoria accumulates small amounts of 3-methylquercetin as an intermediate for the synthesis of 3,3'-dimethyl derivative. In order to isolate the gene encoding quercetin 3-OMT, a cDNA library of *S. tinctoria* was screened using both, degenerate primers designed based on tryptic peptides of the partially purified and biochemically characterized native protein from *S.* tinctoria, as well as an *Arabidopsis* 3'-OMT full-length cDNA clone as heterologous probes. Although the isolation of a full-length cDNA clone containing all the six tryptic peptides of the native flavonol 3-OMT was not forth

coming, 27 non full-length and three full-length putative OMT-like cDNA clones (StOMT1, StOMT2 and StOMT3) were identified. Recombinant StOMT3 crude protein showed catalytic activity with quercetin and 5-hydroxyferulic acid, but not with 3-methylquercetin. StOMT3 may potentially be the gene encoding flavonol 3-OMT of *S. tinctoria*, but given the lack of all the tryptic peptides, its true nature is unclear and awaits further characterization. Further work is currently in progress for the identification of these putative clones.

Phylogenetic trees were constructed based on amino acid sequence data of 60 biochemically characterized OMTs involved in *O*-methylation of plant natural products in order to develop a framework for predicting the putative function and substrate specificity of newly sequenced genes, and to infer the evolutionary history of OMTs in plant species. Two distinct monophyletic lineages were observed, one consisting of the CCoAOMTs and carboxylic acids utilizing OMTs, and the other comprising the phenylpropanoid COMTs, (iso)flavonoids and alkaloids OMTs. The clustering of OMTs on the phylogenetic trees is corroborated not only on the basis of their sequence homology, but also of the structural and regiospecific similarities of their methyl acceptor molecules.

ACKNOWLEDGEMENT

DEDICATION

To my mother for her love.

"I can do everything through Him who gives me strength."

Philippians 4:13

TABLE OF CONTENTS

A. LIST OF FIGURES

B.   LIST OF TABLES

C.   INTRODUCTION

Throughout the course of evolution, plants have acquired complex mechanisms to compensate for their inability to relocate as a defense strategy. In addition to the primary metabolites required for growth and development, plants have developed a sophisticated mechanism for the synthesis of a wide spectrum of low molecular weight organic compounds, collectively known as secondary metabolites. Of these, alkaloids, flavonoids and terpenoids are the most ubiquitous, and they function primarily in enhancing plant defense and reproductive success. Flavonoids, in particular, play important roles in such processes as flower pigmentation, pollen fertility, signal transmission in symbiotic interactions, and as defense compounds with antimicrobial and antifungal activity. Flavonoids are also well known for their potential benefit to human health (Bohm, 1998 and references therein).

The structural diversity found in plant natural products is attributed to a number of substitution reactions, which are catalyzed by substrate-specific, position-oriented enzymes (Ibrahim and Anzellotti, 2003 and references therein). Of these, enzymatic O-methylation, catalyzed by S-adenosyl-L-methionine (AdoMet)-dependent O-methyltransferases, plays a major role in the biosynthesis of secondary metabolites (Ibrahim and Muzac, 2000). Several O-methylated compounds are involved in plant processes such as microbial signaling (Maxwell *et al.*, 1989, 1993), lignification (coniferyl and sinapyl alcohols), chemical defense (phytoalexins), and increased antiviral/antimicrobial activities (Bohm, 1998 and references therein). Plant OMTs could be classified into various classes based on their substrate preferences. In general, OMTs

1

are substrate specific, position-oriented enzymes, as demonstrated with a number of distinct enzymes catalyzing the stepwise *O*-methylation in *Chrysosplenium americanum* (Ibrahim *et al.*, 1987). In other cases, however, plant OMTs have been shown to exhibit a broader substrate specifity utilizing substrates from different classes of secondary metabolites (Gauthier *et al.*, 1998; Frick and Kutchan, 1999; Ibdah *et al.*, 2003).

Much of the work involving flavonoids has focused on methylated flavonoids due to their intrinsic properties as phytoalexins/phytoanticipins, as well as their antimicrobial/antiviral activities. These low molecular weight compounds are synthesized in response to pathogen attack and other stress-related conditions. So far, several OMTs involved in the biosynthesis of isoflavonoids and pterocarpans have been characterized at the biochemical and/or molecular levels. Recently, quercetin 3-*O*-methyltransferase was partially purified, biochemically characterized from *Serratula tinctoria* (Asteraceae), and its internal amino acid sequence information made available (Huang *et al.*, 2004). 3-Methylquercetin has been reported to accumulate, as a phytoanticipin, in response to wounding and insect herbivory in tobacco (Roda *et al.*, 2003), and to act as an antiinflammatory and antiviral agent (Middleton and Kandaswami, 1994; Malhotra *et al.*, 1996), as well as a selective inhibitor of cAMP- and cGMP- phosphodiesterases (PDE) (Ko *et al.*, 2003). 3-Methylquercetin also exhibits inhibitory effects on viral RNA replication (Castrillo and Carrasco, 1987). As an intermediate metabolite, it rarely accumulates under normal conditions. However, it serves as the precursor in the stepwise *O*-methylation of various partially and polymethylated flavonoids. This includes the biosynthesis of 3,3'-dimethylquercetin in *Serratula tinctoria* (Dedaldechamps and Ibrahim, unpublished), 3,7,4'-trimethylquercetin

in apple cell culture (Macheix and Ibrahim, 1984), Calamondin orange peel (Brunet and Ibrahim, 1980), and spinach leaves (Thresh and Ibrahim, 1985), as well as in polymethylated flavonoid synthesis in *Chrysosplenium americanum* (Ibrahim *et al.*, 1987).

The advent of recombinant DNA technology and cell culture technology have facilitated the application of genetic engineering for manipulation of plant secondary metabolism, thus providing the potential to increase productivity, alter the quality traits of food, fodder or horticulturally-valuable plants, as well as enhance the resistance of plants against pests and disease (Sato *et al.*, 2001; Hain and Grimmig, 2000). However, a comprehensive understanding of the regulation of biosynthetic pathways, at both the gene and protein levels, is needed to effectively apply the various genetic engineering approaches. Ideally, this requires the elucidation of all steps involved in the formation of metabolic intermediates and the enzymes involved in their biosynthesis. Considerable efforts have recently been directed towards identifying the genes involved in the biosynthesis of flavonoids including their *O*-methylation (Ibrahim and Muzac, 2000) and the understanding of their biosynthetic regulation.

The *Arabidopsis thaliana* (Brassicaceae) genomic initiative offers a more expeditious means to uncover putative OMT genes in *Arabidopsis* through a direct molecular genetics approach bypassing lengthy enzyme purification and identification procedures. However, the enormous pool of gene sequences present in these mega genomic databases, primarily represent genes with only putative function with little or no experimentally-documented function of their gene products. The identification of putative OMT genes has traditionally been based on their sequence homology to

3

experimentally-characterized OMT gene(s). This technique, in essence, implies sequence similarity as being similarity of function, which is not always the case, and hence can result in the incomplete or erroneous annotation of putative OMTs.

Recently, however, phylogenomic approaches have become a more powerful and reliable tool in enzyme function prediction (Eisen and Wu, 2002). This approach acknowledges that the identification of sequence similarity alone does not reliably reflect an identical function for an uncharacterized gene. It takes advantage of an evolutionary perspective: molecular phylogeny, in the analysis of the sequences being compared. In addition, phylogenomics provides the means with which to reconstruct the evolutionary history of OMTs between plant species, with the intention of inferring whether this gene superfamily has evolved from a common ancestor. The reconstruction of the phylogenetic tree(s) would enable the development of a framework for predicting the putative function and substrate specificity, as well as a classification of newly sequenced genes for the study of their evolutionary relatedness.

D.   AIM OF THE RESEARCH

1. Molecular and Biochemical Characterization of *Arabidopsis thaliana* (Brassicaceae) OMT clones. The goal is to isolate or acquire full length cDNA clones encoding putative OMTs involved in phenolic/flavonoid metabolism of *Arabidopsis*. The concerted efforts in the genomic sequencing project of *Arabidopsis* have created a remarkable opportunity not only to study the evolution and molecular mechanism leading to the diversification of the OMT superfamily, but also provided an expeditious means for the characterization and cloning of putative OMT cDNAs in *Arabidopsis*. This will enable the elucidation of the role of OMTs in various biosynthetic pathways, an important prerequisite step towards genetic engineering.

2. Screening of *Serratula tinctoria* (Asteraceae) cDNA Library. The goal is to isolate a full length cDNA clone encoding the flavonol 3-OMT of *S. tinctoria* for further molecular and biochemical characterization. Given the important roles that 3-methylquercetin plays in plants and its potential uses in human welfare as an antiviral agent, the cloning of the gene encoding flavonol 3-OMT would be of considerable interest for its manipulation in plants. The expression of the recombinant protein would also facilitate the study of its response to plant stress as well as allow the constitutive expression of 3-methylquercetin in economically valuable crops.

3. Phylogenomic Analysis of OMTs. The goal is to analyze biochemically characterized OMT sequences in an evolutionary context to develop a framework phylogenetic tree for predicting the putative function and substrate specificity of newly sequenced genes. This will be invaluable for guiding future research endeavors, and for infering the evolutionary history of OMT gene superfamily.

E. LITERATURE REVIEW

E.1 Functions of Flavonoids in Nature and Their Potential Value to Human Welfare

Flavonoids constitute one of the major groups of plant secondary metabolites; not only because of their ubiquitous occurrence but also for the important roles they play in plant growth, development and adaptation to a changing environment. Chalcones, flavanones, flavones, isoflavones, flavonols and anthocyanidins comprise the major flavonoid subclasses (Figure 1). As a result of this structural diversity, flavonoids can impart numerous physiological activities (Croteau *et al.*, 2000).

Flavonoids are best known as the characteristic red, blue, and purple anthocyanin pigments of plant tissues and flower petals. These compounds serve essential functions in plant reproduction by recruiting pollinators and seed dispersers (Brouillard and Dangles, 1994), protection against UV-B radiation, and as insect feeding attractants, or deterrants (Croteau *et al.*, 2000). Flavonoids also serve as regulators of auxin transport (Faulkner and Rubery, 1992), determinants of pollen tube growth and pollen functionality (Ylstra *et al.*, 1992), and as signal molecules in legume-rhizobium bacterial interactions, thus facilitating nitrogen fixation (Maxwell *et al.*, 1989; Maxwell *et al.*, 1993).

Methylated flavonoids play an active role in many of the processes described above, particularly in their involvement in the *de novo* biosynthesis of phytoalexins. These methylated compounds can act as precursors (i.e. isoflavones) or end products (i.e. pterocarpans), exhibiting antimicrobial or antifungal activities (Dakora and Phillips, 1996; He *et al.*, 1998). Formononetin (7-hydroxy-4'-methoxyisoflavone), for instance, is an essential
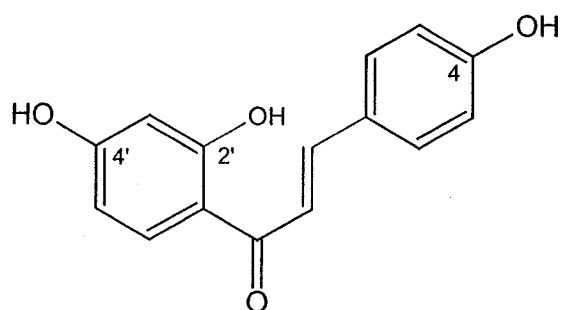
Chalcone
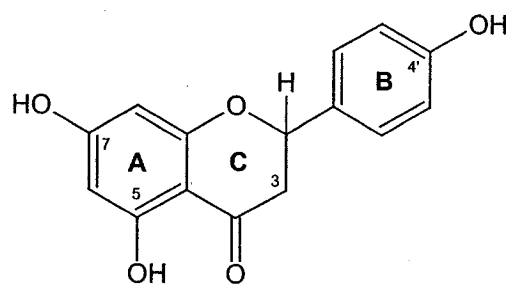
Flavanone
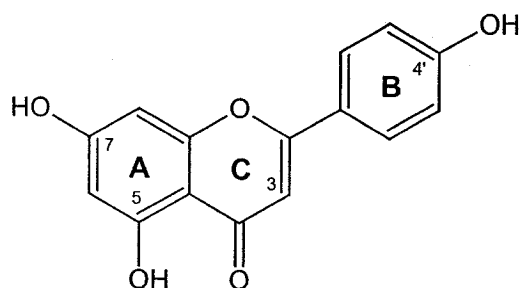
Flavone
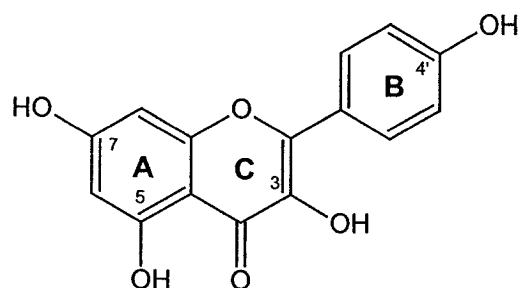
Flavonol
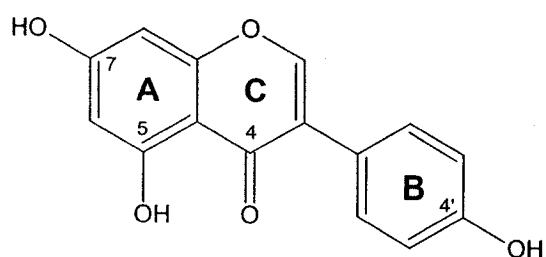
Isoflavone
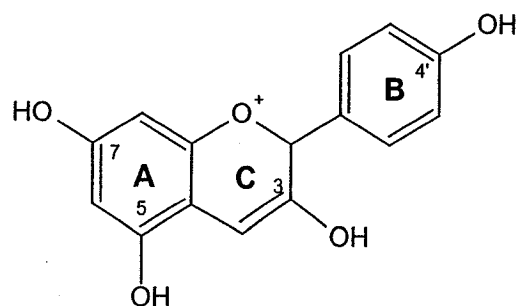
Anthocyanidin

Figure 1: The major classes of flavonoids.

8

intermediate in the biosynthesis of the phytoalexins, namely medicarpin, vestitol, pisatin, and maackiain (Dixon, 1999).

Furthermore, studies assessing effects of flavonoids in mammalian systems revealed that these compounds act as modulators of the immune and inflammatory responses, as antioxidants, as antimicrobial/antiviral agents, as hepatoprotective agents, as enzyme inhibitors, and as anticancer compounds (Middleton and Kandaswami, 1994). Currently, there is a considerable interest in the use of isoflavonoids in cancer prevention. For example, the methylated isoflavonoids of soybean, formononetin (7-hydroxy-4'-methoxyisoflavone) and biochanin A (5,7-dihydroxy-4'-methoxyisoflavone), have been implicated in the inhibition of the growth of human breast cancer cells (Peterson and Barnes, 1991) due to their roles as estrogenic and anticarcinogenic agents (Sarkar and Li, 2003). Interestingly, 3-methylquercetin has also been shown to possess antispasmodic (Karamenderes and Apaydin, 2003) and antiinflammatory activities (Ko *et al.*, 2003). The potential health benefits of fruits, vegetables, green tea and red wine might partly be due to these flavonoid properties as well as other phytochemicals and, as such, have received considerable attention as health-promoting components of human diet and potential for pharmacological use.

E.2   Biosynthesis of Flavonoids

Flavonoids are synthesized from L-phenylalanine as part of the larger phenylpropanoid pathway. Structurally, they consist of two phenolic rings, A and B, connected by a 3-carbon side chain that forms the heterocyclic ring C (Figure 2). The A-ring is derived from malonyl CoA via the polyketide pathway, whereas ring B and the
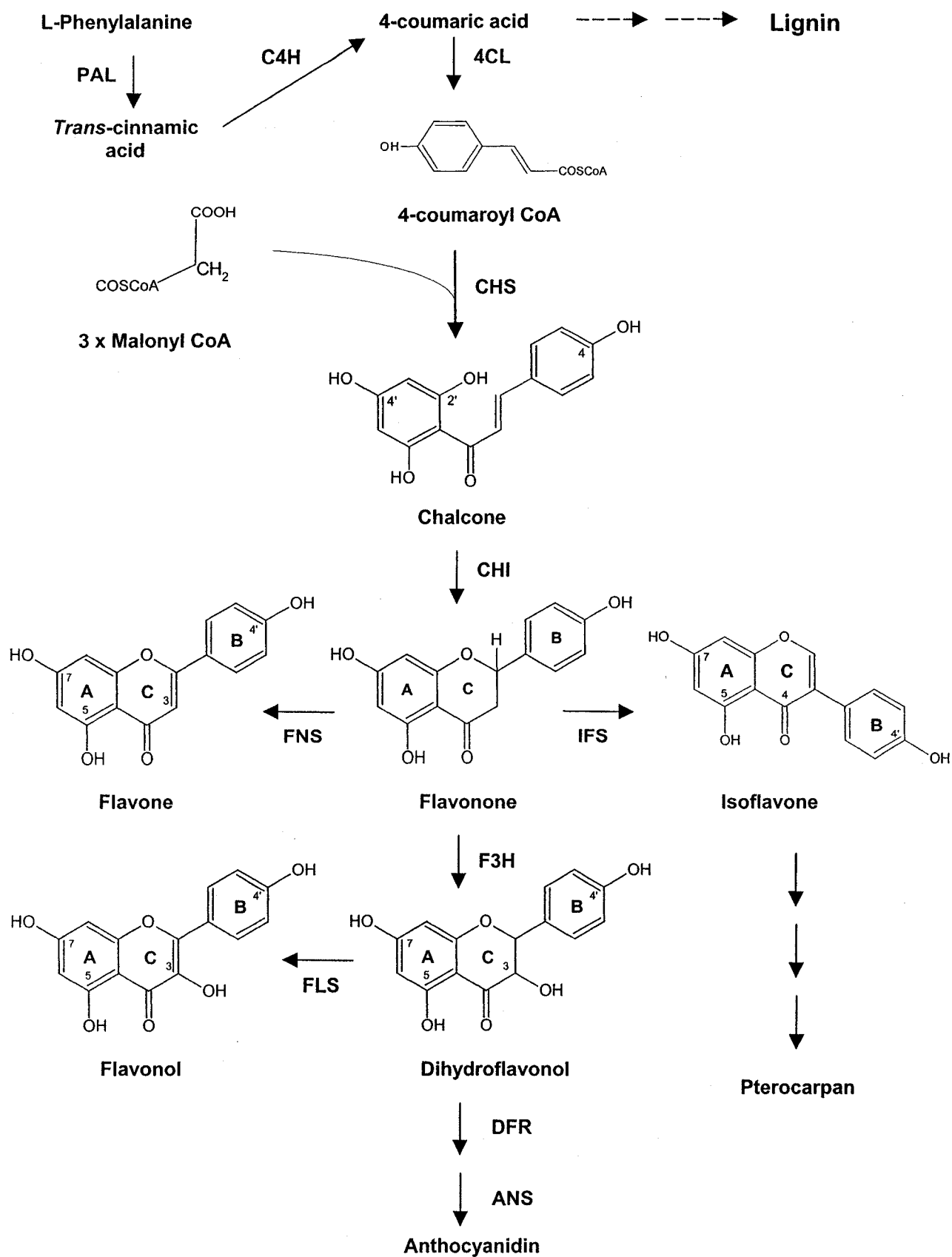
Figure 2: The flavonoid biosynthetic pathway.

3-carbon side chain are derived from L-phenylalanine via the shikimate pathway. The biosynthesis of flavonoids begins with the conversion of phenylalanine to *trans*-cinnamic acid, then to 4-coumaric acid, catalyzed by phenylalanine ammonia lyase (PAL) and cinnamate 4-hydroxylase (C4H), respectively. As a branch-point molecule, 4-coumaric acid can either be committed to a pathway leading to the formation of monolignols involved in lignin bioynthesis, or subsequently esterified with coenzyme A to form 4-coumaroyl CoA, a reaction catalyzed by 4-coumaroyl CoA ligase (4CL). The first committed step in the biosynthesis of flavonoids (Figure 2) involves the stepwise condensation of one molecule of 4-coumaroyl CoA with three molecules of malonyl CoA, catalyzed by chalcone synthase (CHS). This results in the formation of naringenin chalcone (4,2',4',6'-tetrahydroxychalcone), the first stable intermediate of this pathway. Depending on the plant species, chalcone synthase (CHS) may coact with chalcone reductase (CHR) to form the 5-deoxychalcone isoliquiritigenin (4,2',4'-trihydroxychalcone). Beyond CHS, the next step common to most flavonoid biosynthetic pathways is catalyzed by chalcone isomerase (CHI), which catalyzes a stereospecific ring closure forming the 2S-flavanone, naringenin or liquiritigenin.

Flavanones represent the most important branch-point intermediates leading to the formation of other subclasses of flavonoids (Figure 2). Hydroxylation of the C3-position of flavanones produces dihydroflavonols. One pathway involves a stereospecific 3-hydroxylation of naringenin, catalyzed by flavanone 3-hydroxylase (F3H), leading to the formation of dihydroflavonols. Their subsequent dehydrogenation, catalyzed by flavonol synthase (FLS), gives rise to flavonols. Another pathway involves the catalytic desaturation of naringenin creating a C2-C3 double bond, by flavone synthase (FNS),

11

giving rise to flavones. Depending on the plant species, some plants utilizes the dioxygenase flavone synthase I (FNS I), while others utilize the monooxygenase, flavone synthase II (FNS II). The colored pigments, anthocyanidins, are synthesized from dihydroflavonol precursors through the sequential activities of dihydroflavonol reductase (DFR) and anthocyanidin synthase (ANS). The third pathway involves the formation of isoflavonoids through the activities of isoflavone synthase (IFS) and 2-hydroxyisoflavanone dehydratase (IFD). The former enzyme catalyzes a unique C2 to C3 aryl migration and hydroxylation to give rise to 2-hydroxyisoflavanone, which is then acted upon by the latter enzyme to form the isoflavonoids genistein and daidzein. These bioactive compounds are precursors of the pterocarpan phytoalexins commonly found in legume species. For a detailed review on the subject, see Dixon (1999) and Forkmann and Heller (1999).

E.3 Elaboration of Flavonoid Skeletons: *O*-Methylation

Flavonoids owe their numerous biological activities to their extensive structural diversity resulting from a variety of enzyme-catalyzed substitution reactions. From the basic flavonoid structure a wide range of derivatives are formed through hydroxylation, methylation, glycosylation, acylation, prenylation, and sulfonation reactions. For example, hydroxylation of both aromatic rings (A and B) and methylation of their hydroxyl groups leads to the formation of the various aglycones within each flavonoid class. Furthermore, glycosylation of hydroxyl groups can lead to the formation of glycosides which can be further modified to produce di- and oligoglycosides. Glycosylation reactions have been shown to, not only enhance water solubility of the

metabolites but also serve, in combination with acylation, as recognition signals for membrane transport and vacuolar storage as in the case of many anthocyanins. Substitution reactions are common and can occur practically across all flavonoid classes/subclasses. For a detailed account of this topic, see reviews by Heller and Forkmann (1994) and Ibrahim and Anzellotti (2003).

Among these modification reactions, O-methyation plays an important role in its contribution to the understanding of these substitution reactions mainly because it modifies the solubility of flavonoids and their activity. O-Methyation of flavonoids is catalyzed by a family of S-adenosyl-L-methionine (AdoMet)-dependent methyltransferases [OMTs; E.C.2.1.1.6]. O-Methyltransferases, ubiquitous enzymes present across phyla, catalyze the transfer of the methyl group of AdoMet to the hydroxyl group of an acceptor molecule with the concomitant formation of the corresponding methylated derivative and S-adenosyl-L-homocysteine (AdoHcy) as products. Enzymatic O-methylation takes place in all classes/subclasses of flavonoids and occurs on all possible hydroxylated positions on the A and B-ring, the heterocyclic ring, and the 5-hydroxyl group of isoflavones (Khoury et al., 1986). In plants, enzymatic O-methylation of flavonoids not only alters its physiological properties by reducing the chemical reactivity of their phenolic hydroxyl groups, but also modifies their water/lipid solubility, thereby affecting their intracellular compartmentation and the metabolic pathways in which they participate (Laks and Pruner, 1989; Poulton, 1981).

E.4 *O*-Methyltransferases

Interestingly, with more than 80,000 known secondary metabolites (Verpoorte, 2000), it is reasonable to assume that the enzymes catalyzing *O*-methylation reactions possess a broad range of substrate specificities, as is the case with many mammalian methyltransferases. Nevertheless, plant OMTs, despite methylating a diverse spectrum of compounds, have evolved a high degree of selectivity and regiospecificity for their methyl acceptor molecules. This is evident in the substitution pattern of their phenolic hydroxyl groups, as illustrated by the position-specific enzymes in *Chrysosplenium americanum* (Ibrahim *et al.*, 1987; Gauthier *et al.*, 1996) and *Mentha piperata* (Willits *et al.*, 2004), both of which produce a variety of partially methylated flavonoids. In certain cases however, plant OMTs have been shown to operate as multifunctional enzymes, acting on different classes of secondary metabolites, such as phenylpropanoids and flavonoids (Gauthier *et al.*, 1998), catechol, benzylisoquinoline alkaloids, and phenylpropanoids (Frick and Kutchan, 1999), and both caffeoyl esters and flavonoids (Ibdah *et al.*, 2003).

In the past decade, the number of putative OMT cDNA clones have grown in public databases. The corresponding OMT cDNA gene products that have been functionally characterized have revealed that despite their substrate specificities in plants, they share a high amino acid sequence similarity. Furthermore, five consensus motifs (regions I to V, Figure 3) located mostly near the carboxy terminal have been identified for most, if not all plant OMTs (Kagan and Clarke, 1994; Ibrahim, 1997; Ibrahim *et al.*, 1998; Joshi and Chiang, 1998). These motifs are glycine-rich (ca. 28%) and contain several hydrophobic residues. Two additional motifs, specific only to caffeic acid
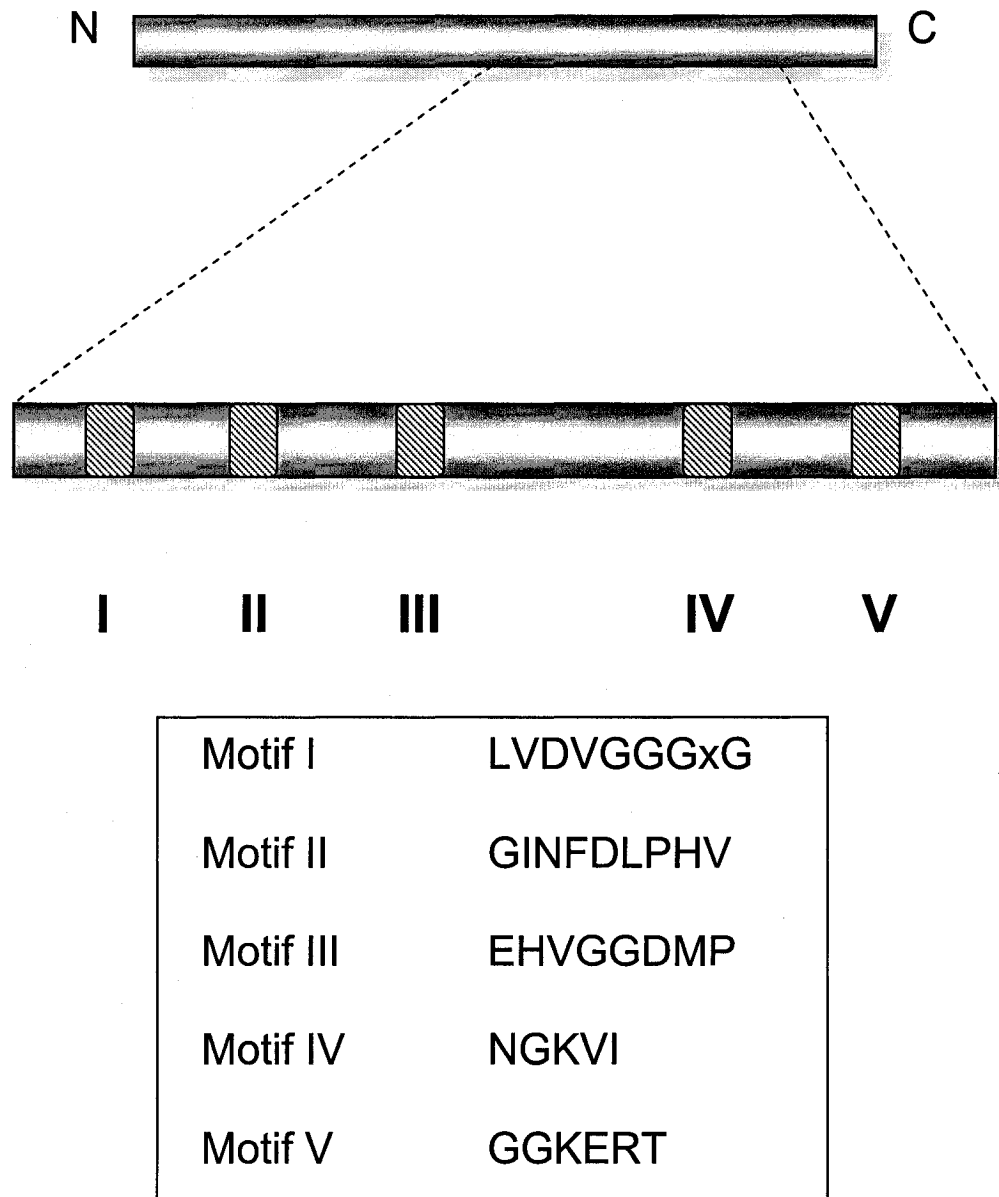
Figure 3: The five conserved signature motifs of plant OMTs.

3-OMTs (COMTs), were also recognized (Selman-Housein *et al.*, 1999). Motifs I and IV have been proposed as the binding sites for AdoMet and the metal ion, respectively (Ibrahim, 1997). These findings indicate that, in general, the methytransferase gene superfamily may contain one or more of these motifs and the degree of sequence homology among them is related to the extent of similarity within the consensus motifs (Ibrahim *et al.*, 1998). Accordingly, plant OMTs constitute a distinct superfamily whose members share at least the five signature motifs (Ibrahim *et al.*, 1998 and references therein; Ibrahim and Muzac, 2000). The elucidation of these common signature motifs will undoubtedly aid in the identification of other yet undiscovered OMT sequences. The crystal structures of two OMTs, COMT and chalcone OMTs, from alfalfa (Zubieta *et al.*, 2001) identified two other motifs located at both the amino and carboxyl ends that are believed to participate in substrate binding.

E.4.1 Classification of *O*-Methylatransferases

The biochemical, molecular and protein 3D-crystal structure information available thus far can provide the basis for different classification schemes. Amino acid sequence alignments and subsequent phylogenetic analyses have demonstrated a good correlation between the extent of sequence identity of these proteins and the different structural classes of their methyl acceptor molecules, such as alkaloids, flavonoids, and phenylpropanoids (Ibrahim *et al.*, 1998; Joshi and Chiang, 1998; Ibrahim and Muzac, 2000). In general, plant OMTs can be divided into two major classes (Joshi and Chiang, 1998). The first class (Group I) constitutes OMTs (CCoAOMT/CCOMT) that utilize the CoA esters of caffeic acid and/or 5-hydroxylferulic acid as preferred substrates. This

class of OMTs resembles, at the physico-chemical level, the mammalian catechol OMTs in having a smaller molecular weight of ~27 to 30 kDa, being divalent cation-dependent (Mg^{2+}) and containing a semblance of the signature motifs characteristic of Group II OMTs. Sequence analyses have shown that Group I OMTs do, however, possess their own set of conserved motifs (Joshi and Chiang, 1998). The second class of OMTs (Group II) consists of OMTs that accepts a broader range of substrates, including alkaloids, flavonoids, and phenylpropanoids. Group II OMTs share common features, including similar molecular mass (~38 to 42 kDa), exhibit no Mg^{2+} requirement for catalytic activity, a high amino acid sequence similarity, and the presence of the five conserved motifs within their protein sequence (Ibrahim *et al.*, 1998).

The elucidation of the protein 3D-crystal structure of rat liver catechol OMT (Vidgren *et al.*, 1994), three plant flavonoid OMTs, the chalcone, isoflavone, and caffeic/5-hydroxyferulic acid OMT (COMT) from alfalfa (Zubieta *et al.*, 2001, 2002), as well as the salicylic acid carboxyl OMT (Zubieta *et al.*, 2003), revealed the nature of the amino acids involved in binding AdoMet and Mg^{2+}, as well as those involved in substrate binding and catalysis. These protein X-ray crystallography studies have not only confirmed the presence of these two classes of OMTs, but have also contributed to the proposal by Zubieta and colleagues (2003) of an entirely new class of OMTs that prefer carboxylated substrates. Members of this enzyme group catalyze the methylation of the aliphatic hydroxyl moiety of carboxyl groups on a number of compounds, mainly benzoic, salicylic, and jasmonic acids, whose methylated products are involved in floral scent production or as signaling molecules.

Based on elaborate computer modeling analyses, Zubieta *et al.* (2001), (2002) and (2003) have established that the chemical determinants within the active sites of OMTs play critical roles in substrate discrimination. The identification of the specific amino acid residues lining the active site cavity have enabled Zubieta and colleagues to differentiate between residues that participate directly in the catalytic cycle from other peripheral residues that are only responsible for the overall architecture of the catalytic pocket. These include the extensive hydrogen bonding and van der Waals interaction that sequester the cosubstrate methyl donor and position its methyl group near the hydroxylated acceptor moieties for transmethylation. Notably, structural studies provide a template for deciphering the catalytic mechanism that underlie substrate recognition in different classes of OMTs and hence shed light on substrate prediction (Zubieta *et al.*, 2003). Furthermore, Zubieta and colleagues also found that dimerization of enzyme monomers, in certain cases, also contributes significantly to substrate selection. The changes of specificities through the formation of heterodimers of similar OMT proteins have also been reported by Frick and Kutchan (1999), who cloned and characterized OMTs methylating both alkaloids and phenylpropanoids pathway.

E.4.2   OMTs Involved in Phenylpropanoid Biosynthesis

Lignin is a major constituent of cell walls and tracheary elements, providing cells with rigidity for structural support and impermeability to water. It also functions as a defense barrier in response to wounding and pathogen attack (Boudet, 1998). Lignin is synthesized by the oxidative polymerization of three phenylpropanoid alcohol monomers: p-hydroxyphenyl, guaiacyl (G), and syringyl (S) units. The G/S ratio in lignin

composition determines its forage digestibility by ruminant animals and degradability as a pulping material, and is thus removed during pulp and paper processing (Dean and Eriksson, 1992). Therefore, the modification of lignin content and/or composition by genetic manipulation would be of significant economic interest.

Among the different enzymatic steps involved in the biosynthesis of these monolignols, methylation of the 3- and 5- hydroxyls of monolignol precursors is critical in determining the content and composition of lignin. The monolignol biosynthetic pathway has been controversial (see review by Ye *et al.*, 2001), but the current view entails a phenylpropanoid "metabolic grid" (Figure 4) leading to G and S units. This model proposes that successive hydroxylation and *O*-methylation reactions occur at various levels of side-chain oxidation, though there may be kinetically favored paths (see review by Dixon *et al.*, 2001). Alternatively, Dixon *et al.* (2001) has challenged this grid model and suggested another model in which metabolic channeling allows for independent pathways to G and S lignin involving various membrane-associated enzymes acting in a concerted manner.

Several COMTs and CCoAOMTs have been biochemically characterized and cloned from a variety of plant species. COMTs utilize caffeic acid and 5-hydroxyferulic acid as substrates, to give rise to the monolignol precursors, ferulic and sinapic acids, respectively (Dixon *et al.*, 2001). In contrast, CCoAOMTs utilize, as preferred substrates, the CoA esters of caffeic and/or 5-hydroxyferulic acids and, in certain cases, caffeoyl glucose may act as a better acceptor than the CoA esters (Ibdah *et al.*, 2003).

Originally, it was believed that methylation of caffeic/ferulic acids were carried out by a 'bifunctional/bispecific' COMT, albeit to different degrees. However,

19
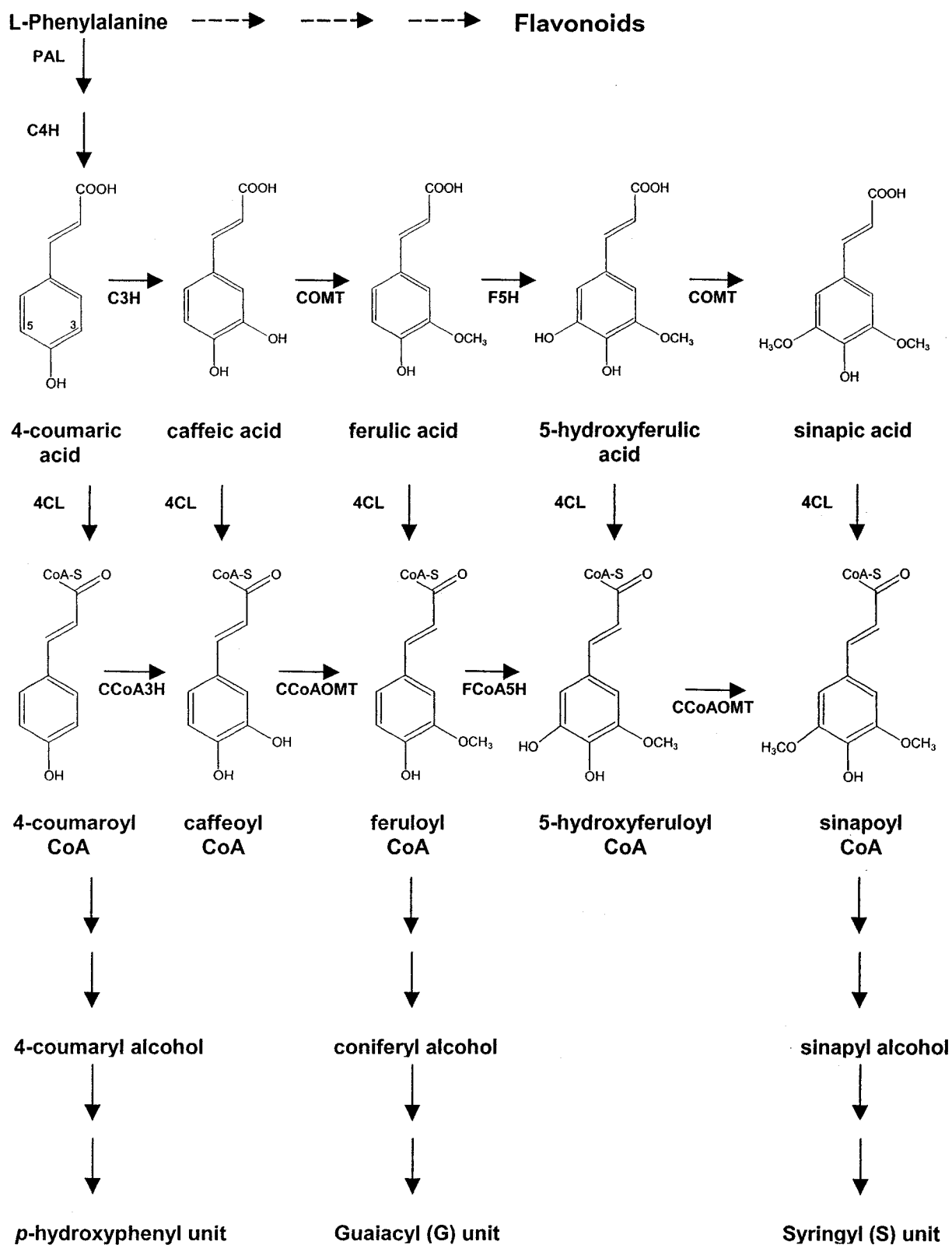
Figure 4: Pathway to monolignols biosynthesis as proposed by Dixon *et al.*, 2001. F5H: ferulate 5-hydroxylase; C3H: coumarate 3-hydroxylase; CCoA3H: coumaroyl CoA 3-hydroxylase; 4CL: CoA ligase; FCoA5H: feruloyl CoA 5-hydroxylase.

De Carolis and Ibrahim (1989) have reported the purification of two isoforms (I and II) of a COMT from *Sinapis alba*, which exhibited distinct substrate preferences. The differential activity of both isoforms towards caffeic acid and ferulic acid is indicative of two distinct enzymes, each having a specific preference for its respective substrate, while at the same time, accepting the other substrate with a reduced affinity. The first cloned COMT (Accession #: X62096) was reported from aspen and is regarded as the type-member of this group (Bugos *et al.*, 1991). Subsequently, other COMTs from alfalfa, poplar, and several other species, including gymnosperms, were reported (Ibrahim and Muzac, 2000 and references therein). These enzymes are typical Group II OMTs, given the presence of almost all of the five signature motifs and no Mg^{2+} requirement for catalytic activity. Exceptions however exist, as demonstrated by several Group II OMTs from *Thalictrum tuberosum* (Frick and Kutchan, 1999) that utilize a wide range of substrates including simple catechol, as well as benzylisoquinoline alkaloids and phenylpropanoids.

The first cloned CCoAOMT (M69184), the type-member of this group, was isolated from elicitor-treated parsley cell cultures (Schmitt *et al.*, 1991). This clone encodes an OMT involved in the methylation of caffeoyl CoA esters. Later, another enzyme (U13151) from *Zinnia* showed that CCoAOMT can accept 5-hydroxyferuloyl CoA as a substrate as well as caffeoyl CoA to the same degree (Ye *et al.*, 1994). To date, a number of CCoAOMT cDNA clones have been isolated from grape (Z54233) (Busam *et al.*, 1997), *Stellaria* (L22203) (Zhang and Chinnappa, 1997), and tobacco (U38612) (Martz *et al.*, 1998) among several other species. Interestingly, Li *et al.* (1997) reported the isolation from loblolly pine (U39301) of a xylem cDNA encoding an enzyme

involved in the methylation of both caffeic and 5-hydroxyferulic acids, as well as their CoA esters to a similar extent. Functionally, this OMT resembles that of a CCoAOMT, but it harbors the five signature motifs characteristic of Group II OMTs with a molecular weight of 42 kDa. The versatile nature of this loblolly pine OMT suggests that it may, in fact, be another member of Group II OMTs. This idea is further supported by the isolation of an OMT cDNA clone from stress-induced Scots pine (Chiron et al., 2000) that has been reported to encode a multifunctional OMT methylating the stilbene pinosylvin, caffeic acid, 5-hydroxyferulic acid, caffeoyl CoA, catechol, and quercetin among other substrates. This pine OMT exhibits all the physico-chemical properties of Group II OMTs.

Another group of phenylpropanoids include the phenylpropenes, chavicol, eugenol and isoeugenol. These are characterized by the absence of a carboxyl function in the 3-carbon side chains, and methylation usually occurs on a hydroxyl group *para* to the side chain. These methylated compounds constitute an important component of floral scent of *Clarkia* species. Recently, cDNA clones encoding OMTs responsible for this methylation have been reported from sweet basil (Gang et al., 2002).

E.4.3 OMTs Involved in Polymethylated Flavonol Biosynthesis

Sequential methylation of flavonols in *Chrysosplenium americanum* represents one of the most elaborate methylation systems in plants. This semi-aquatic weed accumulates a variety of tetra- and penta-*O*-methylated flavonol glucosides, but not the intermediates with lower (mono-, di- or tri-) level of methylation (Collins et al., 1981).

22

This plant seems to accumulate highly methylated metabolites to protect itself against pathogen attack and UV damage.

The initial steps of the biosynthetic pathway (Figure 5) involve the stepwise O-methylation of the parent aglycone, quercetin to give rise to 3-methylquercetin, 3,7-dimethylquercetin, and 3,7,4'-trimethylquercetin. Each of these enzymatic steps is catalyzed by a distinct substrate-specific, position-oriented O-methyltransferase (De Luca and Ibrahim, 1985a; Ibrahim *et al.*, 1987). At this point, 3,7,4'-trimethylquercetin can undergo either of two separate hydroxylations, one of which results in the production of 3,7,4'-trimethyl-2'-hydroxyquercetin, which will ultimately be glucosylated. The second hydroxylation occurs at position 6 of 3,7,4'-trimethylquercetin (Anzellotti and Ibrahim, 2004) giving rise to an intermediate for subsequent methylation and hydroxylation, producing 3,6,7,4'-tetramethyl-2'-hydroxyquercetagetin. O-Glucosylation of position 2' and 5' is catalyzed by two distinct glucosyltransferases (Latchinian *et al.*, 1987; Latchinian and Ibrahim, 1989). Although O-glycosylation is considered a terminal step in flavonoid biosynthesis, it is interesting to note that both glucosylated products can be further methylated resulting in the formation of 2'- and 5'-glucosides of 3,7,4',5'-tetramethylquercetin and 3,6,7,2',4'-pentamethylquercetagetin, respectively, as final products (Khouri *et al.*, 1986).

The 3-OMT enzyme activity, the first committed step in the sequential methylation of flavonols, has also been reported in spinach leaves (Thresh and Ibrahim, 1985) and *Serratula tinctoria* (Huang *et al.*, 2004) among other species. Flavonol OMTs methylating positions 3, 6, 7, 4', and 5' in this pathway have all been partially purified and biochemically characterized from *C. americanum* (De Luca and Ibrahim, 1985a, b;
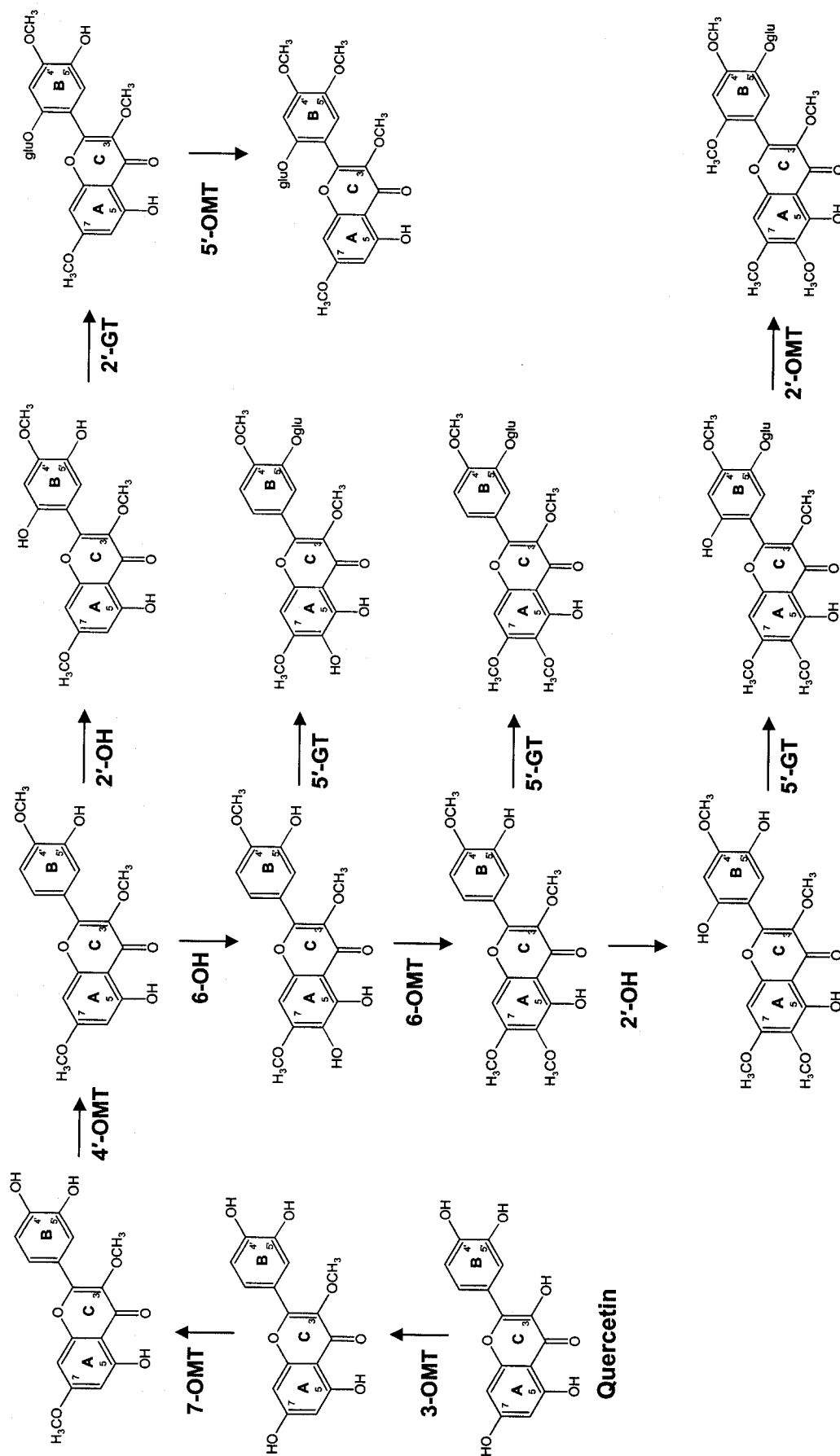
Figure 5: Polymethylated flavonol biosynthesis in *C. americanum*. OH: hydroxylase; GT: Glucosyltransferase.

24

Khouri *et al.*, 1986; Ibrahim *et al.*, 1987). The cDNA clone encoding the 3'/5'-OMT from the same species has also been isolated and characterized (Gauthier *et al.*, 1996).

E.4.4   OMTs Involved in Isoflavonoid Biosynthesis: Phytoalexin Biosynthesis

Phytoalexins are low-molecular weight antimicrobial compounds that are synthesized *de novo* and accumulated by plants as a result of biotic or abiotic elicitation. Many isoflavone OMTs (IOMTs) are involved in the biosynthesis of phytoalexins, such as medicarpin and pisatin. These isoflavonoids, limited primarily to the Leguminosae, serve as precursors (isoflavones) or final products (pterocarpans). They exhibit estrogenic, anti-angiogenic, antimicrobial and anti-cancer activities. As a result, they have received considerable attention as health-promoting components of the human diet (Dixon, 1999; Dixon and Steele, 1999).

Medicarpin, a major pterocarpan phytoalexin in alfalfa, is synthesized via the isoflavonoid branch of phenylpropanoid metabolism. Much of this pathway has recently been elucidated through characterization of the specific enzymes involved and the cloning of their genes. The 4'-*O*-methylation of the B-ring of isoflavones in certain legume species is the prerequisite reaction for further elaboration of the isoflavonoid nucleus to form pterocarpans from isoflavone intermediates, such as daidzein in alfalfa and genistein in chickpea (Dixon, 1999 and references therein). However, the nature of the enzymatic step leading to the introduction of the methyl group at position 9 of medicarpin (4'-position in isoflavones) is still a matter of debate (Figure 6).
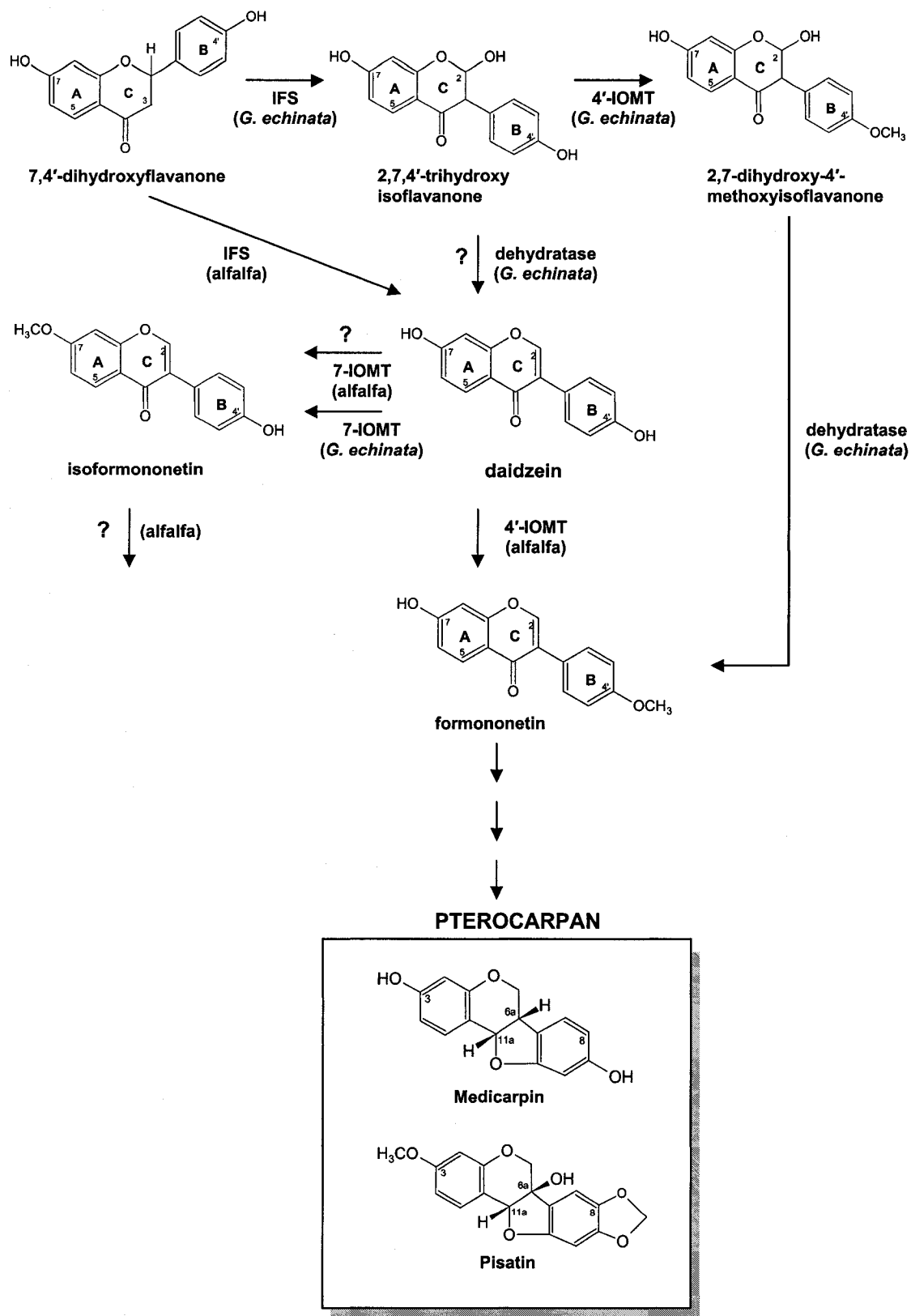
Figure 6: Biosynthesis of pterocarpan phytoalexins in alfalfa and *G. echinata*.
Dotted lines and question marks indicate possible pathways based on the present findings.

The IOMT (U97125) involved in the biosynthesis of medicarpin was first cloned from alfalfa cell cultures upon elicitation with $CuCl_2$ (He *et al.*, 1998). Surprisingly, this recombinant enzyme, as with the native purified protein (He and Dixon, 1996), was demonstrated to methylate daidzein (7,4'-dihydroxyisoflavone) *in vitro* at the 7 position of the A-ring leading to the formation of isoformononetin (4'-hydroxy-7-methoxyisoflavone). This rare compound does not accumulate in alfalfa and thus, is unlikely to be involved in the formation of medicarpin. Interestingly however, when challenged by abiotic elicitation or fungal infection, transgenic alfalfa overexpressing IOMT produces elevated levels of the 4'-*O*-methylated medicarpin and its precursor formononetin (7-hydroxy-4'-methoxyisoflavone) and, as a result, exhibits reduced susceptibility to the fungal infection (He and Dixon, 2000). Accordingly, the regiospecificity of IOMT appears to be different *in vivo* (4'-*O*-methylation of B-ring) as compared to *in vitro* (7-*O*-methylation of A-ring).

In order to decipher the mystery of the pathway, Dixon and co-workers (Dixon, 1999; He and Dixon, 2000; Liu and Dixon, 2001) proposed daidzein 7-IOMT as part of an *in vivo* multienzyme complex or "metabolic channel" that favors the 4'-*O*-methylation over 7-*O*-methylation, where the enzyme complex facilitates the direct transfer, or channeling, of intermediates between active sites producing formononetin and other compounds downstream of the pathway. This hypothesis has been supported by the observation that 7-IOMT fused with green fluorescent protein localized to the endoplasmic reticulum of alfalfa where other membrane-bound enzymes (cytochrome P450s) involved in the isoflavonoid pathway are also localized (Liu and Dixon, 2001). Recently, radiolabeling and isotope dilution studies demonstrated the incorporation of

2,7,4'-trihydroxyisoflavanone, but excluding daidzein, as an intermediate of formononetin biosynthesis in alfalfa (Liu and Dixon, 2001). This observation is in agreement with a previous study illustrating that radio-labeled daidzein is not incorporated into medicarpin in $CuCl_2$-treated seedlings of alfalfa, whereas formononetin is an acceptable precursor (Dewick and Martin, 1979). Nevertheless, 7-IOMT has been proposed to be the enzyme responsible for the 4'-*O*-methylation of 2,7,4'-trihydroxyisoflavanone (Liu and Dixon, 2001), which is supported by the crystal structure of 7-IOMT since it demonstrates that the enzyme can accommodate 2,7,4'-trihydroxyisoflavanone in its active site according to computer modeling studies (Zubieta *et al.*, 2001).

More recently however, a novel IOMT cDNA clone encoding the 4'-IOMT (AB091684), isolated from *Glycyrrhiza echinata*, was shown to methylate 2,7,4'-trihydroxyisoflavanone at the 4' hydroxyl group, but exhibited no 7-IOMT activity (Akashi *et al.*, 2003). Concurrently, a cDNA clone of 7-IOMT (AB091685) was also identified in *G. echinata* and the recombinant protein was shown to mediate the formation of isoformononetin from daidzein (Akashi *et al.*, 2003). This discovery clearly demonstrates, at least in *G. echinata*, that 7-IOMT and 4'-IOMT are two distinct enzymes catalyzing two biochemically different reactions (Figure 6). A cDNA encoding the 4'-IOMT (AB091686) was also detected in *Lotus japonicus* and the recombinant protein is functionally similar to *G. echinata* 4'-IOMT (Akashi *et al.*, 2003). Phylogenetic analyses revealed that both 7-IOMT from alfalfa and *G. echinata* are closely related, while the 4'-IOMT exhibits a higher degree of relatedness to the (+)-6a-hydroxymaackiain 3-OMT of pea (Akashi *et al.*, 2003), which is involved in pisatin biosynthesis.

Pisatin is a major pterocarpan produced by *Pisum sativum* upon microbial infection. The enzyme that catalyzes the terminal step in pisatin biosynthesis involves an OMT that methylates (+)-6a-hydroxymaackiain at the 3-position (3-position of pterocarpans is essentially position 7 of isoflavones) (Dixon, 1999 and references therein). Isolation of the cDNA clone of (+)-6a-hydroxymaackiain 3-OMT (U69554) from fungus-infected pea tissue gave a gene product with a Mr of ca. 40 kDa (Wu *et al.*, 1997) that exhibits the same substrate specificity as the native protein (Preisig *et al.*, 1989).

Although the majority of the phytoalexins characterized so far seem to be derived from isoflavonoids, other flavonoids such as flavanones and flavones may also act as phytoalexins. Examples of proteins involved in their biosynthesis include the flavone apigenin 7-OMT of Barley (X77467) (Christensen *et al.*, 1998), the flavanone naringenin 7-OMT of rice (Rakwal *et al.*, 2000), and the flavonol kaempferol 4'-OMT of carnation (Curir *et al.*, 2003), producing genkwanin, sakuranetin, and kaempferide, respectively. Only the cDNA clone of apigenin 7-OMT (Christensen *et al.*, 1998) from barley leaves has been isolated and its recombinant protein characterized. Nevertheless, all the native enzymes have been shown to be elicitor-induced and exhibit characteristics of Group II OMTs. Likewise, all IOMTs characterized thus far exhibit characteristics of the Group II OMTs.

E.4.5 OMTs Involved in Benzylisoquinoline Alkaloid Biosynthesis

Alkaloids constitute one of the largest groups of natural products represented by such compounds as morphine and codeine. Berberine, a benzylisoquinoline alkaloid, has been found to be associated with an antimicrobial defense mechanism and is also used to treat stomach ache. Berberine biosynthesis in plant cells has been well investigated at the enzyme level (Croteau *et al.*, 2000). The biosynthetic pathway leading from L-tyrosine to berberine involves thirteen different enzymatic reactions and several of the enzymes involved, including three OMTs, have been characterized and cloned from *Coptis japonica* (Figure 7). The three OMTs are the norcoclaurine 6-OMT (D29811) (Sato *et al.*, 1994), 3'-hydroxy-N-methylcoclaurine 4'-OMT (D29812) (Morishige *et al.*, 2000), and (*S*)-scoulerine 9-OMT (D29809) (Takeshita *et al.*, 1995).

Recently, the norcoclaurine 6-OMT (AY26894) and (*R,S*)-reticuline 7-OMT (AY26893) of opium poppy, *Papaver somniferum*, have also been characterized at the molecular and biochemical level and cDNA clones made available (Ounaroon *et al.*, 2003). The poppy norcoclaurine 6-OMT, as in the case of the 6-OMT, 4'-OMT, and 9-OMT of *Coptis*, exhibits particular substrate preference limited to benzylisoquinoline alkaloids. In contrast, the (*R,S*)-reticuline 7-OMT displays a broader substrate specificity, accepting simple catechols, as well as tetrahydrobenzylisoquinoline alkaloids. A similar situation is observed with the four multifunctional alkaloid OMTs (AF064693 to AF064696) from *Thalictrum tuberosum* (Frick and Kutchan, 1999), which methylate various compounds including catechol, hydroxycinnamic acids, and benzylisoquinoline alkaloids. Another alkaloid OMT cDNA that has been characterized is the columbamine OMT (AB073908) from cell cultures of *C. japonica*
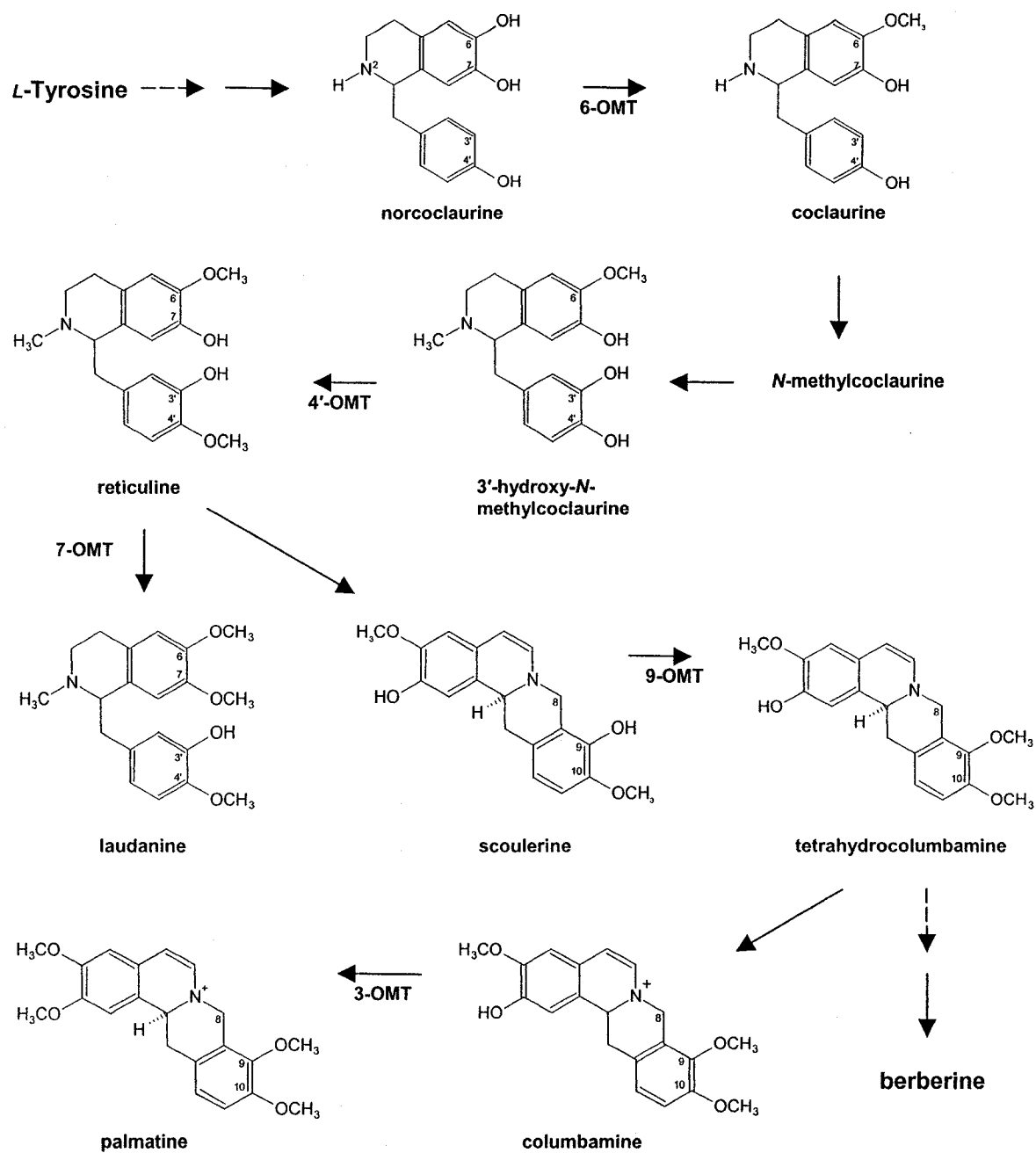
Figure 7: Biosynthetic pathway of berberine and various benzylisoquinoline alkaloids.

(Morishige *et al.*, 2002). This heterologously expressed protein catalyzes the conversion of columbamine to palmatine, thereby elucidating the remaining step in isoquinoline alkaloid biosynthesis in *C. japonica* (Figure 7). The OMTs involved in benzylisoquinoline alkaloids biosynthesis characterized to date all exhibit characteristics of Group II OMTs.

E.5 Utilization of OMTs in Metabolic Engineering

For much of human history, natural products have long been exploited as medicines, flavors, fragrances, dyes, and other uses. To date, they represent an enormous value from an economical standpoint. For the past decades, however, the economic limitation of chemical synthesis, and the pollution that is associated with it, has prompted the innovation of alternative modes for production of these organic compounds (Sato *et al.*, 2001). The advent of recombinant DNA technology and plant transformation techniques made it possible to apply genetic engineering to increase plant productivity and enhance the plants' resistance against microbes and pests. However, it is important to keep in mind that blocking or introducing new pathways, and increasing levels of secondary metabolites, may sometimes drastically alter the physiology of the plants, resulting in plants that are more susceptible to diseases, pest or accumulation of toxic materials (Hain and Grimmig, 2000). Therefore, it is crucial to have a comprehensive understanding of the regulation of a biosynthetic pathway; both at the gene and protein levels including those involved in the methylation of plant natural products.

E.5.1 Impact of COMT and CCoAOMT Gene Inhibition on Lignin Biosynthesis

To better comprehend the methylation pathway leading to the production of monolignols and to assess the feasibility of genetic engineering of COMTs for its effects on the content and composition of lignin, several attempts aimed at suppressing or overexpressing COMT genes have been made. Research involving down-regulation through gene silencing techniques, of COMT in transgenic tobacco, poplar, and alfalfa, indicates that COMT has a greater impact on S lignin than on G lignin (Atanassova *et al.*, 1995; Van Doorsselaere *et al.*, 1995; Guo *et al.*, 2001; Pincon *et al.*, 2001a). Recently, it has also been demonstrated that the loss of S units reflected a true block in S monomer biosynthesis rather than a structural change in the lignin resulting in lower recovery of S units. It is also intriguing that the suppression profile of COMT resembles that of the *bm3* mutant of maize (Barriere *et al.*, 1994). This mutant exhibits low OMT activity and it confers a lower lignin content, altered composition, and better digestibility, indicating that the COMT gene serves as a possible target for genetic manipulation.

Since the proposal of the involvement of CCoAOMTs in monolignol biosynthesis (see review by Ye *et al.*, 2001 and references therein), various attempts have been made to assess the impact of CCoAOMT gene down-regulation on the content and composition of lignin. Taken together, these results suggest that CCoAOMT functions in the biosynthesis of G lignin (at least in tobacco and alfalfa), but is not essential for the synthesis of S lignin (Zhong *et al.*, 1998; Guo *et al.*, 2001; Pincon *et al.*, 2001b). It is also worth noting that although down-regulation of COMT resulted in an almost complete loss of S lignins, suppression of CCoAOMT did not lead to a similar level of reduction in G lignins. This suggests that another methyltransferase is also involved in

the formation of G lignin in both alfalfa and tobacco (Ye *et al.*, 2001; Dixon *et al.*, 2001).

A potential candidate has been proposed to be COMT in view of the ~ 25% reduction in

G lignin that accompanies the larger reduction in S lignin in strongly COMT-suppressed

alfalfa (Dixon *et al.*, 2001). It is clear that more work is required to clarify the number of

different types of OMTs involved in the biosynthesis of G lignin.

The demonstration of the essential roles of COMT and CCoAOMT in

lignification has not only enriched our knowledge of the lignin biosynthetic pathway but

also implies the potential utility of these two OMTs as targets for genetic engineering of

lignin. It is anticipated that CCoAOMT alone, as well as a combination of both

CCoAOMT and COMT, may be ideal targets for modification of lignin content and

composition in trees and crops, in an effort to improve digestibility of forage crops, as

well as simplify the pulping process while reducing ecological damage. For a

comprehensive review of the subject, refer to Whetten *et al.* (1998), Dixon *et al.* (2001),

and Ye *et al.* (2001).

E.5.2   Impact of IOMT Overexpression on Disease Resistance

Recent advances have enabled the elucidation of the complete pathway leading to

isoflavonoid phytoalexin biosynthesis. This has since allowed for attempts to manipulate

the isoflavonoid pathway by genetically modifying the expression of the genes involved.

In particular, He *et al.* (2000) have demonstrated that over-expression of 7-IOMT in

abiotic- and biotic-elicited alfalfa leaves leads to earlier induction and higher

accumulation of formononetin glucoside and medicarpin, than in control lines.

Therefore, these elicitation studies indicate that genetic manipulation of IOMT, an

enzyme activity ascribed as an isoflavone 7-*O*-methyltransferase *in vitro*, leads to modulation of 4'-*O*-methylated isoflavonoids *in vivo* in transgenic alfalfa (He and Dixon, 2000).

In order to evaluate whether the more rapid accumulation of the phytoalexin medicarpin leads to resistance to fungus in plants over-expressing 7-IOMT, the fungus was introduced into wounded leaves of alfalfa in IOMT over-expressing lines. Measurement of the lesion size post-inoculation illustrated a significant reduction in lesion size in IOMT over-expressing lines compared to control lines, with an increase in medicarpin concentration of $\sim$ 370% in the IOMT over-expressing line in comparison to controls. These results clearly demonstrate the relationship between medicarpin content and resistance to *P. medicaginis* (He *et al.*, 2000). The overall result illustrated that over-expression of IOMT leads to markedly increased induction of most of the enzymes directly involved in isoflavonoid biosynthesis but does not appear to affect expression of these enzymes in the absence of the inducing agent.

The results described above clearly demonstrate that the increase in medicarpin leads to effective resistance of alfalfa to the pathogen *P. medicaginis*, providing direct evidence for the involvement of an isoflavonoid compound in disease resistance. The introduction of foreign phytoalexins into plants has been shown to be a feasible genetic manipulation technique by which the recipient plants acquire constitutive resistance. However, such studies do not address the function of the endogenous phytoalexins on the host plant. In the case of IOMT over-expression in alfalfa, the host's natural phytoalexin is overproduced, thereby allowing the evaluation of the potential role of that compound in disease resistance in the natural host (He *et al.*, 2000). It would also be important to

determine the impact of the down-regulation of the IOMT gene on isoflavonoid biosynthesis and on disease resistance. It could be anticipated that, while the over-expression of IOMT confers resistance, the down-regulation of IOMT might result in an increase in susceptibility to pathogen attack.

Furthermore, it is also interesting to study the defence machanism(s) undertaken by microorganisms to protect against phytoalexins. The metabolic strategies of phytoalexin detoxification or degradation are mechanisms by which successful phytopathogens may overcome the resistance response of their host (Dixon, 1999). Microbes could inactivate phytoalexins by way of utilizing enzyme systems in the modifications of these compounds. These modification processes include hydroxylation, demethylation, reduction/oxidation, and hydration, catalyzed by enzymes such as hydroxylases, demethylases, reductases, and hydratases, respectively (Dixon, 1999). Understanding the mechanisms by which the phytopathogen responds to phytoalexins is important for the application of metabolic engineering.

E.6   Phylogenomics

As important as it is to identify or assign putative functions to gene sequences with little or no experimentally-determined functions, the ability to accurately predict their substrate specificity is of equal, if not of, greater importance. Understanding gene functions undoubtedly requires experimental information, but improved prediction capabilities could facilitate the identification of potential substrates with a certain measure of accuracy (Schröder et al., 2002). Most of the methods utilized for gene function prediction rely merely on sequence homology/similarity, as has been done

traditionally. However, it is important to acknowledge that the identification of sequence similarity alone does not reliably reflect an identical function for an uncharacterized gene (Eisen, 1998). It has been shown that genes possessing sequence similarity do in fact exhibit diverse functions. This is exemplified by the fact that *Arabidopsis* OMT 1, which was reported to encode a putative COMT on the basis of its sequence homology/identity to other homologous genes (Zhang *et al.*, 1997) was later identified as a flavonol 3'-OMT gene after its expression in a heterologous system and careful study of its substrate specificity (Muzac *et al.*, 2000).

A fairly recent approach of comparative biology, taking advantage of an evolutionary perspective, has shed new light on the fields of genomics and gene function prediction. This approach, in addition to characterization and quantification of sequence similarity, seeks to elucidate how and why those similarities and differences came to be. Essentially, it questions the reliability of the use of similarity alone and asks how biologically relevant those similarities are, as "similarity itself is not a reliable indicator of evolutionary relatedness" (Eisen and Wu, 2002). Since the processes of convergence and divergence can both lead to sequence homology, in order to infer gene relatedness, it is important to differentiate between orthology and paralogy in functional prediction (Eisen, 1998). This distinction is important since paralogous genes have undergone a duplication event from the ancestral gene, while orthologous genes have not undergone duplication but are the result of speciation events. Thus, determining whether a putative gene is paralogous or orthologous to a gene with a known function would help in assigning a function to the putative gene (Eisen and Wu, 2002). In view of the fact that orthology can be described using phylogeny, and since the function of a gene influences

its rate and pattern of evolution, it is reasonable to hypothesize that the reconstruction of the evolutionary history of a gene (molecular phylogeny) would improve the accuracy of gene function prediction known as phylogenomics (Eisen, 1998; Eisen and Wu, 2002).

The reconstruction of an evolutionary history entails building conventional phylogenetic trees utilizing sequences of the gene of interest, either DNA or protein, such as a putative OMT clone identified by comparative studies and its respective homologs, preferably those whose functional characteristics are determined. The resulting phylogenetic tree is thus a reflection of the pattern of similarity translated into evolutionary relationships. Such trees can then be employed to predict the substrate specificity of the putative OMT based on the topology of the tree itself and the phylogenetic position of the putative OMT with respect to other biochemically-characterized OMTs. The application of this approach and other comparative studies have led Schröder et al. (2002) to successfully predict the function and substrate preference of the gene product of a putative OMT cDNA clone from Catharanthus roseus. The recombinant protein was later heterologously expressed and its activity was experimentally confirmed. A similar approach carried out by Eisen et al. (1997) and Heidelberg et al. (2000) have facilitated the discovery of new genes in their respective genomic projects involving the uvrA and photolyase gene families, respectively. More recently, a novel CCoAOMT from Arabidopsis was also identified based on sequence analysis and the development of a phylogenetic tree (Ibdah et al., 2003).

E.7 Study of the Evolution of OMTs

Plant systematists have long recognized and employed the approach of phylogenetics as a means to assist in the discovery, description, and interpretation of biological diversity. However, this approach has not been very forthcoming in other fields of biology, ironically in the fields of molecular biology and genomics. Therefore, it is interesting to not only discuss the approach of phylogenetics in the prediction of gene function and corresponding substrate prediction, but also to discuss the means by which the reconstruction of the evolutionary history of OMTs could be carried out.

According to one hypothesis on enzyme evolution (Crow and Dove, 1989), methyltransferases may have evolved in such a way that they have selectively "picked up" pre-existing polypeptide subunits and, over time, through amino acid substitution led to the acquisition of new functions. These methytransferases have however, due to functional constraints, retained some conserved motifs required for catalytic activity, and it is believed that lignin OMTs were the first to have evolved. This is due to the fact that the primitive form of lignin may have been present in the first vascular land plants (Lewis *et al.*, 1999; Qiu and Palmer, 1999). These early vascular land plants may have acquired the ability to synthesize this biopolymer for mechanical support, as well as transport of water and solutes. These early OMTs, through duplication, evolutionary divergence, and a process of "fine-tuning" to the preference and regiospecificity of their respective methyl acceptors, may have given rise to the present variety of OMTs that are involved in the methylation of various secondary metabolites (Ibrahim and Muzac, 2000).

The study of the evolution of metabolic enzymes has traditionally been based on the function of the enzyme of interest and the relative position of this enzyme in a

particular metabolic pathway. In general, it is reasonable to assume that enzymes upstream of a biosynthetic pathway would be more ancestral to those that are located downstream in the pathway. It is also interesting to note that upstream enzymes in a pathway are less subject to acquire mutations as compared to the enzymes that participate downstream in the pathway, as in the case of the flavonoid biosynthetic pathway (Rausher *et al.*, 1999). This is believed to be due to the fact that upstream enzymes participate in reactions that are located at branch-points of different metabolic pathways, resulting in metabolites that are subsequently acted upon by downstream enzymes. Changes in the functions of these upstream enzymes would, therefore, lead to a cascade of reactions affecting downstream enzymes.

As more OMT genes are identified and biochemically characterized, and their sequences made available in public databases, it would be interesting to study the evolution of these OMTs using molecular sequences (molecular phylogeny) as the character state instead of an enzyme's function alone, as has been previously done. In particular, the examination of the convergent versus divergent evolution of genes and gene functions of OMTs would no doubt tell us more about how this large gene family has resulted in the present-day diversity. Knowledge of the phylogenetic relationships of these OMTs would allow us to infer the rate of evolutionary change and the ages and diversification patterns of the various lineages.

To facilitate the study of the phylogenetic relationship between the taxa of a phylogenetic tree, it is useful to take advantage of a technique known as the reconciled tree. This involves the super juxtaposition of a gene tree and an organismal tree, an analytical tool that was originally utilized to account for the incongruence between

genetic and organismal phylogeny (Goodman *et al.*, 1979). Analysis of the reconciled tree would reveal information about the duplication events and an approximation as to where in evolutionary time they occurred, as well as allow processes such as gene losses or extinctions to be readily inferred. More significantly, this method simplifies the identification of orthologous from paralogous genes (Page, 2000). Therefore, the utilization of reconciled trees would provide new insights regarding the phylogenetic relationship between the taxa, which would not otherwise be possible if either a gene or organismal phylogeny alone were to be used. Nevertheless, in order to embed a gene tree into a species tree, there is a "cost" involved in the reconciliation process. This cost may result in the loss of information, such as the modification of gene duplication or deletion events, in the construction of the reconciled tree. Therefore, the "fitness" of a reconciled tree requires the incorporation of an optimality criterion that selects the best species tree for a particular gene tree; one that would accommodate the gene tree with the least cost (Page, 2000; Slowinski and Page, 1999). One could consult "The Angiosperm Phylogeny Group" (classification published in Annals of the Missouri Botanical Garden 85: 531-553 (1998) or http://www.mobot.org/MOBOT/Research/APweb/welcome.html for more information as to how to obtain a species tree.

Similarly, phylogenetic analyses can be employed as the basis for classification of the OMT gene superfamily, since their analysis as such has indicated that plant OMT genes can be grouped according to a functional trait that reflects the substrate preferences of their encoded proteins (Joshi and Chiang, 1998; Ibrahim *et al.*, 1998). Ideally, in order for such an endeavor to take form, more biochemically-characterized OMT sequences must first be made available, which again implies that novel OMT genes need to be

discovered. The classification of this OMT gene superfamily would not only facilitate in their taxonomic categorization, but more importantly will provide insights as to the selection of OMTs for metabolic engineering of commercially valuable secondary metabolites.

F.   MATERIALS AND METHODS

F.1   Chemicals and Buffers

S-adenosyl-L-[^{3}H]methionine (80Ci/mmol) was purchased from American Radiolabelled Chemicals, Inc. (St. Louis, MO). The enzyme substrates used were from our laboratory collection. [α-^{32}P]dCTP (3000Ci/mmol) and [γ-^{32}P]dATP (3000Ci/mmol) were purchased from Perkin Elmer Life and Analytical Sciences (Boston, MA).

The following buffers were used: **A,** sonication/lysis buffer, 50 mM NaH_2PO_4 (pH 7.8), 300 mM NaCl; **B,** assay buffer I, 50 mM NaH_2PO_4 (pH 7.8), 300 mM NaCl, 10% glycerol, 14 mM β-mercaptoethanol, 10 mM EDTA; **C,** assay buffer II, 50mM NaH_2PO_4 (pH 7.8), 300 mM NaCl, 10% glycerol, 14 mM β-mercaptoethanol; **D,** denaturation solution, 0.5 M NaOH, 1.5M NaCl; **E,** neutralization solution, 1 M Tris HCl, 1.5 M NaCl (pH 7.5); **F,** hybridization buffer (Church and Gilbert, 1984), 0.5M phosphate buffer (pH 7.2), 7%(w/v) SDS, 10mM EDTA; **G,** wash buffer I, 5x SSC (sodium chloride/sodium citrate), 0.1% SDS (sodium dodecyl sulfate); **H,** wash buffer II, 1x SSC, 0.1% SDS; **I,** TMAC (tetramethylammonium chloride) hybridization buffer, 3M TMAC, 0.1M NaPO$_4$ (pH 6.8), 1mM EDTA (pH 8.0), 0.6% SDS, 100 μg/ml denatured Salmon sperm DNA, 5x Denhardt's solution (100x solution: 5 g Ficoll 400, 5 g Polyvinyl pyrrolidone, 5 g BSA); **J,** TMAC wash buffer, 3M TMAC, 50 mM Tris.Cl (pH 8.0), 0.2% SDS; **K,** wash buffer III, 2x SSC, 0.1% SDS.

F.2   Molecular and Biochemical Characterization of *Arabidopsis* OMTs

F.2.1   DNA Sequencing

Polymerase Chain Reaction (PCR) products were purified using a Qiagen PCR

purification kit, and electrophoresed on 1% agarose gel with ethidium bromide (0.33

μg/mL) at 7 v/cm for 40 min. Amplified fragments were visually quantified employing a

GeneSnap 4.00-Gene Genius Bio Imaging System (Syngene) and the digital image files

were analyzed utilizing Gene Tools software from Syngene. The quantity of DNA was

estimated using a 1kbp DNA Standard Marker (MBI Fermentas). The purified, amplified

DNA was directly sequenced using ABI Big Dye Terminator v. 3.0 and 3.1 Cycle

Sequencing Ready Reaction kit and an ABI310 automated genetic analyzer (Applied

Biosystems, Foster City, CA). Each region was sequenced between two to seven times.

The thermal cycling profile of sequencing reactions were: Denaturation, 96°C, 10 sec;

Annealing, 50°C, 5 sec; Extension, 60°C, 4 min; repeated for 25 cycles. Alternatively,

DNA sequencing was done by the Hospital for the Sick Children (Toronto, ON) using

slab gel method and/or Sheldon Biotechnology Centre (McGill University) using the

dideoxy-mediated chain termination method (Sanger *et al.*, 1977). Primers employed for

sequencing reactions were either gene-specific or vector-specific primers such as

universal primers M13 Forward and Reverse, T7 and SP6.

F.2.2   *Arabidopsis thaliana* Full Length cDNA Clones

Two full-length putative OMT cDNA clones of *Arabidopsis* ecotype Columbia,

GenBank Accession # AF344315 and AF344316, were obtained from The *Arabidopsis*

Information Resource (TAIR). The two clones are in pUNI-T3-D/V5-His-TOPO vector

in *E. coli* strain PIR1 under the acquisition numbers C00113 and C00114, respectively.

F.2.3   Cloning Strategies

(a) Subcloning into prokaryotic expression system

The two putative OMT clones (C00113 and C00114) were sub-cloned into the pTrc-HisB

bacterial expression vector (Invitrogen).   Upon verification of the two cDNA clones by

sequencing reaction, the following specific primers were designed for sub-cloning:

AO315F: 5' -GAG CGG GTA CCA TGG GAT ACC TTT TCG- 3'

AO315R1: 5' -GGC GGA GCG AAT TCT CAC ACG TTT TAT TTA CAG AAC- 3'

AO316F: 5' -GAG CGG GTA CCA TGG GAT ACC TTT TTC- 3'

AO316R1: 5' -GGC GGA GCT TCG AAT CAC ACA TTT TAT TTA CAG AAT- 3'

These primers individually contain a restriction site at their 5' end; *Kpn*I site for AO315F

and AO316F, *Eco*RI site for AO315R1 and AO315R2, and *Bst*BI /*Bsp*1191for AO315R1.

These primers were used to carry out PCR amplification of the two clones utilizing *Pfu*

Turbo-high fidelity DNA polymerase (Stratagene) to maximize fidelity.

The optimized PCR conditions for clone C00113 utilizing primers AO315F and

AO315R1, were carried out for the first round, following a first denaturation step at 94°C

for 5 min, as follows: Denaturation, 94°C, 30 sec, Annealing, 63°C, 30 sec; Extension,

72°C, 1.5 min. This was repeated for a total of 30 cycles, with a final extension time of 5

min.   For clone C00114, primers AO316F and AO316R1 were used and the thermal

cycling optimized conditions were the same as with C00113, expect that the annealing step was carried out at $65^{\circ}C$ for 30 sec.

The ~1.2-kb *KpnI–EcoRI* fragment of C00113 and ~1.2-kb *KpnI–BstBI* fragment of C00114, containing the full-length cDNA sequences of both putative OMTs, were digested accordingly and ligated into the corresponding recognition sites of the digested pTrc-HisB expression vector. The DNA inserts were positioned downstream and in-frame with a sequence that encodes a $(His)_6$ tag. Both the pTrc-HisB-C00113 and the pTrc-HisB-C00114 constructs were transformed into *E. coli* $CaCl_2$-competent Top 10 cells (Invitrogen) by heat shock (Sambrook *et al.*, 1989), and selected on ampicillin (100 µg/mL) containing LB-media (Luria-Bertani). The presence of the DNA inserts in their respective vectors was confirmed by bacterial-colony PCR, and further verification was performed by DNA sequencing.

(b) Subcloning into eukaryotic expression system

Heterologous protein expression in the *Pichia pastoris* system (EasySelect pPicZ-His; Invitrogen) can be either intracellular or secreted. Secretion of the recombinant protein requires the presence of an in-frame signal peptide. In this instance, the $(His)_6$ tag is fused to the C-terminus end of the recombinant protein. The DNA fragment was amplified by PCR using gene-specific primers, PicZ315F and PicZ315R, or PicZα315F and PicZα315R for intracellular and secreted systems, respectively. For the pPicZ construct, the initiation ATG was part of the yeast consensus sequence (AATA<u>ATG</u>TCT). For the pPicZα construct, the insert was cloned in-frame with the N-terminal signal sequence and C-terminal tag. The pPicZ-315 and pPicZα-315 constructs

were transformed into *Pichia* cells by electroporation according to manufacturer's instructions. Putative multi-copy recombinant clones were selected by plating cells on YPD-media containing increasing concentrations of the antibiotic Zeocin (500, 1000, 2000 μg/mL). *Pichia* transformants were analyzed according to their Mut (methanol utilization) phenotype on minimal media following the protocol instructions. Transformation of the yeast strain (X-33) with linearized constructs favors single crossover recombination at the *AOX1* locus; therefore, most of the transformants should be Mut$^+$. However with the presence of the *AOX1* sequences in the plasmid, there is a chance that recombination will also occur in the 3' *AOX1* region, disrupting the wild-type *AOX1* gene and creating Mut^S transformants. Hence, loss of the *AOX1* gene alters the ability of the cell to metabolize methanol. Testing cell growth on minimal media containing dextrose or methanol will allow the confirmation of the Mut$^+$ phenotype, where there is little growth with dextrose but normal growth on methanol-containing media. Yeast colony-PCR was also used to verify the presence of the insert utilizing *AOX1* specific primers; 5' AOX1 site: 5'–GAC TGG TCC CAA TTG ACA AGC–3' and 3' AOX site: 5'–GCA AAT GGC ATT CTG ACA TCC–3'. Further verification of the insert was done by sequencing reaction.

Gene specific primers designed for the sub-cloning of C00113 are as follows:

PicZ315F: 5'–GCG GAA TTC GCA ATA ATG TCT ATG GGA TAC CTT TTC GA–3'

PicZ315R: 5'–GGA GGG TAC CTT TAC AGA ACT CAA TAA TCC AG–3'

PicZα315F: 5'–GCG GAA TTC ATG GGA TAC CTT TTC GA–3'

PicZα315R: 5'–GAG GCG GCC GCT TTA CAG AAC TCA ATA ATC CAG–3'

These primers individually contain a restriction site at their 5' end, *EcoR*I site for

PicZ315F and PicZα315F, *Kpn*I and *Not*I site for PicZ315R and PicZα315R,

respectively. *Pfu* Turbo-high fidelity DNA polymerase (Stratagene) was used to carry

out the PCR amplification of the C00113 clone. PicZ315F and PicZ315R or PicZα315F

and PicZα315R were used to prime the 5' direction and 3' direction, respectively. The

optimal PCR condition carried out for the first round, following a first denaturation step

at 94°C for 5 min, was as follows: Denaturation, 94°C, 5 min; Annealing, 62.7°C, 30 sec;

Extension, 72°C, 1.5 min; repeated for a total of 30 cycles, with a final extension time of

5 min. The thermal cycling conditions optimized for PicZα315F and PicZα315R were

the same, except that the annealing step was at 65.1°C for 30 sec. The PCR products were

transferred into multiple cloning sites of pPicZ and pPicZα plasmids, and transformed

into *E. coli* XL1-Blue strains by heat-shock in order to isolate sufficient amount of

plasmid for transformation into yeast.

F.2.4   Heterologous Expression of Recombinant Proteins

(a) Recombinant expression in E. coli

In order to characterize the expressed protein, a one-mL aliquot of an overnight culture of

E. coli Top 10 cells carrying the expression plasmid was used to inoculate 50 mL of fresh

LB medium containing 100 µg/mL ampicillin. The culture was incubated at 37°C until

an OD_{600} of ~ 0.7 to 0.9 was attained, followed by the addition of isopropyl-β-D-

thiogalactopyranoside (IPTG) to a final concentration of 2 mM, after which the culture

was allowed to grow for an additional 4 h. The bacterial extracts were prepared

according to the manufacturer's protocol for purification of recombinant (His)$_6$-tagged proteins on Ni-nitrilo triacetic resin (Ni-NTA, Qiagen). Briefly, the bacterial cells were pelleted and stored at -80°C overnight, they were then resuspended in buffer A, pelleted again, then treated with lysozyme to a final concentration of 1 mg/ml for 30 min on ice. The cells were then subjected to sonication on ice, and the lysate cleared by centrifugation at 10,000 g for 20 min at 4°C. The soluble fraction was either used as the crude enzyme preparation for initial characterization of substrate preference, or purified on Ni-NTA resin for subsequent SDS-PAGE analysis of the recombinant protein and further enzyme assays. Alternatively, induction of protein expression (2 mM IPTG final concentration) was carried out at either room temperature or 30°C for 16 hours, in order to ensure that the protein did not form inclusion bodies or accumulate in the insoluble fraction of the cell lysate.

(b) Recombinant expression in P. pastoris

Recombinant protein expression in *P. pastoris* is dependant on the *AOX1* gene promoter, with methanol as the inducer. In the absence of glucose, which acts as a repressor, recombinant protein production was induced by the addition of 0.5% methanol every 24 hours over a period of 3 to 4 days, for both intracellular and secretory systems. A buffered culture medium, containing 100 mM potassium phosphate, pH 8.0, was used for cell growth and protein induction in order to limit enzyme inactivation and prevent protein degradation. In the case of intracellular recombinant protein production, cells were lysed using glass beads, and their extracts were prepared according to manufacturer's instructions. Secreted proteins were concentrated by ammonium sulfate

precipitation, and the protein that precipitated between 35 to 70% salt saturation was pelleted and used for activity determination and subsequent purification. Cleared extracts were also subjected to purification on Ni-NTA resin.

F.2.5   Partial or Affinity Purification of Fusion Proteins

Crude protein lysate, if not used directly for enzyme assays, was subjected to partial purification with ammonium sulfate precipitation in order to concentrate the protein for subsequent enzyme assays. Upon precipitation, the appropriate cut-off of the precipitated fraction was transferred to PD-10 desalting columns (Sephadex G-25; Amersham) for the removal of salts for enzyme assays. The desalted-protein fraction can be further purified by affinity chromatography utilizing Ni-NTA resin according to manufacturer's specifications. Briefly, the lysate was incubated with the resin that was previously equilibrated with buffer A for 2 to 16 h at 4°C and the mixture was loaded onto a column. After flow-through, the column was thoroughly washed with buffer A, followed by a second wash with the same buffer containing 25 mM imidazole. The recombinant protein was eluted with 250 mM imidazole in buffer A and, at this point, can be either used for SDS-PAGE, or concentrated on a Centricon (YM 10 membrane; Amicon), and buffer exchanged against buffer B for immediate use in enzyme assays. Alternatively, the affinity-purified protein fraction was immediately subjected to a buffer exchange against buffer B on a PD-10 column for use in enzyme assays.

F.2.6 *O*-Methyltransferase Assays

Enzyme activity was assayed with buffer B using 50 to 500 μM substrates (dissolved in DMSO), 25 nCi [^{3}H] AdoMet (55mCi/mmol) as the co-substrate, and 10 to 100 μg protein in a total volume of 100 μL. MgCl$_2$ at a final concentration of 1.0 mM was added to certain enzyme assays to account for any Group I OMT activity that might require divalent cations. The reaction was started by the addition of protein, incubated for 30 min at 30°C, and terminated by the addition of 10 μL of 6N HCL. The reaction products were extracted with 500 μL of a benzene-ethyl acetate mixture (1:1, v/v) or 500 μL of a diethyl ether-ethyl acetate mixture (1:1, v/v) or 100 μL of hexane depending on the nature of substrates used. An aliquot of the organic phase, containing the labeled reaction product, was mixed with a toluene-based scintillation fluid and counted for radioactivity in a Wallac LKB 1217 Rackbeta or a Winspectra 1414 Liquid Scintillation counter.

F.2.7 Polyacrylamide Gel Electrophoresis and Western Blot Analysis

SDS-PAGE was performed on 12.5% acrylamide gels according to (Laemmli, 1970) using molecular weight markers (Bio-Rad) ranging from 14.4 to 97.4 kDa for calibration. After electrophoresis, proteins were stained with Coomassie Brilliant Blue (R-250). For Western analysis, the proteins were submitted to a semi-dry electrophoretic transfer onto nitrocellulose membranes. Upon blocking the membrane with non-fat dry 3% milk, the blots were probed with IgG-purified *Arabidopsis* anti-3'-OMT (1:500 dilution) as primary antibody (Muzac *et al.*, 2000) and alkaline phosphatase-conjugated goat-anti-rabbit IgG's (1:5000; Bio-Rad) as secondary antibody. Alkaline phosphatase

activity was revealed in the dark using nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad) as substrates. Alternatively, blots were probed with anti-His (1:5000; Amersham) and HRP (Horseradish peroxidase)-linked goat-anti-mouse IgG's (1:3000; Amersham), or IRDye800-conjugated goat-anti-mouse IgGs (1:5000; Rockland Immunochemicals). Peroxidase activity was visualized in a chemiluminescent reaction detected by exposure to light-sensitive autoradiography film. In the case of the IRDye800 system, detection was carried out using IRDye technology and the Odyssey workstation provided by Li-Cor Inc. All detection procedures were performed according to manufacturer's instructions.

F.2.8 Protein Determination

Protein concentration was determined employing the Bio-Rad protein assay based on the method of (Bradford, 1976) according to manufacturer's protocols, with bovine serum albumin as the standard protein.

F.3 cDNA Library Screening of *Serratula tinctoria*

F.3.1 Library Construction

A bacterial cDNA library derived from the roots of *Serratula tinctoria* was prepared by Dr. Jean Danyluk (Université du Québec à Montréal; UQAM). Briefly, two month-old roots of *Serratula* were extracted for the isolation of mRNA using RNAwiz according to manufacturer's instructions (Ambion) and mRNA was isolated from total RNA using oligo-dt column (Ammersham). For cDNA synthesis, the superscript plasmid system with Gateway technology and cloning kit (Invitrogen) was used, except that the

precipitation steps without yeast carrier tRNA were replaced by the Qiaquick PCR purification procedure (Qiagen). cDNAs were directionally cloned into the pCMT-Sport6 vector with the *Sal* I adaptor on the 5' end and the *Not* I adaptor on the 3' end. Six million primary transformants were amplified and frozen as glycerol stocks.

F.3.3 Primers

Internal peptide information was obtained from partially-purified quercetin 3-OMT of *S. tinctoria* by LC-MS/MS analyses (Huang *et al.* 2004). Oligonucleotide primers for PCR-based or hybridization screening were designed based on the amino acid sequences of five out of the six tryptic peptides obtained from the partially-purified native protein. Degenerate primers were used in order to account for codon degeneracy and, in each case, primers were designed utilizing minimal degeneracy in the genetic code. Listed below are the primers used for this experiment:

DegPep1F: 5'–AC(C/T)AC(C/T)ATGATGCA(C/T)AG(A/G)CTTAA(A/G)–3'

DegPep4F: 5'–GG(A/T)AGTAG(C/T)GACCATAG(C/T)AC(C/T)ATGAG(C/T)ATGAA(A/G)AAA–3'

DegPep5F: 5'–GG(A/T)GTTATTAT(C/T)CT(C/T)GCTGCACT(G/T)CCTAA(A/G)–3'

PEP1F: 5'–AC(A/C/G/T)AC(A/C/G/T)ATGATGCA(C/T)(C/A)G–3'

PEP3.1F: 5'–ATGGA(A/G)TC(A/C/G/T)TGGTA(C/T)CA(C/T)–3'

PEP3.2F: 5'–ATGGA(A/G)AG(C/T)TGGTA(C/T)CA(C/T)–3'

PEP4A.1F: 5'–GG(A/C/G/T)ATGTC(A/C/G/T)GA(C/T)CA(C/T)–3'

PEP4A.2F: 5'–GG(A/C/G/T)ATGAG(C/T)GA(C/T)CA(C/T)–3'

PEP4B.1F: 5'–AC(A/C/G/T)ATGTC(A/C/G/T)ATGAA(A/G)AA(A/G)–3'

PEP4B.2F: 5'–AC(A/C/G/T)ATGAG(C/T)ATGAA(A/G)AA(A/G)–3'

PEP5.1F: 5'–GA(C/T)AT(T/C/A)AT(T/C/A)TT(A/G)GC(A/C/G/T)GC–3'

PEP5.1R: 5'–GC(A/C/G/T)GC(C/T)AA(A/G/T)AT(A/G/T)AT(A/G)TC–3'

PEP5.2F: 5'–GA(C/T)AT(T/C/A)AT(T/C/A)CT(A/C/G/T)GC(A/C/G/T)GC–3'

PEP5.2R: 5'–GC(A/C/G/T)GC(A/C/G/T)AG(A/G/T)AT(A/G/T)AT(A/G)TC–3'

PEP6.1F: 5'–GC(A/C/G/T)TT(A/G)AT(T/C/A)CA(C/T)AA(A/G)–3'

PEP6.1R: 5'–(C/T)TT(A/G)TG(A/G/T)AT(C/T)AA(A/C/G/T)GC–3'

PEP6.2F: 5'–GC(A/C/G/T)CT(A/C/G/T)AT(T/C/A)CA(C/T)AA(A/G)–3'

PEP6.2R: 5'–(C/T)TT(A/G)TG(A/G/T)AT(A/C/G/T)AG(A/C/G/T)GC–3'

Sport6F1: 5'–ACACCACAGAAGTAAGG–3'

Sport6F2: 5'–AAGCTAGCAGTTTTCC–3'

F.3.4   Screening of the cDNA Library

(a) PCR-based screening

An aliquot of plasmid DNA (ca.10 ng) from the cDNA library was used as a template in

PCR reactions with different combinations of primers. Due to the complex combinations

of primers used (Figure 15), not all combinations are listed. Examples of primer

combinations include gene specific forward primer (PEP1F) and vector specific reverse

primer (Sport6F2); or gene specific primers alone (forward primer PEP1F and reverse

primer PEP6.1R). PCR amplification was carried out using "hot-start" Platinum *Taq*

DNA Polymerase (Invitrogen) which reduces amplification at ambient temperature, thereby increasing the sensitivity, specificity, and yield of the PCR reaction. The absence of endonuclease and exonuclease activity of this polymerase also helps to prevent proof-reading activity, which is not desired when using degenerate primers. Thermal cycling programs were as follows: Activation, 94°C for 2 min (for activation of the polymerase); Denaturation, 95°C for 20 sec; Annealing, 55°C for 30 sec; Extension, 72°C for 1 min; repeated for 10 ~ 12 cycles (depending on the situation) with annealing temperatures that progressively decreased by 1°C each cycle. This touch-down step is followed by another 40 cycles of amplification with a 20-sec denaturation at 95°C, 30-sec annealing at 38 ~ 48°C (depending on the *Tm* of the reaction primers), 1 min extension at 72°C, and a final extension of 5 min at 72°C. Amplified DNA fragments were diluted and employed as templates for a second round of PCR amplifications with another series of different primer sets, or nested primers, when possible.

Following electrophoresis, PCR fragments from the second round of amplification were verified by Southern analyses, with [α-^{32}P]dCTP-labelled *Arabidopsis* 3'-OMT full-length cDNA as a heterologous probe for hybridization. Selected PCR products, with the expected fragment sizes, were then subjected to blunt end treatment with *Pfx*-DNA Polymerase prior to cloning into the pCR4Blunt-TOPO vector using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). Ligated constructs were then transformed into *E. coli* strain Top10 by heat shock. Plasmids were then extracted for verification of the presence of inserts, either by restriction analysis or colony-PCR. Putatively positive clones were subsequently sequenced for further characterization. Alternatively, PCR

fragments were ligated to the pGEM-T vector using the pGEM-T cloning kit (Promega) and transformed into *E. coli* strain XL1-Blue by heat shock.

(b) Hybridization screening/Colony lift screening

The bacterial cDNA library was screened using i) *Arabidopsis* 3'-OMT full-length cDNA probe according to (Sambrook *et al.*, 1989) or ii) degenerate oligonucleotide probes (DegPep1F, DegPep4F, and DegPep5F) according to (Duby *et al.*, 2000). Approximately 50,000 cells were plated in duplicates of twelve LB-plates and subsequently transferred to 0.45 micron 137 mm nylon membranes (Osmonics Inc.) for hybridization to radio-labelled probes. Following transfer, cells were subjected to lysis with 10% SDS for 3 min, denaturation with buffer D for 7 min, neutralization with buffer E for 7 min, soaked with 2X SSC for 7 min and then dried for 30 min, before DNA was fixed to membranes by UV crosslinking using a Stratalinker at 120,000 μJ/cm^2. Screening with single oligonucleotide probes, as in the case of the *Arabidopsis* 3'-OMT full-length cDNA, involves random priming of the DNA fragment with [α-^{32}P]dCTP using Klenow as part of the Multiprime DNA labelling system (Amersham). Wet membranes were first pre-hybridized for 3 hours with buffer F at 65°C, then hybridized overnight with the radiolabeled probe at the same temperature. Subsequently, membranes were washed with buffer G for 20 min at 42°C and an additional 15 min for the temperature to increase from 42°C to 65°, follow by two more rinses with buffer H at 65°C for 20 min prior to detection. Positive clones were subsequently isolated, replated at a lower density onto smaller dishes and the process was repeated. At this point, putative positives were

isolated and analyzed by restriction digest and subjected to Southern analysis prior to sequencing for final identification.

Library screening utilizing degenerate primers as probes involves labeling of oligonucleotides with $[\gamma\text{-}^{32}P]dATP$ using T4 Polynucleotide kinase (Invitrogen). Since these probes consist of multiple oligonucleotides that are shorter in length, the conditions for optimal hybridization differ from those with longer, homogeneous probes. In TMAC (tetramethylammonium chloride), the melting temperature of an oligonucleotide is a function of length independent of base composition. Therefore, erroneous hybridization due to high G-C content of some of the oligonucleotides can be reduced (Duby *et al.*, 2000). The membranes with fixed DNA were pre-hybridized with buffer I at ca. 57 to $62^\circ C$ (a temperature 7 to $10^\circ C$ lower than the Tm of the reaction primers) for 2 to 3 h. Upon addition of $[\gamma\text{-}^{32}P]dATP\text{-}DegPep1F$ probes, membranes were hybridized at 57 to $62^\circ C$ for 40 to 60 h. After hybridization, membranes were washed with buffer J, two washes at room temperature (the first wash for 5 min and the latter for 15 min), followed by a third wash at $60^\circ C$ for 1 h. Final rinses, 3 times, 10 min each, were carried out at room temperature with buffer K prior to detection. Radio-labeled probes were then stripped off the membranes for subsequent hybridization with $[\gamma\text{-}^{32}P]dATP\text{-}DegPep4F$ and $[\gamma\text{-}^{32}P]dATP\text{-}DegPep5F$. Positive clones were purified, as described previously, by one more round of screening and followed by Southern analysis before sequencing for further verification.

F.3.5 Southern Blot Analysis

Standard Southern analysis was carried out according to Sambrook *et al.* (1989). PCR products or digested plasmid DNA were subjected to electrophoresis on 0.8% 1X TAE agarose gel with ethidium bromide (0.33 μg/mL) at 7 v/cm for 40 min. After successive treatments with 0.25M HCl (acidification), twice 10 min each; buffer D (denaturation), twice 15 min each; and buffer E (neutralization), twice 20 min each, the DNA was transferred by capillary action overnight to a charged-nylon membrane (Hybond-N$^+$, Amersham) using 20X SSC solution. Following transfer, the membrane was either baked at 80°C for 2 h or subjected to UV-crosslinking. The membrane was then ready for hybridization with radiolabeled probes as previously described (random or terminal labelling), and membranes were exposed to Kodak X-Ray film (BioMax Ms Film) or Phosphor-Imager technology (Bio-Rad).

F.4   Phylogenomic Analysis of OMTs

Three major methods, namely maximum parsimony, distance and maximum likelihood, are frequently employed for the reconstruction of phylogenetic trees. These methods each differ from one another in their assumptions and algorithms of character state optimization (Page and Holmes, 1998). The maximum likelihood method requires extensive computing power for analysis of large number of amino acid sequences, and therefore, only the former two methods were carried out.

F.4.1   Multiple Sequence Alignment

Genes that are selected for the study include only OMTs that have been biochemically characterized. Sixty such OMTs from diverse plant species were selected by surveying publications on search engines such as PubMed (http://www.ncbi.nlm.nih.gov), and by browsing the sequence databases such as EMBL, GenBank, and SwissProt. The amino acid sequences of these OMTs were aligned using multiple sequence alignment programs, CLUSTAL-W (Thompson *et al.*, 1994) at http://www.ebi.ac.uk/clustalw, using the PHYLIP output format. Upon completion of sequence alignment, the data were saved as TEXT-FILE ONLY format prior to transfer into a data exploration program, MacClade 4.03 (Maddison and Maddison, 2001), and into a phylogenetic tree-building and analysis program, PAUP 4.0, beta 8 (Swofford, 2001). All the alignments were visually inspected and manually edited, as needed, in MacClade 4.03 before tree reconstruction.

F.4.2   Phylogenetic Analysis and Statistics

A) Parsimony Analysis

Maximum parsimonious (MP) phylogenetic trees were constructed using a heuristic search with equal character weights, gaps treated as missing, multi-state taxa interpreted as uncertainty, starting tree was obtained via stepwise addition, and random addition of sequences with 1000 replicates for the data set. Tree bisection-reconnection (TBR) was used as the branch swapping algorithm. Strict and 50% majority rule consensus trees were obtained. The phylogenetic trees were rooted using non-plant OMT sequences as

59

out-groups. Bootstrap analysis with a fast heuristic search based on 1000 replicates was performed to assess the robustness of the tree branches.

The character consistency index (CI) is equal to m_i / s_i (Maddison and Maddison, 2001 and references therein) where m_i is the minimum conceivable number of steps for character i on any tree (always equal to 1 for the binary data), and s_i is the number of reconstructed steps for character i on this tree. The consistency index for all characters on a tree is the minimum possible tree length divided by the observed tree length, with decreasing values indicating increasing homoplasy. The retention index (RI) for all characters in a tree is calculated as the (maximum possible tree length minus actual tree length) divided by (maximum possible tree length minus minimum possible tree length). Since invariant characters are not included in the RI, if the characters in the data matrix are parsimony-informative and are completely congruent with each other as well as with the tree, then the RI will have a value of 1. However, if the data are maximally homoplastic on the tree, the RI will have a value of 0 (Maddison and Maddison, 2001 and references therein). The rescaled consistency index (RC) for all characters on a tree is the CI multiplied by the RI. Invariant characters are not included in the RC. As with the RI, the RC ranges from 0 to 1, with higher RC values indicating that characters in the data set are more congruent with each other and with the tree (Maddison and Maddison, 2001 and references therein). The CI, RI and RC's values were calculated with one of the options implemented in PAUP 4.0, beta 8.

b) Distance Analysis

The optimality criterion employed for the distance method was Neighbor-Joining (Saitou and Nei, 1987). The aligned amino acid sequences were analyzed using default options of the program and all characters were weighted equally, gaps treated as missing characters, multi-state taxa interpreted as uncertain, and the distance measure was set to be equal to mean character difference. The robustness of the tree branches was tested utilizing bootstrap analysis with neighbor-joining search algorithm of 1000 replicates.

G. RESULTS

G.1 Molecular and Biochemical Characterization of *Arabidopsis* OMT Clones

G.1.1 *Arabidopsis thaliana* cDNA Full Length Clones

The two putative OMT cDNA sequences (GenBank Accession # AF344315 and AF344316) employed in this study were first identified by BLAST search. This involved querying the TAIR database (The *Arabidopsis* Information Resources; http://www.arabidopsis.org) with the amino acid sequence of the flavonol 3'/5'-*O*-methyltransferase (Gauthier *et al.*, 1996) to retrieve putative OMT sequences from the *Arabidopsis* database. The BLAST results showed numerous ESTs (Expressed Sequence Tags) annotated as potential OMTs based on sequence homology. AF344315 and AF344316 were subsequently chosen based on the fact that they exhibit the highest score; a calculation of the sequence similarity of these two putative OMTs relative to the functionally characterized flavonol 3'/5'-*O*-methyltransferase of *C. americanum* (GenBank Accession # U16794).

Both AF344315 and AF344316 sequences contain an open reading frame of 1131 bp that would encode a 373 amino acid protein with a deduced molecular mass of ~ 41 kDa. The two sequences share a 92% similarity, at both the nucleotide and amino acid levels, and 91% identity at the amino acid level. The putative molecular weights obtained are representative of Class II OMTs that do not require Mg^{2+} for enzyme activity and, as shown in Figure 8, both AF344315 and AF344316 possess the five characteristic motifs (Ibrahim *et al.* 1998) that have been associated with plant OMTs, albeit with slight modifications. AF344316 differs from AF344315 by Glu-6 to

```
CLUSTAL W (1.82) multiple sequence alignment


AF344315    MGYLFEETLSSNPKTPIVVDDDNELGLMAVRLANAAAFPMVLKASLELGVFDTLYAEASR 60
AF344316    MGYLFQETLSSNPKTPIVVDDDNELGLMAVRLANAAAFPMVLKAALELGVFDTLYAAASR 60
            *****:*****************************************:*********** ***


AF344315    TDSFLSPSEIASKLPTTPRNPGAPVLLDRMLRLLASYSMVKCEKVSVGKGERVYRAEPIC 120
AF344316    TDSFLSPYEIASKLPTTPRNPEAPVLLDRMLRLLASYSMVKCGKALSGKGERVYRAEPIC 120
            ******* ************* ******************* *. ************


AF344315    RFFLKNNIQDIGSLASQVIVNFDSVFLNTWAQLKDVVLEGGDAFGRAHGGMKLFDYMGTD 180
AF344316    RFFLKDNIQDIGSLASQVIVNFDSVFLNTWAQLKDVVLEGGDAFGRAHGGMKLFDYMGTD 180
            *****:******************************************************


                                           Motif I              Motif II
AF344315    ERFSKLFNQTGFTIAVVKKALEVYQGFKGVNVLVDVGGGVGNTLGVVTSKYPNIKGINFD 240
AF344316    ERFSKLFNQTGFTIAVVKKALEVYEGFKGVKVLVDVGGGVGNTLGVVTSKYPNIKGINFD 240
            ***********************:*****:*:****************************


                    Motif III                                    Motif IV
AF344315    LTCALAQAPSYPGVEHVAGDMFVDVPTGDAMILKRILHDWTDEDCVKILKNCWKSLPENG 300
AF344316    LTCALAQAPSYPGVEHVAGDMFVDVPTGDAMILKRILHDWTDEDCVKILKNCWKSLPENG 300
            ************************************************************


                              Motif V
AF344315    KVVVIELVTPDEAENGDINANIAFDMDMLMFTQCSGGKERSRAEFEALAAASCFTHCKFV 360
AF344316    KVVVIELVTPDEAENGDINANIAFDMDMLMFTQCSGGKERSRAEFEALAAASGFTHCKFV 360
            ***************************************************** ******


AF344315    CQAYHCWIIEFCK 373
AF344316    CQAYHCWIIEFCK 373
            *************
```

Figure 8: The amino acid sequences of *Arabidopsis* AF344315 and AF344316. The five signature motifs of OMTs are underlined in bold. * Represents strictly conserved regions.

Gln, Ser-45 to Ala, Glu-57 to Ala, Ser-68 to Tyr, Gly-82 to Glu, Glu-103 to Gly, Val-105 to Ala, Ser-106 to Leu, Val-107 to Ser, Asn-126 to Asp, Gln 205 to Glu, Asn-211 to Lys, Cys-353 to Gly.

BLAST searches of GenBank, EMBL and Swiss-Prot show that AF344315 and AF344316 are similar to other OMTs exhibiting amino acids sequence similarity and identity greater than 60% and 40%, respectively (Table 1), to other biochemically characterized OMTs (Figure 9). These OMTs include the flavonol 3'-OMT of *Arabidopsis* (U70424; Muzac *et al.*, 2000), an OMT utilizing hydroxycinnamic acids and flavonoids of Periwinkle (AY028439; Schröder *et al.*, 2002), the multifunctional OMTs from *Thalictrum tuberosum* (AF064693 to AF064696; Frick and Kutchan, 1999) and, in particular, the phloroglucinol OMT of *Rosa chinensis* (AB121046; Wu *et al.*, 2004). The BLAST analysis suggests that the gene products of AF344315 and AF344316 could potentially utilize a variety of substrates, ranging from small dihydroxyphenols to flavonoids, where biochemical characterization may ultimately confirm their substrate preferences.

G.1.2   Expression of Recombinant C00113 and C00114 in *E. coli*

To further characterize the gene products of clone C00113 (Accession #AF344315) and clone C00114 (AF344316), they were subcloned into the *E. coli* expression vector pTrc-His. These constructs allow for the expression of fusion proteins with N-terminally located $(His)_6$-tags under the control of the *trc* (*trp-lac*) promoter. Upon induction with 2 mM IPTG for 4 h, the solubility of the fusion proteins was assessed and enzyme assays were carried out. The level of fusion protein expression was

Table 1: The overall relationship (% similarity and identity) of the two putative *Arabidopsis* cDNA clones (AF344315 and AF344316) with other biochemically characterized OMTs from various plant species.

	AB121046[a]		U70424[b]		AY028439[c]		AF064695[d]	
	Similarity	Identity	Similarity	Identity	Similarity	Identity	Similarity	Identity
AF344315	69%	50%	67%	48%	67%	47%	69%	46%
AF344316	63%	46%	63%	45%	63%	44%	63%	43%

[a] Phloroglucinol OMT of *Rosa chinensis*
[b] Flavonol 3'-OMT of *Arabidopsis*
[c] An OMT utilizing hydroxycinnamic acids and flavonoids of Periwinkle
[d] A multifunctional OMT from *Thalictrum tuberosum* utilizing substrates from both phenylpropanoids and alkaloids
(Refer also to **Figure 9**)

```
CLUSTAL W (1.82) multiple sequence alignment

U70424      MGSTAETQLT------PVQVTDDEAALFAMQLASASVLPMALKSALELDLLEIMAKNG-- 52
AY028439    MGSANPDNKN------SMTKEEEEACLSAMRLASASVLPMVLKSAIELDLLELIKKSGP- 53
AF064694    MGST-QKNHN------LTPEEEEEACLHAMQLASASVLPMVLKAAIELSVLEIIAKAGQ- 52
AB121046    MRSTSPQTIDSTPMSHPQQEEEEELGKQAIRLANVVILPMVLKSALELNVIDIIWGAGD- 59
AF344315    MGYLFEETLSSNPKTPIVVDDDNELGLMAVRLANAAAFPMVLKASLELGVFDTLYAEASR 60
AF344316    MGYLFQETLSSNPKTPIVVDDDNELGLMAVRLANAAAFPMVLKAALELGVFDTLYAAASR 60
                 *              ::*     *:**    :** **::**  ::: :

U70424      --SPMSPTEIASKLPTK--NPEAPVMLDRILRLLTSYSVLTCSNRKLSGDGVERIYGLGP 108
AY028439    -GAYVSPSELAAQLPTQ--NPDAPVMLDRILRLLASYSVLNCTLKDLPDGGIERLYSLAP 110
AF064694    -GAYVAPTEIASQLSTS--NSQAPIILDRILRLLASYKVLTCNLRTLEDGGVERLYGLAP 109
AB121046    -GESLSPSDIAAQLPTK--NSNAPAVLDRMLRLLASHSILKCSARTGSDGQVERLYSAGP 116
AF344315    TDSFLSPSEIASKLPTTPRNPGAPVLLDRMLRLLASYSMVKCEKVSVGKG--ERVYRAEP 118
AF344316    TDSFLSPYEIASKLPTTPRNPEAPVLLDRMLRLLASYSMVKCGKALSGKG--ERVYRAEP 118
                 ::*  ::*::*.*   *. ** :***:****:*:. ::.*      **:*    *

U70424      VCKYLTKNEDG--VSIAALCLMNQDKVLMESWYHLKDAILDGGIPFNKAYG-MSAFEYHG 165
AY028439    VCKFLTKNEDG--VSMAALLLMNQDKVLMESWYHLKDAVLEGGIPFNKAYG-MTAFEYHG 167
AF064694    VCKFLVKNEDG--VSIAPLVLMNQDKVLMESWYHLKDAVLDGGIPFNKAYG-MTAFEYHG 166
AB121046    ICKFLVKDQNGGTRSVGPLFLLHHDKVFMEGWFHLNDAILEGGIPFNRAYG-MTAFEYPE 175
AF344315    ICRFFLKNNIQDIGSLASQVIVNFDSVFLNTWAQLKDVVLEGGDAFGRAHGGMKLFDYMG 178
AF344316    ICRFFLKDNIQDIGSLASQVIVNFDSVFLNTWAQLKDVVLEGGDAFGRAHGGMKLFDYMG 178
            :*::: *::    *:.   ::: *.*:::  *  :*:*.:*:.** .*..:*:*  *.:*

U70424      TDPRFNKVFNNGMSNHSTITMKKILETYKGFEGLTSLVDVGGGIGATLKMIVSKYPNLKG 225
AY028439    KDPRFNKVFNQGMSNHSTIIMKKILEIYQGFQGLKTVVDVGGGTGATLNMIVSKYPSIKG 227
AF064694    TDPRFNKVFNRGMADHSTITMKKLLELYKGFEGLKSVVDVGGGTGATINMIVTKHPTIKG 226
AB121046    TDERFNRVFNQAMSNHTTLILKKILDVYRGFEGINVLVDVGGGIGVTLNLITNKYPHIKG 235
AF344315    TDERFSKLFNQ--TGFTIAVVKKALEVYQGFKGVNVLVDVGGGVGNTLGVVTSKYPNIKG 236
AF344316    TDERFSKLFNQ--TGFTIAVVKKALEVYEGFKGVKVLVDVGGGVGNTLGVVTSKYPNIKG 236
            .* **.::**.   :..:   :** *: *.**:*:. :******  *: ::.   *:* :**

U70424      INFNLPHVIEDAPSHPGIEHVGGDMFVSVPKGDAIFMKWICHDWSDEHCVKFLKNCYESL 285
AY028439    INFDLPHVIEDAPSYPGVDHVGGDMFVSVPKGDAIFMKWICHDWSDAHCLKFLKNCHEAL 287
AF064694    INFDLPHVIDDAPAYPGVEHIGGDMFVSVPKGDAIFMKWILHDWSDEHSVKFLKNCYESI 286
AB121046    INFDLPHVLADAPSYPGVEHVGGDMFKSVPQGDAIFMKWILHDWSDEHCLTLLKNCCKSL 295
AF344315    INFDLTCALAQAPSYPGVEHVAGDMFVDVPTGDAMILKRILHDWTDEDCVKILKNCWKSL 296
AF344316    INFDLTCALAQAPSYPGVEHVAGDMFVDVPTGDAMILKRILHDWTDEDCVKILKNCWKSL 296
            ***:*.   .: :**::**::*:*:.**** .** ***::.:* * ***:* .  : :****  :::

U70424      PEDGKVILAECILPETPDS-SLSTKQVVHVDCIMLAHNPGGKERTEKEFEALAKASGFKG 344
AY028439    PENGKVILAECLLPEAPDS-TLSTQNTVHVDVIMLAHNPGGKERTEKEFEALAKGAGFRG 346
AF064694    PADGKVIIVESILPVYPET-NLASNACFQLDNIMLAHNPGGKERTEKDFEALSAKAGFTG 345
AB121046    PSSGKVIFVESILPEVPDS-TVTSNIVCEQDLLMFTQNPGGKERTKKEYEALALKSGFSR 354
AF344315    PENGKVVVIELVTPDEAENGDINANIAFDMDMLMFTQCSGGKERSRAEFEALAAASCFTH 356
AF344316    PENGKVVVIELVTPDEAENGDINANIAFDMDMLMFTQCSGGKERSRAEFEALAAASGFTH 356
            * .***:. * : *   .:.  : ::       *  :*::: .*****:.  ::*****:    : *

U70424      IKVVCDAFGVNLIELLKKL 363
AY028439    FIKVCCAYNSWIMELLK-- 363
AF064694    FKIVCGAFGSWVMEFCK-- 362
AB121046    LEVVCSAYNSWVMEFHK-- 371
AF344315    CKFVCQAYHCWIIEFCK-- 373
AF344316    CKFVCQAYHCWIIEFCK-- 373
            .* ** *:     ::*: .
```

Figure 9: Multiple sequence alignment of AF344315 and AF344316 with different biochemically characterized OMTs (refer to Table 3 for substrate preferences) from various plant species. The color scheme represents amino acids grouped according to their physico-chemical properties.
* **Represents strictly conserved regions.**

monitored by Western blot analysis of cell lysates. Membranes were probed with *Arabidopsis* anti-3'-OMT and detected using the systems described in Materials and Methods section. As shown in Figure 10A, the antibody immuno-recognized a 45 kDa protein band in the crude cell lysates carrying the pTrc-His-C00113 construct after IPTG induction. The $(His)_6$-tag and amino acids resulting from the cloning process account for an additional ~ 3.5 kDa in the fusion protein. Partial purification by ammonium sulfate precipitation and affinity chromatography on a metal ligand affinity resin were carried out and protein expression was monitored by immunoblotting (Figure 10 B & C).

The expression profile of recombinant C00114 protein was different from that of recombinant C00113 in that the level of expression was lower at the condition optimized for recombinant C00113 at 37 $^\circ$C for 4 h. Hence, after discovering that the expressed protein was present in inclusion bodies (Figure 11 A), induction by IPTG was carried out at either room temperature or 30°C for 16 h, which was optimal for the expression of C00114 (Figure 11 B). The Western blot profile of a Ni-NTA-purified recombinant protein shows a band at the expected molecular weight of 45 kDa (Figure 11 C).

G.1.3   Determination of Recombinant OMT Activity in *E. coli*

Enzyme assays were first carried out with recombinant proteins from crude cell lysates of both C00113 and C00114 using a variety of potential phenolic substrates from different classes; including benzoic acids, hydroxycinnamic acids, coumarins, flavonoids, isoflavonoids, and indolic substrates (Table 2). When crude preparations of the recombinant proteins did not exhibit OMT activity, partially purified and affinity-purified proteins were utilized. Nevertheless, this yielded no specific OMT activity regardless of
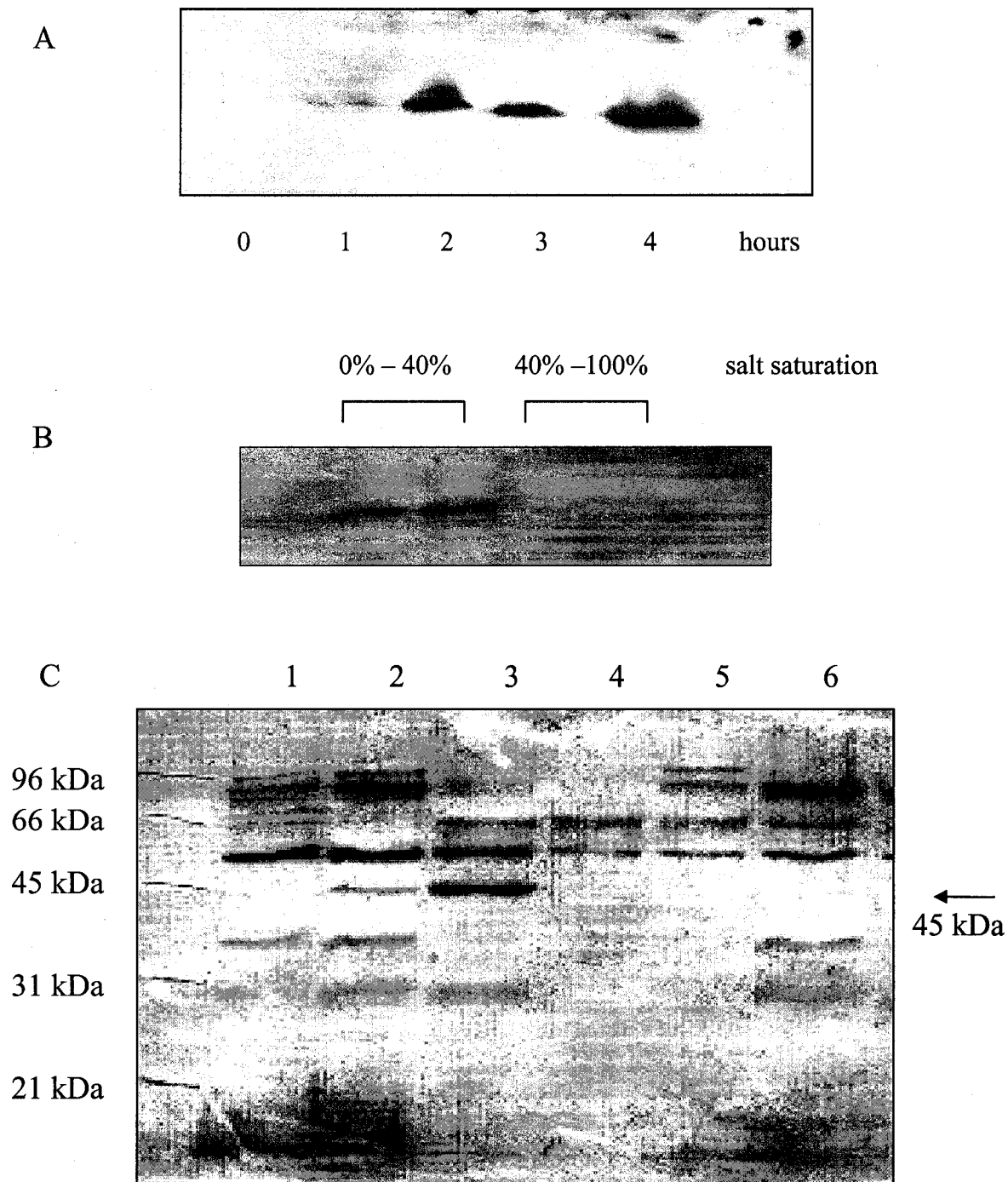
Figure 10: Expression of recombinant C00113 protein in *E. coli* as detected by immunoblotting

A) Time course induction at 37 °C for 4 h.
B) Partial purification by ammonium sulphate precipitation (in duplicates).
C) Purification on Ni-NTA; Lane: **1**, Non-induced (flow-through);
   2, Induced (flow-through); **3**, Eluate; **4** to **6**, Washes.
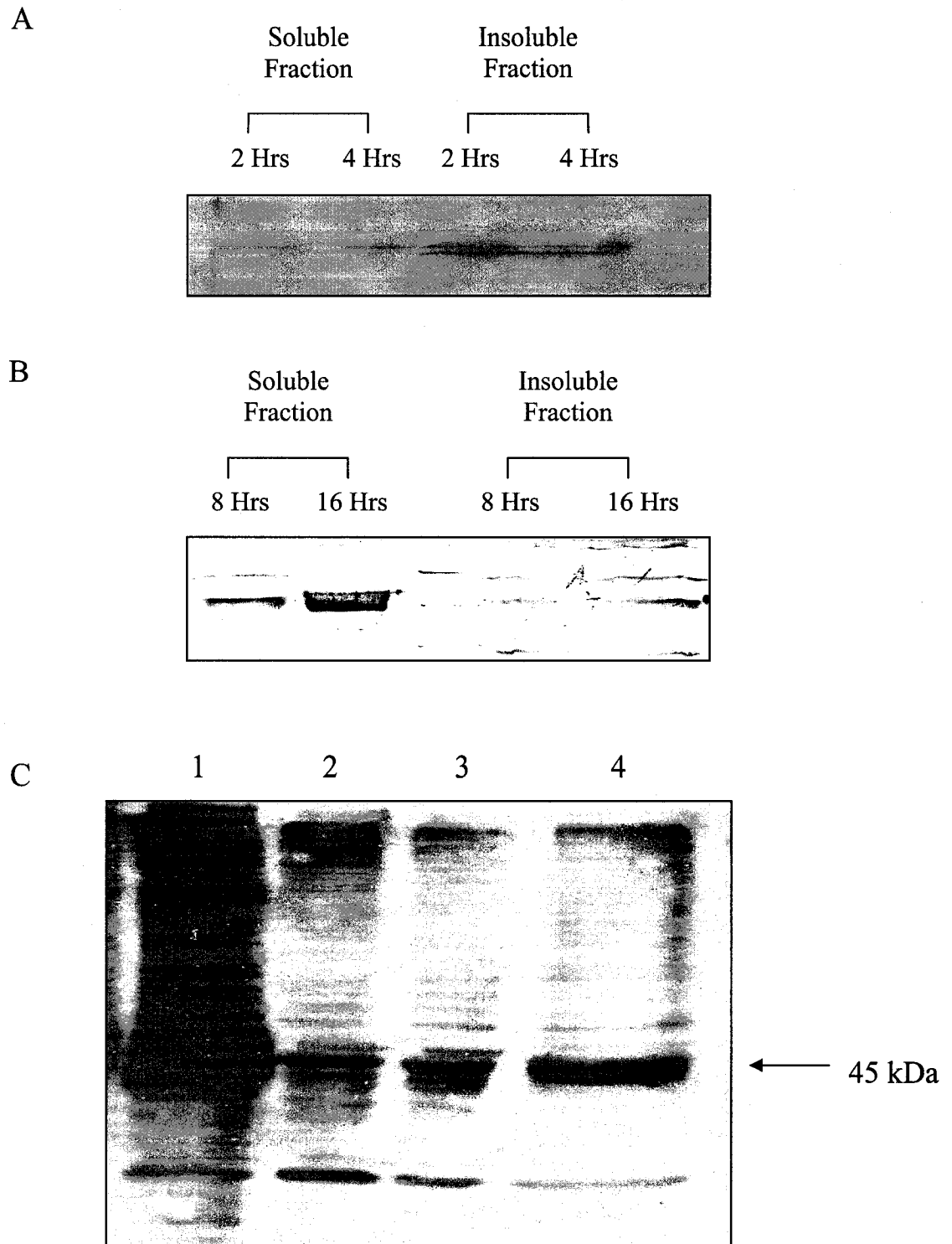
Figure 11: Expression of recombinant C00114 protein in *E. coli* as detected by immunoblotting

A) Time course induction at 37°C for 4 hours.

B) Time course induction at 30°C for 16 hours.

C) Purification on Ni-NTA; Lane: **1**, Flow-through; **2**, Wash #1; **3**, Wash #2; **4**, Eluate.

Table 2: OMT assay of C00113 and C00114 recombinant proteins using various substrates [a].

Substrate[b]	Protein[c]
Hydroxycinnamic acids:	
5-Hydroxyferulic acid	A.P, C00114, *E. coli*
Caffeic acid	C.L, C00113, *P. pastoris*
Hydroxycinnamoyl CoA-ester:	
Caffeoyl CoA	A.P, C00114, *E. coli*
Benzoic acids:	
Benzoic acid	A.P, C00114, *E. coli*
Salicylic acid	A.P, C00113, *P. pastoris*
2,3-Di-hydroxybenzoic acid	C.L, C00113, *P. pastoris*
2,4-Di-hydroxybenzoic acid	P.P, C00114, *E. coli*
2,5-Di-hydroxybenzoic acid	C.L, C00113, *E. coli*
2,6-Di-hydroxybenzoic acid	P.P, C00113, *P. pastoris*
3,4-Di-hydroxybenzoic acid	A.P, C00114, *E. coli*
Coumarins:	
Daphnetin	A.P, C00113, *P. pastoris*
Esculetin	A.P, C00114, *E. coli*
(iso)Flavonoids:	
Quercetin	C.L, C00113, *P. pastoris*
3,7,4'-Trimethylquercetin	P.P, C00113, *E. coli*
Apigenin	C.L, C00113, *P. pastoris*
Luteolin	A.P, C00113, *P. pastoris*
Naringenin	P.P, C00114, *E. coli*
Genistein	A.P, C00113, *P. pastoris*
Indolic:	
4-Hydroxyindole	A.P, C00114, *E. coli*
5-Hydroxyindole	A.P, C00113, *P. pastoris*
Indole 3-ethanol	C.L, C00114, *E. coli*
Simple sugar and di-/trihydric phenols:	
Myo-inositol	A.P, C00114, *E. coli*
phloroglucinol	C.L, C00113, *E. coli*
Catechol	C.L, C00114, *E. coli*

[a] No specific activity was observed to be significantly above controls (less than 5%).

[b] Refer to Figure 21 for the chemical structures of various compounds utilized by OMTs.

[c] For illustration purposes, different levels of protein purification (C.L: crude lysate; P.P: partially-purified; A.P: affinity-purified) from either of C00113 or C00114 and either of *E. coli* or *P. pastoris* recombinant proteins were randomly chosen and depicted.

the substrate, and a series of different conditions were attempted in order to identify any potential enzymatic activity. These conditions include different substrate concentrations (50 μM to 500 μM), varying the amounts of recombinant protein (10 μg to 100 μg), modifying buffer systems (in the presence/absence of EDTA or Mg^{2+}), as well as different reaction product extraction methods (benzene-ethyl acetate mixture, diethyl ether-ethyl acetate mixture, or hexane). However, no OMT activity was detected (Table 2), and it is not clear whether the recombinant proteins are inactive or the appropriate substrates have yet to be identified.

G.1.4 Expression of C00113 in *P. pastoris*

The expression of C00113 in a prokaryotic system may not provide the appropriate conditions for eukaryotic protein maturation. Therefore, such post-translational modifications and protein folding limitations can be addressed by expression of the recombinant protein in a eukaryotic system, *P. pastoris*. This system is relatively easy to manipulate and displays many of the advantages of higher eukaryotic expression systems such as protein processing, folding and modification. Through the enzymatic activity of alcohol oxidase (AOX1), *P. pastoris* metabolizes methanol as its sole carbon source. The promoter regulating the production of AOX1 drives heterologous protein expression in the presence of methanol and absence of glucose. Furthermore, depending on the presence of a signal peptide located N-terminally to the heterologous protein, production of recombinant proteins can be either intracellular or secreted. Since *P. pastoris* secretes very low levels of native proteins, secreted recombinant proteins have the advantage of a higher yield to purity ratio as compared to intracellular expressed

A

Day   0   1   2   3   4   5

B

96 kDa

66 kDa

45 kDa   ←——— 45 kDa

31 kDa

21 kDa

1   2   3   4   5
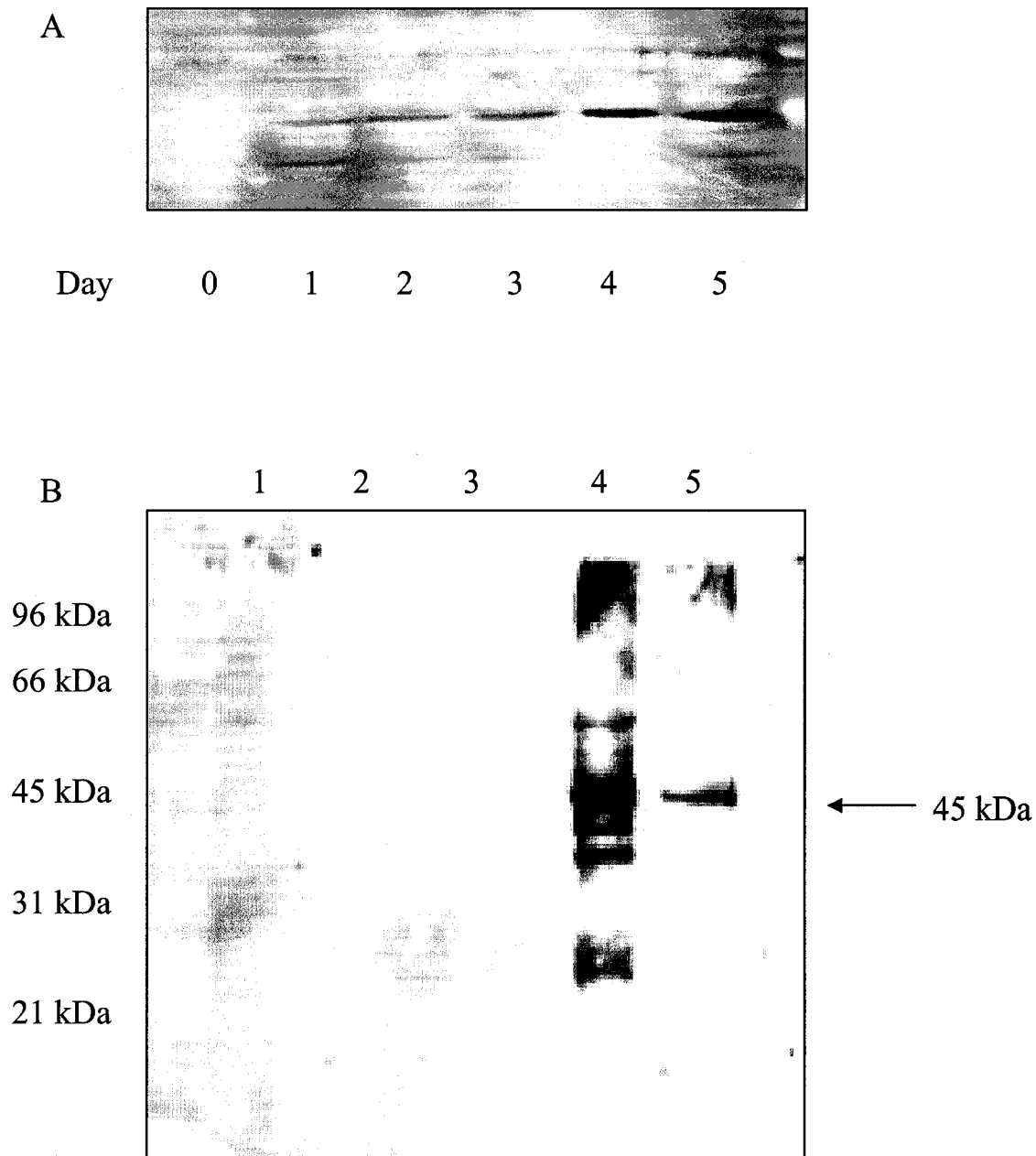
Figure 12: Expression of intracellular recombinant C00113 protein in *P. pastoris* as detected by immunoblotting

   A) Time course induction at 30°C for 5 days.
   B) Purification on Ni-NTA; Lane: **1**, Flow-through; **2** and **3**, Washes; **4** and **5**, Eluates.

A

Day   0   1   2   3   4   5

B   1   2   3   4   5   6

← 45 kDa
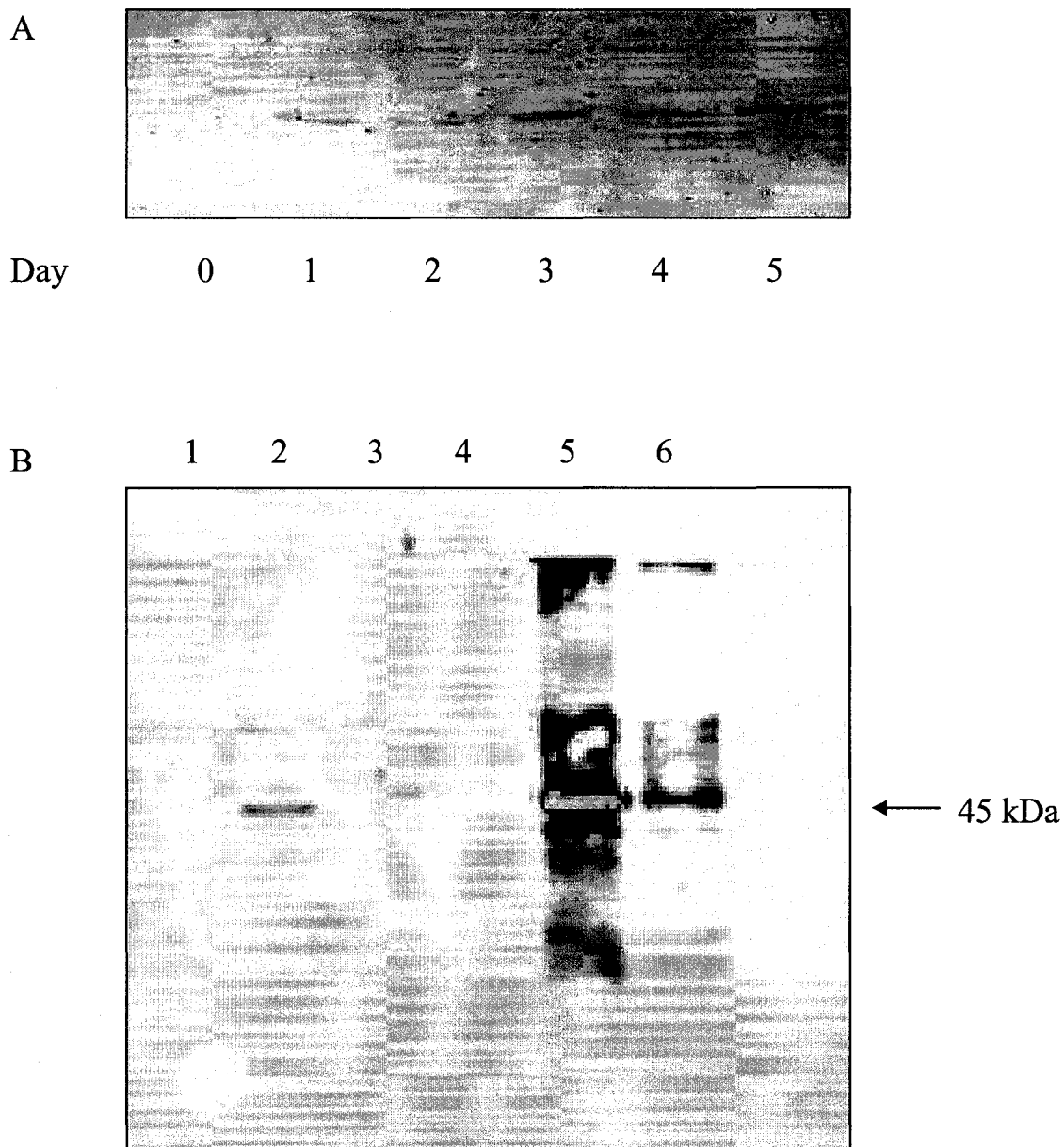
Figure 13: Expression of secretory recombinant C00113 protein in *P. pastoris* as detected by immunoblotting

A) Time course induction at $30^{\circ}C$ for 5 days.

B) Purification on Ni-NTA; Lane: **1**, Flow-through; **2** and **3**, Washes; **4** to **6**, Eluates.

protein. Purification of recombinant proteins was carried out using a metal-affinity ligand; although the polyhistidine tag is located at the N-terminal for the intracellular expression construct and at the C-terminal for the secretory system.

Recombinant protein induction was carried out in a buffered culture medium as described in the Materials and Methods. In addition, the media (BMGY/BMMY) contained protein stabilizing factors, such as peptone and yeast extracts that have been demonstrated to reduce proteolysis, particularly of secreted proteins (Brierley et al., 1994; Clare et al., 1991). Methanol-induced expression of recombinant protein was carried out over a 5-day period, for both intracellular and secretory systems. Protein expression, monitored by Western blot analysis using anti-$(His)_6$ antibodies, was optimal 4 to 5 days post-induction in both systems (Figure 12 A & 13 A). Ni-NTA purified recombinant proteins of both systems showed a similar protein band of 45 kDa (Figure 12 B & 13 B).

G.1.5   Determination of Recombinant Protein Activity in P. pastoris

Intracellular recombinant protein was subjected to buffer-exchange on a PD-10 column against buffer B for immediate use in enzyme assays. The secreted recombinant protein was first subjected to ammonium sulfate concentration, followed by buffer exchange against buffer B, before enzyme assays were carried out. When no OMT activity was detected for both the crude and partially-purified proteins against all the potential substrates tested (Table 2), recombinant proteins from both systems were subjected to further purification on Ni-NTA resin and conditions for enzymes assay were

altered as in the case for *E. coli*. These manipulations, nevertheless, did not produce any OMT activity above control levels (Table 2).

G.2   Screening of *Serratula tinctoria* cDNA Library

G.2.1   PCR-based Screening

To facilitate the isolation of a putative flavonol 3-OMT clone from *S. tinctoria*, degenerate oligonucleotide primers based on tryptic peptide sequences obtained from partially purified native 3-OMT (Huang *et al.*, 2004), were designed for use in a PCR-based screening of a *S. tinctoria* cDNA bacterial library. This strategy (Figure 14) involved the use of numerous primers and primer combinations as outlined in Figure 15.

In order to minimize the number of false positives, two consecutive rounds of screening were carried out. Whenever possible, primer combinations that would generate the longest potential fragments were utilized. Certain combinations, however, did not amplify any fragments and/or produced false positives regardless of optimization. The first round of screening involved the use of primer sets consisting of either one degenerate, peptide-specific forward primer (Pep1F or Pep3.1F and so on until Pep5.2F) and reverse vector primer Sport6F2 or two peptide-specific primers (forward primers, Pep1F or Pep3.1F and so on until Pep4B.2F, with reverse primers, Pep6.1R or Pep6.2R or Pep5.1R or Pep5.2R).

PCR fragments from the first round of screening were then used as templates for a second round of screening. The second round of screening involved the use of nested primer sets, either peptide-specific primers with vector primers, or peptide-specific primers alone. Nested primer combinations include forward primers, Pep3.1F or Pep3.2F

```
┌─────────────────────────────────────────────────────┐
│  First round of PCR screening using different primers │
│              and primer combinations                  │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│   Second round of PCR screening using yet another     │
│        set of primers and primer combinations         │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│  Amplified fragments subjected to Southern analysis   │
│   using a radio-labelled *Arabidopsis* 3'-OMT full-   │
│        length cDNA as a heterologous probe            │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│   Putative positive fragments cloned into pCR4Blunt-  │
│    TOPO vector and plasmid analysis for verification  │
│     of the presence of inserts, either by restriction │
│              analysis or colony-PCR                    │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│  Further verification of putatively positive clones by│
│               sequencing reaction                     │
└─────────────────────────────────────────────────────┘
```

Figure 14: PCR-based cloning strategy.
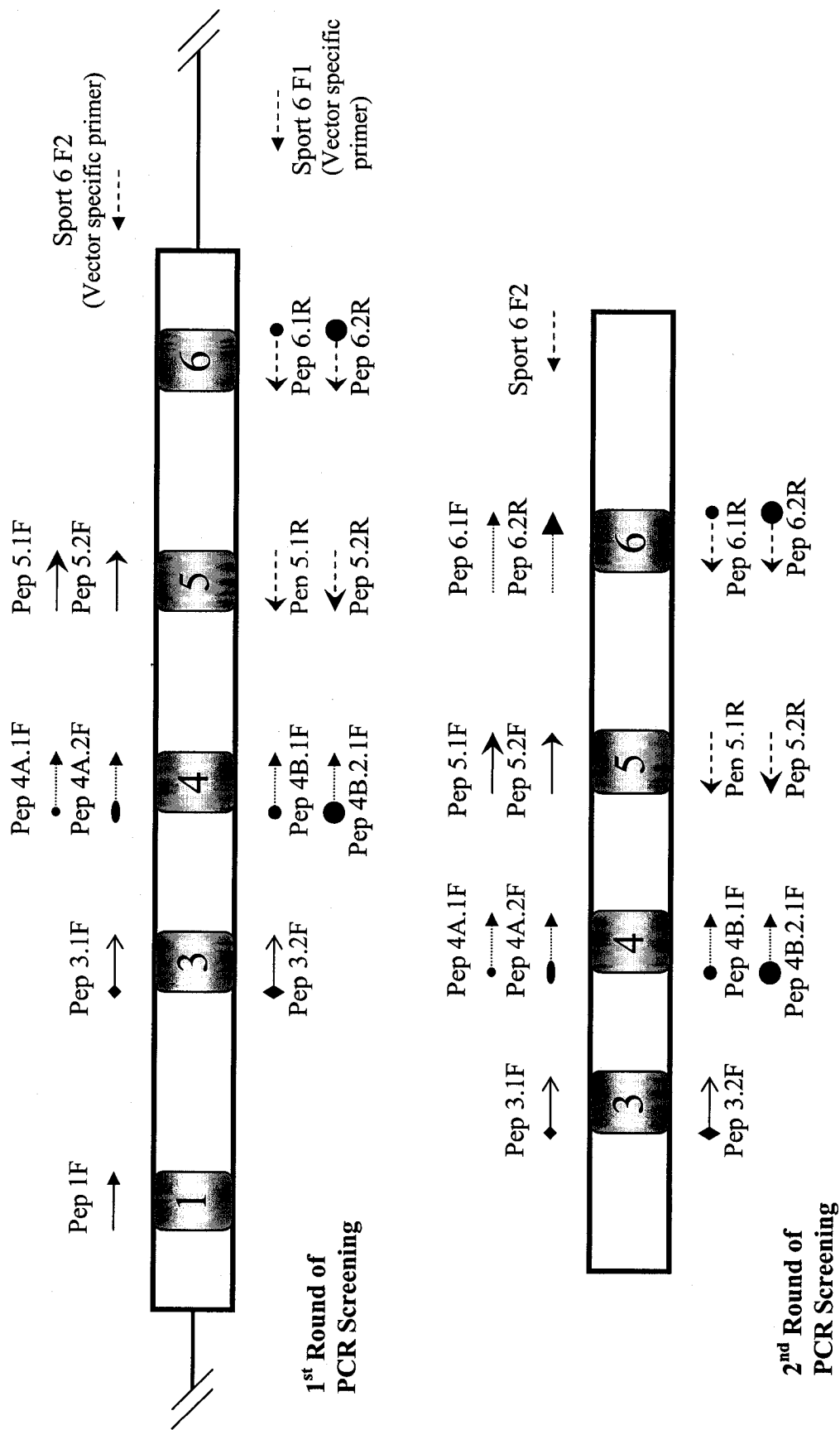
Figure 15: Primers and primer combinations used for the PCR-based screening strategy.

77

and so on until Pep6.2F, with reverse primers, vector primer Sport6F1 or Pep5.1R or Pep5.2R or Pep6.1R or Pep6.2R). For illustration purpose, if Pep1F and Sport6F2 were employed for the first round of amplification, the resulting fragment was then used as a template for a second round of amplification using a set of nested primers, such as Pep3.1F and Sport6F1 or Pep4A.1 and Pep6.1R and so forth.

Following Southern analysis with [α-^{32}P]dCTP-labelled *Arabidopsis* 3'-OMT full-length cDNA as a heterologous probe for hybridization (a choice probe simply because of its availability and being biochemically characterized (Muzac *et al.*, 2000)), any putatively positive fragments were then cloned and sequenced. The sequence information obtained was analyzed by BLAST and further verified to assess whether any of the six peptides obtained from the partially purified native protein of *S. tinctoria* were present in these sequences. When no such match was found, modifications of the PCR conditions were attempted for each primer set combination as described previously, thereby creating a multitude of PCR reactions that covered most possible combinations of primers and PCR conditions. These modifications include, different annealing temperatures (from 38°C to 48°C) for both rounds of screening, different numbers of cycles, (10 to 12 cycles for the touchdown step and 40 to 45 cycles following touchdown step), and different amounts of primers. Despite these efforts, PCR products containing sequence matches to the expected peptides (based on the combination of primers used) of *S. tinctoria* 3-OMT native protein could not be identified.

Although the isolation of quercetin 3-OMT clone has been unsuccessful, several putative OMT cDNA-like clones have been identified. The initial BLAST analysis of the 27 clones obtained showed them to be closely related to caffeic acid OMTs. This is not

unexpected since the lignin OMT in *S. tinctoria* utilizes both caffeic acid and 5-hydroxyferulic acids as substrates. Nonetheless, these putative OMT cDNA clones could very well encode OMTs that catalyze *O*-methylation of different classes of substrates. Further analysis of the sequences revealed that they all contain most, if not all, of the five characteristic motifs of plant OMTs (Ibrahim *et al.* 1998). Furthermore, certain sequences exhibit partial matches to some of the peptide sequences obtained from the *S. tinctoria* 3-OMT. Given the fact that none of the vector primers were located at the 5' end, full-length cDNA fragments could not be isolated using this strategy. Therefore, in order to carry out further analysis of these clones, specific primers will have to be designed from each of the clones for further isolation of missing sequences belonging to either the 5' end and/or 3' end from the cDNA library.

G.2.2   Hybridization Screening/Colony Lift Screening

The cDNA library (~50,000 clones) was also screened with (i) *Arabidopsis* 3'-OMT full-length cDNA random-labeled probe; or (ii) terminally-labeled degenerate oligonucleotide probes (DegPep1F, DegPep4F, and DegPep5F). The former probe was used because it is biochemically characterized and is readily available with the hope of isolating any full-length putative OMT clones, while the latter probes were used as a final attempt at isolating the 3-OMT clone of *S. tinctoria.* Due to the nature of the oligonucleotides used, the protocols for the screening strategy differed as described in Materials and Methods and as outlined in Figure 16. In order to minimize the number of false positives in each case, two rounds of screening were carried out, allowed by verification by Southern analysis before any putatively positive clones were sequenced.
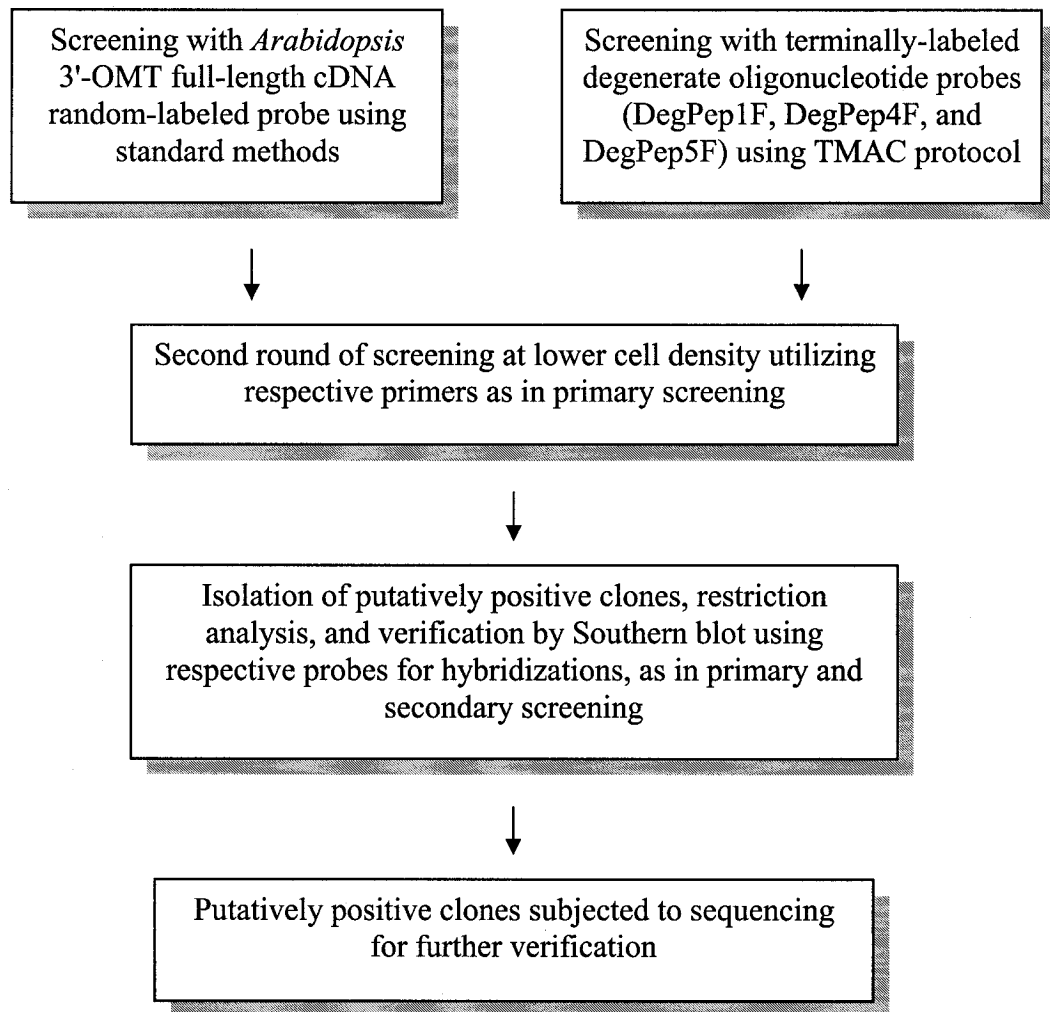
Figure 16: Hybridization screening/colony lift screening strategies.

Screening with the *Arabidopsis* 3'-OMT full-length cDNA random-labeled probe under high stringency conditions yielded 11 putative OMT clones upon analysis of the sequence using BLAST. Interestingly, further analysis revealed that 9 out of the 11 cDNA clones have a 100% match with tryptic peptide 3 and a partial match with tryptic peptide 4 of the native 3-OMT of *S. tinctoria*. These 9 clones are most likely derived from the same mRNA species as local alignment of their nucleotide sequences show that they are 93% to 99% identical. Among these 11 clones, 6 provided sequences sufficient for contig development and revealed three full-length cDNA clones, StOMT1, StOMT2 and StOMT3, respectively. At the nucleotide level, StOMT1 and StOMT2 are 99% identical in their coding region, 94% in its 5' untranslated region, and 97% in its 3' untranslated region. StOMT3 is 79% identical to both StOMT1 and StOMT2 and shows homology in their 5' and 3' untranslated regions. The first ATG at the 5' ends of the three clones contains the sequence AAA<u>A</u>ATGGC<u>T</u>, consistent with the consensus sequence (TA<u>AA</u>C<u>A</u>ATGGCT) for translation initiation sites proposed for plant genes (Joshi, 1987). All three clones have TAA as the predicted stop codon (Figure 17, 18, 19).

Attempts to isolate the 3-OMT clone of *S. tinctoria,* employing the TMAC (Tetramethylammonium chloride) protocol with degenerate terminal-labeled probes, were not successful. The two successive rounds of screening, carried out under both stringent and relaxed conditions, though yielded many putative positives, sequencing results confirmed their being false positives.

```
ccttcctcaagacctttccggattataattccgtccattcaaattgaagctgccactaac
P   S   S   R   P   F   R   I   I   I   P   S   I   Q   I   E   A   A   T   N
tcaattacaagttagaatatatataaaatttaaaaaatggcttcacagaagaatcatcat
S   I   T   S   -   N   I   Y   K   I   -   K   M   A   S   Q   K   N   H   H
gaggaggaggaggaggaagctttcctgtttgctatgcaattggcaagtgcgtcggtgtta
E   E   E   E   E   A   F   L   F   A   M   Q   L   A   S   A   S   V   L
cccatggtgttgaaatctgccatagacttggatttgcttgagatcattgccaaagctggc
P   M   V   L   K   S   A   I   D   L   D   L   L   E   I   I   A   K   A   G
cctggtgcctatgtgtcccctgcatacctggctgctcagctcccaaaggcagacaacccg
P   G   A   Y   V   S   P   A   Y   L   A   A   Q   L   P   K   A   D   N   P
gaggcacccgtgatgctcgacaggatatgccgcttgttggctacctactctgtcttgacc
E   A   P   V   M   L   D   R   I   C   R   L   L   A   T   Y   S   V   L   T
tgtaccctcaataaggtagacgatgatgttgttgagcgtctctatggtctggccccggtc
C   T   L   N   K   V   D   D   D   V   V   E   R   L   Y   G   L   A   P   V
tgcaaattcttggtgaagaatgaagatggtgtttccaatgcacctctcttgcttatgaac
C   K   F   L   V   K   N   E   D   G   V   S   N   A   P   L   L   L   M   N
caggataagattctaatggagagctggtatcatctgaaagatgctgttttggatggtgga
Q   D   K   I   L   M   E   S   W   Y   H   L   K   D   A   V   L   D   G   G
attccattcaacaaggcttatggcatgagtgctttcgagtatcatgggaaggatgataga
I   P   F   N   K   A   Y   G   M   S   A   F   E   Y   H   G   K   D   D   R
ttcaacaaggtcttcaacagtggcatgtttaaccactccaccatgaccatgaagaagatc
F   N   K   V   F   N   S   G   M   F   N   H   S   T   M   T   M   K   K   I
ctggatgtgtatcctggtttcgaagggggttcaaaccctggtagatgttggtggtggcacc
L   D   V   Y   P   G   F   E   G   V   Q   T   L   V   D   V   G   G   G   T
ggtgccagtctcagtatgatcctctccaagcatccttcaatcaagggtatcaattttgat
G   A   S   L   S   M   I   L   S   K   H   P   S   I   K   G   I   N   F   D
ttgccacatgtcattcaagacgcaaccacttatcctggtattgagcatgttggcggggac
L   P   H   V   I   Q   D   A   T   T   Y   P   G   I   E   H   V   G   G   D
atgtttgaaagcgtgccaaaaggagatgccattttcatgaagtggatctgccatgactgg
M   F   E   S   V   P   K   G   D   A   I   F   M   K   W   I   C   H   D   W
agcgatgcacactgtgtgaaattcctcaacaactgttacaaggcactccctgagaatgga
S   D   A   H   C   V   K   F   L   N   N   C   Y   K   A   L   P   E   N   G
aaagtgatcgtggcggaatgcatccttccggagactccagactcgagcctggccaccaaa
K   V   I   V   A   E   C   I   L   P   E   T   P   D   S   S   L   A   T   K
aacgtggtgcacatcgacgtcatcatgttggcccacaacccccggtggcaaagagaggacc
N   V   V   H   I   D   V   I   M   L   A   H   N   P   G   G   K   E   R   T
gagaaagaattccaagcattggccaaggctgccggcttcaacggattccgtaaggcctgc
E   K   E   F   Q   A   L   A   K   A   A   G   F   N   G   F   R   K   A   C
tctgcactcaacacctggcttatggagtttaccaaaTAAttttaataattcttcttcttt
S   A   L   N   T   W   L   M   E   F   T   K   -   F   -   -   F   F   F   F
ttcttttttctttttacaaaaacatatgcgccttcttttctcttcaactgagtgattttaa
F   F   F   F   L   Q   K   H   M   R   L   L   F   S   S   T   E   -   F   -
tttgtttgtttatccaatacaaattattaatcatatcatatcttatatgaatatgaatat
F   V   C   L   S   N   T   N   Y   -   S   Y   H   I   L   Y   E   Y   E   Y
gatgatcatgcatgtaccccccactgcttgtgaattaattcaatcaataaatacaaacatg
D   D   H   A   C   T   P   T   A   C   E   L   I   Q   S   I   N   T   N   M
ttttttttt
F   F   F
```

Figure 17: Nucleotide and deduced amino acid sequence of StOMT1. The translation initiation site is depicted in bold and underlined. The stop codon is indicated in upper-case letter in bold.

```
ttccgtccattcaaattgaagctgccactaactcaattacaagttagaatatatatat
F  R  P  F  K  L  K  L  P  L  T  Q  L  Q  V  R  I  Y  I  Y
aaaattttaaaaatggcttcacagaagaatcatcatgaggaggaagctttcctgtttgct
K  I  L  K  M  A  S  Q  K  N  H  H  E  E  E  A  F  L  F  A
atgcaattggcaagtgcgtcggtgttacccatggtgttgaaatctgccatagacttggat
M  Q  L  A  S  A  S  V  L  P  M  V  L  K  S  A  I  D  L  D
ttgcttgagatcattgccaaagctggccctggtgcctatgtgtcccctgcatacctggct
L  L  E  I  I  A  K  A  G  P  G  A  Y  V  S  P  A  Y  L  A
gctcagctcccaaaggcagacaacccggaggcacccgtgatgctcgacaggatatgccgc
A  Q  L  P  K  A  D  N  P  E  A  P  V  M  L  D  R  I  C  R
ttgttggctacctactctgtcttgacctgtaccctcaataaggtagacgatgatgttgtt
L  L  A  T  Y  S  V  L  T  C  T  L  N  K  V  D  D  D  V  V
gagcgtctctatggtctggcccccggtctgcaaattcttggtgaagaatgaagatggtgtt
E  R  L  Y  G  L  A  P  V  C  K  F  L  V  K  N  E  D  G  V
tccaatgcacctctcttgcttatgaaccaggataagattctaatggagagctggtatcat
S  N  A  P  L  L  L  M  N  Q  D  K  I  L  M  E  S  W  Y  H
ctgaaagatgctgtttttggatggtggaattccattcaacaaggcttacggcatgagtgct
L  K  D  A  V  L  D  G  G  I  P  F  N  K  A  Y  G  M  S  A
ttcgagtatcatgggaaggatgatagattcaacaaggtcttcaacagtggcatgtttaac
F  E  Y  H  G  K  D  D  R  F  N  K  V  F  N  S  G  M  F  N
cactccaccatgaccatgaagaagatcctggatgtgtatcctggtttcgaaggggttcaa
H  S  T  M  T  M  K  K  I  L  D  V  Y  P  G  F  E  G  V  Q
accctggtagatgttggtggtggcaccggtgccagtctcagtatgatcctctccaagcat
T  L  V  D  V  G  G  G  T  G  A  S  L  S  M  I  L  S  K  H
ccttcaatcaagggtatcaattttgatttgccacatgtcattcaagacgcaaccacttat
P  S  I  K  G  I  N  F  D  L  P  H  V  I  Q  D  A  T  T  Y
cctggtattgagcatgttggcggggacatgtttgaaagcgtgccaaaaggagatgccatt
P  G  I  E  H  V  G  G  D  M  F  E  S  V  P  K  G  D  A  I
ttcatgaagtggatctgccatgactggagcgatgcacactgtgtgaaattcctcaacaac
F  M  K  W  I  C  H  D  W  S  D  A  H  C  V  K  F  L  N  N
tgttacaaggcactccctgagaatggaaaagtgatcgtggcggaatgcatccttccggag
C  Y  K  A  L  P  E  N  G  K  V  I  V  A  E  C  I  L  P  E
actccagactcgagcctggccaccaaaaacgtggtgcacatcgacgtcatcatgttggcc
T  P  D  S  S  L  A  T  K  N  V  V  H  I  D  V  I  M  L  A
cacaaccccggtggcaaagagaggaccgagaaagaattccaagcattggccaaggctgcc
H  N  P  G  G  K  E  R  T  E  K  E  F  Q  A  L  A  K  A  A
ggcttcaacggattccgtaaggcctgctctgcactcaacacctggcttatggagtttacc
G  F  N  G  F  R  K  A  C  S  A  L  N  T  W  L  M  E  F  T
aaaTAAttttaataattcttcttctttttcttttacaaaaacatatgcgccttcttttc
K  -  F  -  -  F  F  F  F  F  F  L  Q  K  H  M  R  L  L  F
tcttcaactgagtgattttaatttgtttgtttatccaatacaaattattaatcatatcat
S  S  T  E  -  F  -  F  V  C  L  S  N  T  N  Y  -  S  Y  H
atcttatatgaatatgaatatgatgatcatgcatgtacccccactgcttgtgaattaatt
I  L  Y  E  Y  E  Y  D  D  H  A  C  T  P  T  A  C  E  L  I
caatcaataaatacaaacatgttttttttc
Q  S  I  N  T  N  M  F  F  F
```

Figure 18: Nucleotide and deduced amino acid sequence of StOMT2. The translation initiation site is depicted in bold and underlined. The stop codon is indicated in upper-case letter in bold.

```
caacaatataatcatattcactaaacttatatttgcctcttcttttaaaaaatagttctgt
 N  N  I  I  I  F  T  K  L  I  F  A  S  S  F  K  K  -  F  C
aaaatgggttcatcatcagcatatgtgaatgttccggtggaagcaaaggaggacgaatcg
 K  M  G  S  S  S  A  Y  V  N  V  P  V  E  A  K  E  D  E  S
ttcctattcgccatgcaattggcgagtgcatccgtcctaccgatggccttgaaaaccgcc
 F  L  F  A  M  Q  L  A  S  A  S  V  L  P  M  A  L  K  T  A
atcgagctcgatttgctagaaaccatagccaaagcggggcccagcgcttccctttcatcc
 I  E  L  D  L  L  E  T  I  A  K  A  G  P  S  A  S  L  S  S
tcccagctggcagctcagctccccttggttaacaaccccgaagcacccgtgatgctggac
 S  Q  L  A  A  Q  L  P  L  V  N  N  P  E  A  P  V  M  L  D
cgcatctgcagtcttctggctagccattccgtgctcacttgtagcctaaaggaccttgac
 R  I  C  S  L  L  A  S  H  S  V  L  T  C  S  L  K  D  L  D
gatggtggcgttgagcggctctacggtttggccccggtttgcaagttcttgaccaagaac
 D  G  G  V  E  R  L  Y  G  L  A  P  V  C  K  F  L  T  K  N
caagatggtgtgtcaatggcacccctcttgctcatgaaccaagataaggttctcatggag
 Q  D  G  V  S  M  A  P  L  L  L  M  N  Q  D  K  V  L  M  E
agctggtaccatttgaaagatgcggttttggacggtggaattccattcaacaaggcatat
 S  W  Y  H  L  K  D  A  V  L  D  G  G  I  P  F  N  K  A  Y
ggtatgtctgcctttgagtaccatggcaaagatgccagattcaacaaggtatttaacagt
 G  M  S  A  F  E  Y  H  G  K  D  A  R  F  N  K  V  F  N  S
ggcatgtttaatcattccaccatgaccatgaaaaagattgtagatttgtatgatggtttc
 G  M  F  N  H  S  T  M  T  M  K  K  I  V  D  L  Y  D  G  F
aatggcctcaccacccttgtggacgttggtggtggcaccggtgctagccttaacatgatc
 N  G  L  T  T  L  V  D  V  G  G  G  T  G  A  S  L  N  M  I
attgctaaacacacttcgctcaaggggatcaactttgatttgccccatgttattgaagac
 I  A  K  H  T  S  L  K  G  I  N  F  D  L  P  H  V  I  E  D
gctaccacttatcctggtattgagcatgttggaggcaatatgttcgaaagtgtgccaaaa
 A  T  T  Y  P  G  I  E  H  V  G  G  N  M  F  E  S  V  P  K
ggagatgccatctttatgaagtggatacttcatgattggagtgatgcacattgcctgaaa
 G  D  A  I  F  M  K  W  I  L  H  D  W  S  D  A  H  C  L  K
tttctcaagaactgctacgaagcacttccagccaacgggaaagtgatcgtggctgaatgc
 F  L  K  N  C  Y  E  A  L  P  A  N  G  K  V  I  V  A  E  C
attcttccagaggctccagactcgaccacggccacccaaaatgtggtgcacattgatgtg
 I  L  P  E  A  P  D  S  T  T  A  T  Q  N  V  V  H  I  D  V
atcatgttggctcataatcccggaggcaaagagagaaccgagaaggaatttgaggcctta
 I  M  L  A  H  N  P  G  G  K  E  R  T  E  K  E  F  E  A  L
gccaagggagccggattcaaaagtttcaacaaggctgcttgtgctcttaatacttgggtt
 A  K  G  A  G  F  K  S  F  N  K  A  A  C  A  L  N  T  W  V
atggaattttgtaaaTAGatttgagtttaagcatatgtgggtatggttgtggtttgatat
 M  E  F  C  K  -  I  -  V  -  A  Y  V  G  M  V  V  V  -  Y
gatggttcttttttggaaacaagcatggctttgattctaataaatatactgattttttata
 D  G  S  F  L  E  T  S  M  A  L  I  L  I  N  I  L  I  F  I
gttgatgagattcaaaaaaaaaaaaaagggcggccgc
 V  D  E  I  Q  K  K  K  K  R  A  A
```

Figure 19: Nucleotide and deduced amino acid sequence of StOMT3. The translation initiation site is depicted in bold and underlined. The stop codon is indicated in upper-case letter in bold.

G.2.3 Analysis of the Deduced StOMTs Amino Acid Sequences

The deduced amino acid sequences of StOMT1, StOMT2, and StOMT3 revealed

that their respective polypeptides, translated from the longest ORF beginning from first

start ATG to the first stop codon, contained 360, 357, and 364 amino acid residues,

respectively. The predicted molecular weights obtained for the three putatives OMT

clones were representative of OMTs that do not require the addition of Mg^{2+} for enzyme

activity, and these deduced sequences further revealed the presence of the five

characteristic motifs of plant OMTs (Ibrahim et al., 1998) (Figure 20). These preliminary

analyses indicate that the three clones exhibit the characteristics of Group II OMTs.

Comparison of the deduced amino acids of the three clones also showed that StOMT1

and StOMT2 exhibit sequence identity of 98% with one another, while showing 76% and

77% with StOMT3, respectively. Interestingly, the only difference between StOMT1 and

StOMT2 is that the latter contains three additional glutamic acid residues in succession at

position 12, 13 and 14, as compared to StOMT1 (Figure 20).

BLAST analysis revealed that all three clones are closely related to other plants

OMTs. StOMT1 exhibited 89% sequence similarity and 80% sequence identity with the

hydroxycinnamic acid OMT of Zinnia (U19911; Ye and Varner, 1995) followed by the

multifunctional OMT of periwinkle (AY028439; Schröder et al., 2002), which showed

89% and 78% sequence similarity and identity, respectively. Likewise, StOMT3 showed

93% sequence similarity and 86% sequence identity with the caffeic acid OMT of Zinnia

(U19911) followed by the multifunctional OMTs of C. americanum (U16793; Gauthier et

al., 1998) exhibiting 83% and 71% sequence similarity and identity, respectively.

```
StOMT1, StOMT2, StOMT3

CLUSTAL W (1.82) multiple sequence alignment


StOMT1    MASQKNHHEE----EEEEAFLFAMQLASASVLPMVLKSAIDLDLLEIIAKAGPGAYVSPA 56
StOMT2    MASQKNHH-------EEEAFLFAMQLASASVLPMVLKSAIDLDLLEIIAKAGPGAYVSPA 53
StOMT3    MGSSSAYVNVPVEAKEDESFLFAMQLASASVLPMALKTAIELDLLETIAKAGPSASLSSS 60
          *.*..  :         *;*;**************.**;**;***** ******.* :*.:


StOMT1    YLAAQLPKADNPEAPVMLDRICRLLATYSVLTCTLNKVDDDVVERLYGLAPVCKFLVKNE 116
StOMT2    YLAAQLPKADNPEAPVMLDRICRLLATYSVLTCTLNKVDDDVVERLYGLAPVCKFLVKNE 113
StOMT3    QLAAQLPLVNNPEAPVMLDRICSLLASHSVLTCSLKDLDDGGVERLYGLAPVCKFLTKNQ 120
           ******  .:***********  ***;:*****:*:..:**. **************.**;


StOMT1    DGVSNAPLLLMNQDKILMESWYHLKDAVLDGGIPFNKAYGMSAFEYHGKDDRFNKVFNSG 176
StOMT2    DGVSNAPLLLMNQDKILMESWYHLKDAVLDGGIPFNKAYGMSAFEYHGKDDRFNKVFNSG 173
StOMT3    DGVSMAPLLLMNQDKVLMESWYHLKDAVLDGGIPFNKAYGMSAFEYHGKDARFNKVFNSG 180
          ****  **********;********************************** *********


StOMT1    MFNHSTMTMKKILDVYPGFEGVQTLVDVGGGTGASLSMILSKHPSIKGINFDLPHVIQDA 236
StOMT2    MFNHSTMTMKKILDVYPGFEGVQTLVDVGGGTGASLSMILSKHPSIKGINFDLPHVIQDA 233
StOMT3    MFNHSTMTMKKIVDLYDGFNGLTTLVDVGGGTGASLNMIIAKHTSLKGINFDLPHVIEDA 240
          ************;*;* **;*: ************* .**;;*** * *:**********;**


StOMT1    TTYPGIEHVGGDMFESVPKGDAIFMKWICHDWSDAHCVKFLNNCYKALPENGKVIVAECI 296
StOMT2    TTYPGIEHVGGDMFESVPKGDAIFMKWICHDWSDAHCVKFLNNCYKALPENGKVIVAECI 293
StOMT3    TTYPGIEHVGGNMFESVPKGDAIFMKWILHDWSDAHCLKFLKNCYEALPANGKVIVAECI 300
          ***********;****************** ********;***;***;*** *********


StOMT1    LPETPDSSLATKNVVHIDVIMLAHNPGGKERTEKEFQALAKAAGFNGFRKACSALNTWLM 356
StOMT2    LPETPDSSLATKNVVHIDVIMLAHNPGGKERTEKEFQALAKAAGFNGFRKACSALNTWLM 353
StOMT3    LPEAPDSTTATQNVVHIDVIMLAHNPGGKERTEKEFEALAKGAGFKSFNKAACALNTWVM 360
          ***;***;. **;************************;****.***;;*.**..*****;*


StOMT1    EFTK 360
StOMT2    EFTK 357
StOMT3    EFCK 364
          ** *
```

Figure 20: Amino acids sequence alignment of StOMT1, StOMT2 and StOMT3 depicting the five OMTs signature motifs (bold and underlined) and sequence matches with peptides 3 and 4 (bold and italic) of the native S. *tinctoria* flavonol 3-OMT.

G.3   Phylogenomic Analysis of OMTs

Of the several hundred OMTs found in the databases (EMBL, GenBank and SwissProt), we have chosen 60 entries, that have been biochemically characterized to date (December 2004), for phylogenomic analysis. These include nine OMTs that utilize caffeoyl CoA esters (CCoAOMTs), 25 OMTs that utilize phenylpropanoid compounds (COMTs), 16 OMTs for flavonoids, four OMTs for carboxylic acids, and six OMTs for alkaloids, as substrates (Table 3 and Figure 21). Except for the first nine sequences which belong to Group I, all the others are Group II OMTs.

G.3.1   Maximum Parsimony Analysis

The analysis of the 60 OMT sequences, based on maximum parsimony searches, yielded two equally parsimonious trees with a length of 6374, consistency index (CI) of 0.544, retention index (RI) of 0.650, and a rescaled consistency index (RC) of 0.354. Figure 22 shows a strict consensus of these two trees and reveals that the 60 OMTs form two distinct clades. The first lineage (lineage A) includes the CCoAOMTs clade, sister to the carboxylic acid OMTs; whereas the other (lineage B), comprises the phenylpropanoid COMTs, (iso)flavonoid and alkaloid OMTs. Within lineage A, the CCoAOMTs and the carboxylic acid OMTs form two distinct subclades, each with a bootstrap support of 100%. In addition, the chickweed and ice plant OMTs group together forming a clade basal to other CCoAOMTs with a bootstrap value of 74%.

Table 3
Structural-functional features of some representative O-methyltransferase cDNA clones derived from different organisms, involved in the methylation of the indicated substrates

Accession number	Organism	Number of amino acids	Preferred substrates	Biological significance	Reference
Hydroxycinna-moyl CoA-esters OMTs					
U20736	*Medicago sativa* (Alfalfa)	247	Caffeoyl CoA > 5HFCoA > 5HF	Lignin biosynthesis	Inoue *et al.*, 1998
AY145521	*Mesembryanthemum crystallinum* (Ice plant)	237	Caffeoyl esters ≈ flavonoids	Phenylpropanoid metabolism	Ibdah *et al.*, 2003
U38612	*Nicotiana tabacum* (Tobacco)	239	Caffeoyl CoA, 5HFCoA	Lignin biosynthesis	Martz *et al.*, 1998
M69184	*Petroselium crispum* (Parsley)	241	Caffeoyl CoA	Disease resistance	Schmitt *et al.*, 1991
AF036095	*Pinus taeda*	259	Caffeoyl CoA > 5HFCoA	Lignin biosynthesis	Li *et al.*, 1999
AJ224894	*Populus balsamifera*	247	Caffeoyl CoA	Lignin biosynthesis	Meyermans *et al.*, 2000
L22203	*Stellaria longipes* (Chickweed)	241	Caffeoyl CoA	Phenylpropanoid metabolism	Zhang and Chinnappa, 1997
Z54233	*Vitis vinifera*	242	Caffeoyl CoA > 5HF-CoA	Induced resistance response	Busam *et al.*, 1997

Accession	Species	Size	Substrate	Pathway	Reference
U13151	*Zinnia elegans*	245	Caffeoyl CoA	Lignin biosynthesis	Ye et al., 1994

Hydroxycin-namic acids OMTs

Accession	Species	Size	Substrate	Pathway	Reference
AY443006	*Ammi majus*	354	Bergaptol	Furanocoumarin biosynthesis	Hehmann et al., 2004
AY443007	*Ammi majus*	365	Methyl caffeate > 5HF > caffeic acid	Phenylpropanoid metabolism	Hehmann et al., 2004
AY028439	*Catharanthus roseus* (periwinkle)	363	5HF > caffeic acid > flavonoids	Phenylpropanoid and flavonoid metabolism	Schröder et al., 2002
U16793	*Chrysosplenium americanum*	343	Flavonoids ≈ 5HF > caffeic acid	Flavonoid and phenylpropanoid metabolism	Gauthier et al., 1998
U86760	*Clarkia breweri*	368	Isoeugenol > eugenol	Floral scent biosynthesis	Wang et al., 1997
U54767	*Hordeum vulgare* (Barley)	376	Caffeic acid > catechol	Lignin biosynthesis	Lee et al., 1997
AF139533	*Liquidambar styraciflua* (Sweet-gum)	367	Coniferyl aldehyde	Lignin metabolism	Osakabe et al., 1999
AF010291	*Lolium perenne*	360	Caffeic acid	Lignin metabolism	McAlister et al., 1998
M63853	*Medicago sativa* (Alfalfa)	365	Caffeic acid > 5HF	Lignin biosynthesis (defense response)	Gowri et al., 1991

Accession	Species	No.	Substrate	Function	Reference
M87340	*Mesembryanthenum crystallinum* (Ice-plant)	365	*Myo*-inositol	Osmotic stress	Vernon and Bohnert 1992
X71430	*Nicotiana tabacum* (Tobacco)	365	o-Dihydric phenols	Defence response	Pellegrini *et al.*, 1993
AF435007	*Ocimum basilicum* (Basil)	356	Chavicol	Phenylpropene biosynthesis	Gang *et al.*, 2002
AF435008	*Ocimum basilicum* (Basil)	357	Eugenol	Phenylpropene biosynthesis	Gang *et al.*, 2002
U39301	*Pinus taeda*	381	Caffeic acid > 5HF and their CoA esters (bifunctional)	Lignin biosynthesis	Li *et al.*, 1997
X62096	*Populus tremuloides*	365	5HF > caffeic acid	Lignin metabolism	Bugos *et al.*, 1991
AB121046	*Rosa chinensis*	371	Phloroglucinol >>Resorcinol	Floral scent	Wu *et al.*, 2004
AF502433	*Rosa hybrida*	367	Orcinol > Orcinol monomethyl ether	Floral scent	Lavid *et al.*, 2002
AF502434	*Rosa hybrida*	366	Orcinol monomethyl ether > Guaiacol > Orcinol	Floral scent	Lavid *et al.*, 2002
AY177404	*Secale cereale* (Rye)	355	Daphnetin	Cold acclimation	Ndong *et al.*, 2003
AF064693	*Thalictrum tuberosum*	364	o-Dihydric phenols, BIQ alkaloids (Multifunctional)	Antimicrobial alkaloids	Frick and Kutchan 1999

Accession	Species		Substrate	Function	Reference
AF064694	Thalictrum tuberosum	362	o-Dihydric phenols	Antimicrobial alkaloids	Frick and Kutchan 1999
AF064695	Thalictrum tuberosum	362	o-Dihydric phenols	Antimicrobial alkaloids	Frick and Kutchan 1999
AF064696	Thalictrum tuberosum	364	o-Dihydric phenols, BIQ alkaloids (Multifunctional)	Antimicrobial alkaloids	Frick and Kutchan 1999
M73235	Zea mays	364	Caffeic acid	Lignin metabolism	Collazo et al., 1992
U19911	Zinnia elegans	354	5HF > caffeic acid	Lignin biosynthesis	Ye and Varner 1995
Carboxylic acid OMTs					
At5g55250	Arabidopsis thaliana	386	Indole-3 acetic acid	Plant growth hormone synthesis	Zubieta et al., 2003
AY008434	Arabidopsis thaliana	389	Jasmonic acid	Jasmonate-regulated plant responses	Seo et al., 2001
AF198492	Antirrhinum majus	364	Benzoic acid	Floral scent	Dudareva et al., 2000
AF133053	Clarkia breweri	359	Salicylic acid > benzoic acid	Floral scent	Ross et al., 1999
Flavonoid OMTs					
U70424	Arabidopsis thaliana	363	Quercetin > myricetin	Flavonoid metabolism	Muzac et al., 2000
AY127568	Catharanthus roseus (periwinkle)	348	Myricetin > dihydromyricetin	Flavonol 3'/5' methylations	Cacace et al., 2003

Accession	Species	No.	Substrate	Function	Reference
AY343489	Catharanthus roseus (periwinkle)	359	Homoeriodictyol >> Chrysoeriol >	Flavanones 4' methylation	Schröder et al., 2004
U16794	Chrysosplenium americanum	343	Partially methylated flavonols	Flavonoid biosynthesis	Gauthier et al., 1996;
AB091684	Glycyrrhiza echinata	367	2,7,4'-TriOH-isoflavanone	Phytoalexin biosynthesis	Akashi et al., 2003
AB091685	Glycyrrhiza echinata	357	Daidzein	Phytoalexin biosynthesis	Akashi et al., 2003
X77467	Hordeum vulgare	390	Apigenin > naringenin	Phytoalexin biosynthesis	Christensen et al., 1998
AB091686	Lotus japonicus	365	2,7,4'-TriOH-isoflavanone	Phytoalexin biosynthesis	Akashi et al., 2003
L10211	Medicago sativa	372	Isoliquiritigenin	Induction of Nod genes	Maxwell et al., 1993
U97125	Medicago sativa	352	6,7,4'-TriOH-isoflavone > daidzein >> pterocarpan	Phytoalexin biosynthesis	He et al., 1998
AY337457	Mentha piperita (peppermint)	344	Flavones > flavonols >> isoflavone	Flavonoid metabolism	Willits et al., 2004
AY337458	Mentha piperita (peppermint)	344	Flavones > flavonols	Flavonoid metabolism	Willits et al., 2004
AY337459	Mentha piperita (peppermint)	366	Various flavonols	Flavonoid metabolism	Willits et al., 2004
AY337460	Mentha piperita (peppermint)	364	Various flavonoids	Flavonoid metabolism	Willits et al., 2004

AY337461	*Mentha piperita* (peppermint)	343	Various flavonoids	Flavonoid metabolism	Willits *et al.*, 2004
U69554	*Pisum sativum*	360	6a-OH-maackiain > (+)maackiain > (+)medicarpin	Pterocarpan biosynthesis	Wu *et al.*, 1997
Alkaloid OMTs					
AB073908	*Coptis japonica*	351	Columbamine > (R,S) scoulerine	Isoquinoline alkaloid biosynthesis	Morishige *et al.*, 2002
D29809	*Coptis japonica*	364	(R,S) scoulerine	Tetrahydro-BIQ alkaloid biosynthesis	Takeshita *et al.*, 1995
D29811	*Coptis japonica*	347	(R,S) 6-OH norcoclaurine	BIQ alkaloid biosynthesis	Morishige *et al.*, 2000
D29812	*Coptis japonica*	350	Various BIQ alkaloids	BIQ alkaloid biosynthesis	Morishige *et al.*, 2000
AY268893	*Papaver somniferum*	355	Guaiacol > (R,S) reticuline 2.5 fold higher	Isoquinoline alkaloid biosynthesis	Ounaroon *et al.*, 2003
AY268894	*Papaver somniferum*	346	(R,S) norcoclaurine	Isoquinoline alkaloid biosynthesis	Ounaroon *et al.*, 2003

Abbreviations: CoA, coenzyme A; 5HF, 5 hydroxyferulic acid; BIQ, benzylisoquinoline.
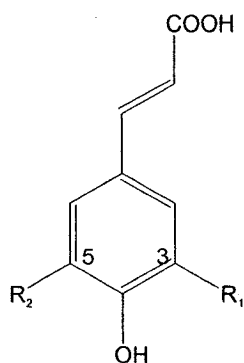
Hydroxycinnamic acids (HCA)
Caffeic acid: R_1 = OH, R_2 = H
5-Hydroxyferulic acid: R_1 = OMe,
R_2 = OH
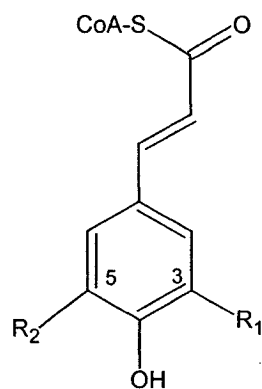
Hydroxycinnamoyl CoA-esters
Caffeoyl-CoA: R_1 = OH, R_2 = H
5-Hydroxyferuloyl-CoA: R_1 = OMe,
R_2 = OH
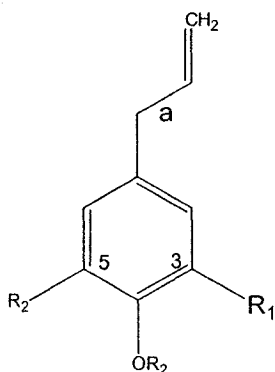
Phenylpropenes
Chavicol: R_1 = R_2 = H
Eugenol: R_1 = OMe, R_2 = H
Isoeugenol: R_1 = OMe, R_2 = H,
a = Double bond
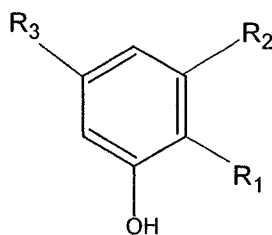
Dihydric phenols (DHP)
Catechol: R_1 =OH, R_2=R_3= H
Guajacol: R_1 = oMe, R_2 = R_3 = H
Orcinol: R_1 =H, R_2 =OH, R_3=OMe
Phloroglucinol: R_1 = R_2 = R_3 = OH
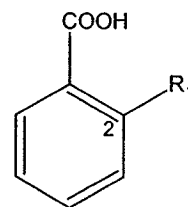
Carboxylic acids
Benzoic acid: R_1 = H
Salicylic acid: R_1 = OH
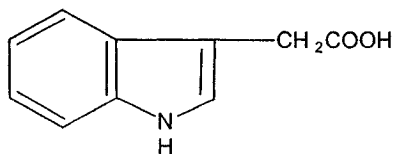
Indole-3-acetic acid
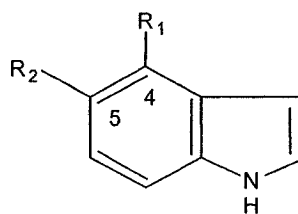
4-hydroxIndole: R_1 = OH, R_2 = H
5-hydroxIndole: R_1 = H, R_2 = OH
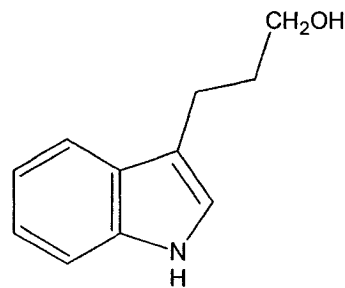
Indole 3-ethanol

Figure 21 A: Various secondary metabolites utilized by OMTs.

94
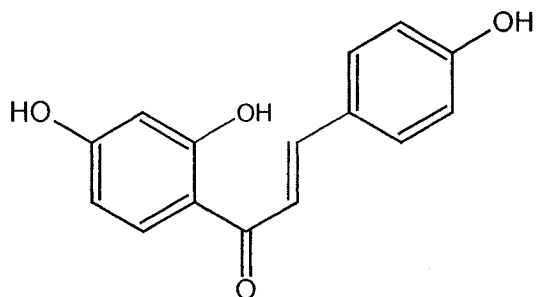
Chalcone
Isoliquiritigenin
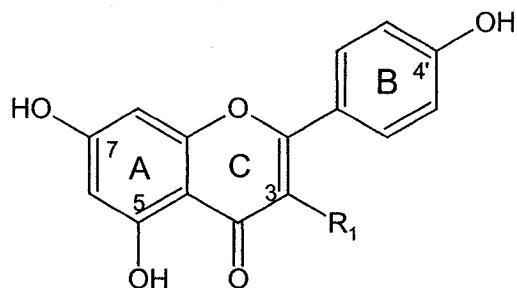
Flavonoids
Flavone: R₁ = H, Flavonol: R₁ = OH
Flavanone: no double bond between C2 and C3
Flavonol: R₁ = OH; no double bond between C2 and C3
Quercetin: R₁ = OH, 5' = OH
Apigenin: R₁ = H
Luteolin: R₁ = H, 5' = OH
Naringenin: no double bond between C2 and C3, R₁ = H
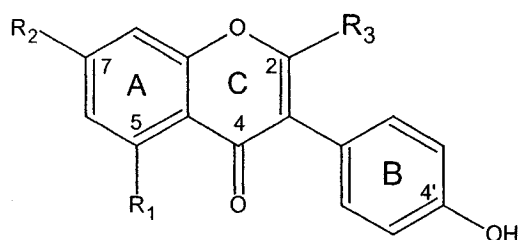
Isoflavonoids
Genistein: R1, R3 = H, R2 = OH
Daidzein: R1, R2 = OH, R3 = H
2, 7, 4'-triOH-isoflavanone: R1 = H, R2 = R3 = OH (no double bond between C2 and C3)
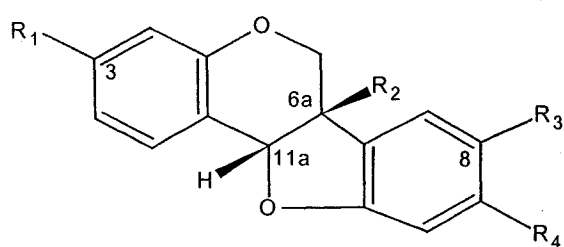
Pterocarpan
(+)6a-hydroxymaackiain: R₁ = OH, R₂ = OH, R₃ = R₄ = OCH₂O
Medicarpin: R₁ = OH, R₂ = R₃ = H, R₄ = OMe
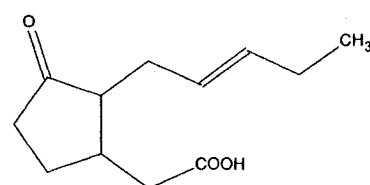
Jasmonic acid

Figure 21 B: Various secondary metabolites utilized by OMTs.

95
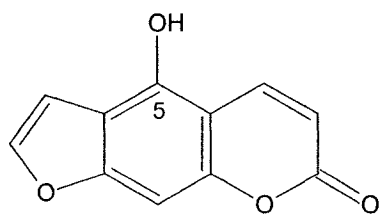
Furanocoumarin
Bergaptol
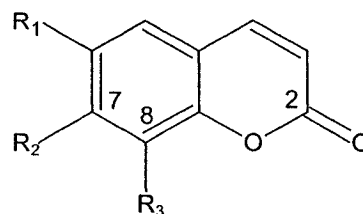
Coumarins
Daphnetin: R_1 = H, R_2 = R_3 =OH
Esculetin: R_1 = R_2 = OH, R_3 = H
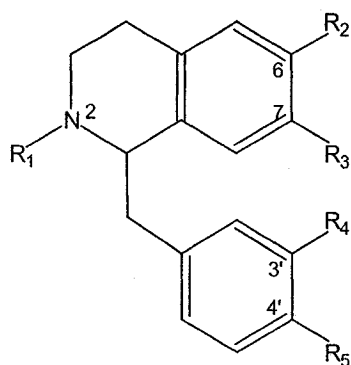
Benzylisoquinoline alkaloids (BIQA)
(R,S) norcoclaurine: R_1 = R_4 = H,
R_2 = R_3 = R_5 = OH,
(R,S) recticuline: R_1 = Me,
R_2 = R_5 = OMe, R_3 = R_4 = OH
Coclaurine : R_1 = R_4 = H, R_2 = OMe,
R_3 = R_5 = OH, R_4 = H
Laudanosoline: R1 = Me,
R2 = R3 = R4 = R5 = OH

Tetrahydrobenzylisoquinoline alkaloids (TBIQA)
Scoulerine: R_1 = OH, R_2 = OMe
Columbamine: R_1 = R_2 = OMe

Myo-inositol

Figure 21 C: Various secondary metabolites utilized by OMTs.

96

A

Vitis (CCoA) Z54233
Populus (CCoA) AJ224894
Alfalfa (CCoA) U20736
Tobacco (CCoA) U38612
Zinnia (CCoA) U13151
Pinus (CCoA) AF036095
Parsley (CCoA) M69184
Chickweed (CCoA) L22203
Ice plant (CCoA) AY145521
Arabidopsis (Jasmonic) AY008434
Clarkia (Salicylic) AF133053
Antirrhinum (Benzoic) AF198492
Arabidopsis (Indole acetic) At5g55250

B

Alfalfa (HCA) M63853
Populus (HCA) X62096
Liquidambar (HCA) AF139533
Chrysosplenium (3'/5'-Flavonol) U16794
Chrysosplenium (HCA) U16793
Arabidopsis (3'-Flavonol) U70424
Clarkia (Isoeugenol) U86760
Lolium (HCA) AF010291
Zea (HCA) M73235
Zinnia (HCA) U19911
Ammi (Furanocoumarin) AY443006
Peppermint (3'-Flavonoid) AY337460
Periwinkle (HCA) AY028439
Ammi (HCA) AY443007
Thalictrum (DHP/HCA/BIQA) AF064696
Thalictrum (DHP/HCA/BIQA) AF064693
Thalictrum (DHP/HCA) AF064695
Thalictrum (DHP/HCA) AF064694
Ice plant (m-Inositol) M87340
Rose (Phloroglucinol) AB121046
Tobacco (DHP) X71430
Coptis (9-Scoulerine) D29809
Alfalfa (2'-Chalcone) L10211
Barley (HCA) U54767
Pinus (CCoA/HCA) U39301
Papaver (7-Reticuline) AY268893
Barley (7-Flavonoid) X77467
Rye (8-Coumarin) AY177404
Periwinkle (3'&5'-Flavonol) AY127568
Periwinkle (4'-Flavonol) AY343489
Ocimum (Chavicol) AF435007
Ocimum (Eugenol) AF435008
Rose (Orcinol) AF502433
Rose (Methyl Orcinol) AF502434
Peppermint (8-Flavone) AY337459
Pisum (3-Pterocarpan) U69554
Glycyrrhiza (4'-Isoflavanone) AB091684
Lotus (4'-Isoflavanone) AB091686
Alfalfa (7-Isoflavone) U97125
Glycyrrhiza (7-Isoflavone) AB091685
Coptis (6-Norcoclaurine) D29811
Papaver (6-Norcoclaurine) AY268894
Coptis (4'-Methylcoclaurine) D29812
Peppermint (7-Flavonoid) AY337457
Peppermint (7-Flavonoid) AY337458
Peppermint (4'-Flavonoid) AY337461
Coptis (Columbamine) AB073908
Bos A42106
Human U11098
Streptomyces U10405

Figure 22: Strict consensus of two parsimonious trees of 60 OMTs. Non-plant OMTs were used as out-groups to root the tree. Branch-lengths and bootstrap values are represented by numbers above (in parenthesis) and below (in bold) each branch, respectively.

97

The majority of the OMTs within the B-lineage constitute a number of distinct subclades, which include 26 clones, mostly encoding the methylation of *o*-diphenol (catechol) and COMTs, with the exception of one chalcone, one furanocoumarin, one inositol, three phenylpropene (chavicol, eugenol, isoeugenol), two alkaloid, and three flavonoid OMTs. Other distinct subclades include 21 OMTs (with bootstrap values in brackets), catalyzing the methylation of five isoflavonoid/pterocarpan OMTs (58%), three *Coptis-Papaver* alkaloid OMTs (86%), three peppermint flavonoid OMTs (96%), two *Ocimum* phenylpropene OMTs (100%), two rose OMTs (100%), and the barley-rye and periwinkle OMTs (78% and 99%, respectively).

G.3.2 Neighbor-Joining Analysis

The same 60 OMT sequences were used for distance based neighbor-joining analysis and resulted in a single tree (Figure 23). The topology of the distance tree was similar to that of maximum parsimony tree, with few exceptions. As in the case of maximum parsimony tree, two distinct monophyletic lineages, namely A and B, were observed. The bootstrap supports for these lineages were stronger with bootstrap values of 100% and 99% for lineages A and B, respectively. Lineage A is identical to that of the maximum parsimony tree (Figure 22).

Likewise, the topology of lineage B resembles the corresponding lineage in the maximum parsimony tree, with some differences that involve the separation and/or grouping of some taxa in each subclade within the larger clade(s). For example the subclades of *Pinus* (U39301) and *Papaver* (AY268893), of *Lolium* (AF010291) and *Zea*

Vitis (CCoA) Z54233
Populus (CCoA) AJ224894
Tobacco (CCoA) U38612
Alfalfa (CCoA) U20736
Parsley (CCoA) M69184
Zinnia (CCoA) U13151
Pinus (CCoA) AF036095
Chickweed (CCoA) L22203
Ice plant (CCoA) AY145521
Arabidopsis (Jasmonic) AY008434
Clarkia (Salicylic) AF133053
Antirrhinum (Benzoic) AF198492
Arabidopsis (Indole acetic) At5g55250

A

Alfalfa (HCA) M63853
Populus (HCA) X62096
Liquidambar (HCA) AF139533
Arabidopsis (3'-Flavonol) U70424
Chrysosplenium (3'/5'-Flavonol) U16794
Chrysosplenium (HCA) U16793
Periwinkle (HCA) AY028439
Ammi (HCA) AY443007
Zinnia (HCA) U19911
Clarkia (Isoeugenol) U86760
Thalictrum (DHP/HCA/BIQA) AF064696
Thalictrum (DHP/HCA/BIQA) AF064693
Thalictrum (DHP/HCA) AF064695
Thalictrum (DHP/HCA) AF064694
Lolium (HCA) AF010291
Zea (HCA) M73235
Ammi (Furanocoumarin) AY443006
Peppermint (3'-Flavonoid) AY337460
Tobacco (DHP) X71430
Ice plant (m-Inositol) M87340
Rose (Phloroglucinol) AB121046
Alfalfa (2'-Chalcone) L10211
Coptis (9-Scoulerine) D29809
Barley (HCA) U54767
Barley (7-Flavonoid) X77467
Rye (8-Coumarin) AY177404
Pisum (3-Pterocarpan) U69554
Glycyrrhiza (4'-Isoflavanone) AB091684
Lotus (4'-Isoflavanone) AB091686
Alfalfa (7-Isoflavone) U97125
Glycyrrhiza (7-Isoflavone) AB091685
Ocimum (Chavicol) AF435007
Ocimum (Eugenol) AF435008
Rose (Orcinol) AF502433
Rose (Methyl Orcinol) AF502434
Peppermint (8-Flavone) AY337459
Periwinkle (3'&5'-Flavonol) AY127568
Periwinkle (4'-Flavonol) AY343489
Coptis (6-Norcoclaurine) D29811
Papaver (6-Norcoclaurine) AY268894
Coptis (4'-Methylcoclaurine) D29812
Coptis (Columbamine) AB073908
Pinus (CCoA/HCA) U39301
Papaver (7-Reticuline) AY268893
Peppermint (7-Flavonoid) AY337457
Peppermint (7-Flavonoid) AY337458
Peppermint (4'-Flavonoid) AY337461

B

Human U11098
Bos A42106
Streptomyces U10405

Figure 23 A: Neighbor-joining tree of 60 OMTs. Non-plant species were used as out-groups to root the tree. Numbers below branches represent bootstrap values.

99

Z54233
AJ224894
U38612
U20736
M69184
U13151
AF036095
L22203
AY145521
AY008434
AF133053
AF198492
At5g55250
M63853
X62096
AF139533
U70424
U16794
U16793
AY028439
AY443007
U19911
U86760
AF064696
AF064693
AF064695
AF064694
AF010291
M73235
AY443006
AY337460
X71430
M87340
AB121046
L10211
D29809
U54767
X77467
AY177404
U69554
AB091684
AB091686
U97125
AB091685
AF435007
AF435008
AF502433
AF502434
AY337459
AY127568
AY343489
D29811
AY268894
D29812
AB073908
U39301
AY268893
AY337457
AY337458
AY337461
Bos
Human
Streptomyc

0.1 changes

Figure 23 B: Neighbor-joining tree of 60 OMTs showing branch lengths. Non-plant OMTs were used as out-group to root the tree. Refer to Table 3 for identification of clones.

(M73235), and of the two periwinkle flavonol OMTs (AY127568 and AY343489) occupy a more basal position in the neighbor-joining tree than the maximum parsimony trees. *Zinnia* OMT (U19911), however, has separated from the clade with periwinkle (AY028439) and *Ammi* (AY443007) occupying a less basal position in the neighbor-joining tree as compared to the maximum parsimony tree. The neighbor-joining tree, nevertheless, resembles the maximum parsimony tree as the major distinct subclades are very similar to the maximum parsimony tree. These subclades include (with bootstrap values in brackets) the five isoflavonoid/pterocarpan OMTs (99%), four *Thalictrum* multifunctional OMTs (AF064693 to AF064696) (bootstrap value 100%), three alkaloids OMTs (D29811, D29812, AY268894) (100%), three peppermints OMTs (AY337457 to AY337461) (100%), two *Ocimum* phenylpropene OMTs (AF435007, AF435008) (100%), rose OMTs (AF502433, AF502434) (100%), and the barley-rye OMTs (X77467,AY177404) (99%).

Both the parsimony (Figure 22) and the neighbor-joining trees (Figure 23) revealed two monophyletic lineages (A and B); with lineage A consisting of CCoAOMTs and carboxylic OMTs, and lineage B comprising mainly the phenylpropanoid COMTs, (iso)flavonoids and alkaloids. It is interesting to note that within lineage A, members of the carboxylic OMTs, which methylate carboxyl groups of various acids are closely related to those of CCoAOMTs, which methylate phenolic hydroxyl groups of various substrates. This difference in substrate preference of the carboxylic OMTs could be attributed to the significant changes in amino acid residues (274 branch length value) from their common ancestor to that of the CCoAOMTs. Furthermore, it is not unexpected to find that the chickweed and ice plant OMTs group together, forming a

101

subclade basal to the rest of the CCoAOMTs. These two CCoAOMTs represents two closely related enzymes, as the ice plant OMT gene was originally cloned on the basis of sequence homology to chickweed OMT. Their location in this lineage may be further corroborated by the fact that both OMTs utilize Caffeoyl-CoA and caffeoyl glucose esters as their preferred substrates, although the ice plant OMT is a multifunctional enzyme that catalyzes the methylation of flavonols as well (Ibdah *et al.*, 2003).

The majority of members within lineage B constitute a number of distinct subclades, with 99% bootstrap support (Figure 23) and 139 amino acids changes (Figure 22) from its nearest ancestor. These include most of the COMTs, OMTs which methylate ring B of flavonoids that are related to phenylpropanoids, including the four *Thalictrum* multifunctional OMTs, as well as other OMTs from the side branch of phenylpropanoids. Other distinct subclades include the isoflavonoid/pterocarpan OMTs (99% Bootstrap support; Figure 23), the three peppermint flavonoid OMTs (96% Bootstrap value; Figure 22), and the barley-rye OMTs (99% Bootstrap value; Figure 23). Members of the latter subclade catalyze the 7-*O*-methylation of flavones and position 8 of daphnetin, a member of the coumarin family of compounds. These two biosynthetically related substrates are induced by pathogen attack and cold acclimation, respectively. The five isoflavone-pterocarpan OMTs represent an interesting example of being clustered in a distinct clade with all the enzymes sharing similar regiospecificities in their methyl acceptor molecules. In fact, position 3 of pterocarpans (*Pisum* OMT) is stereochemically similar to position 7 of isoflavones (Alfalfa and *Glycyrrhiza* OMTs), and resembles position 4' of *Glycyrrhiza* and *Lotus* isoflavones. Since both 4' and 7 positions are *para* to the heterocyclic ring,

their hydroxyl groups exhibit similar negative electron densities with similar methyl acceptor properties.

The six benzylisoquinoline (BIQ) alkaloid-specific OMTs from *Coptis* and *Papaver* are located in two different subclades of the B lineage (Figure 22 and 23). One of these, the *Coptis* scoulerine 9-OMT (92% Bootstrap value; Figure 23) is located between the alfalfa chalcone and barley COMT; its segregation from the other BIQ alkaloid OMTs may be ascribed to the fact that its methyl acceptor, scoulerine is a tetrahydro derivative and that methylation takes place at position 9 of the ring, thus resembling phenylpropanoids (Figure 21A and C). However, *Coptis* columbamine (the methylated product of scoulerine; 9-methyl-scoulerine) 3-OMT, which methylates position 3, is grouped together with three of the norcoclaurine OMTs (Figure 23). These three OMTs, the *Coptis/Papaver* norcoclaurine 6-OMTs and the 3' -hydroxy-*N*-methylcoclaurine-4' -OMT occupy a distinct clade, and correspond well with their strict substrate- and position specificities. In contrast, the sixth member of the alkaloid OMT family, *Papaver* reticuline 7-OMT is associated with a *Pinus* COMT in an unusual clade that reflects their distinct substrate preferences (Figure 22 and 23). This alkaloid OMT is a permissive enzyme that methylates a number of tetrahydro-BIQ alkaloids and the simple phenol guaiacol for which it exhibits a higher preference. In contrast, the pine enzyme is a novel, bifunctional COMT that utilizes both caffeic and 5-hydroxyferulic acids as well as their CoA esters as substrates.

Similar to alkaloid OMTs, the four, position-specific peppermint OMTs are located in two distinct subclades where the 8-OMT is associated with the rose/*Ocimum* OMTs (Figure 22 and 23), whereas the 7- and 4'-OMT are clustered within a distinct

subclade. It could be hypothesized that segregation of these OMTs is correlated with the methylation of *meta* or *para* hydroxyl groups, and warrants further investigation. It is interesting to note that, among the phenylpropene OMTs, the *Clarkia*-isoeugenol OMT is associated with the flavonol-COMT while the other phenylpropene (*Ocimum* chavicol and eugenol) OMTs are located in another subclade (Figure 22 and 23). Likewise, the three rose dihydric phenol OMTs are segregated, with the phloroglucinol grouped with the ice plant myo-inositol, while the other two OMTs (orcinol and methyl orciol) clustered closely with the *Ocimum* chavicol and eugenol (Figure 22 and 23).

G.3.3   Prediction of Putative Substrates for OMT-Like Clones Through Phylogenomics

The sequences of two *Arabidopsis* OMT-like clones (AF344315 and AF344316) were included into the same matrix of the 60 biochemically characterized OMT sequences for analysis. Analyses based on maximum parsimony searches gave rise to eight equally parsimonious trees with a tree length of 6536, consistency index (CI) of 0.537, retention index (RI) of 0.649, and a rescaled consistency index (RC) of 0.349. The strict consensus of these eight trees (Figure 24) reveals that the 62 OMTs form several monophyletic lineages including the two distinct clades previously described in the maximum parsimony and neighbour-joining trees of the 60 biochemically characterized OMTs. The overall topology of the tree in Figure 24 is comparable to Figures 22 and 23. The two *Arabidopsis* putative OMT clones cluster within lineage B, in a subclade that includes the ice plant myo-inositol (M87340) and rose phloroglucinol (AB121046) OMTs.

Vitis (CCoA) Z54233
Populus (CCoA) AJ224894
Alfalfa (CCoA) U20736
Tobacco (CCoA) U38612
Parsley (CCoA) M69184
Zinnia (CCoA) U13151
Pinus (CCoA) AF036095
Chickweed (CCoA) L22203
Ice plant (CCoA) AY145521
Arabidopsis (Jasmonic) AY008434
Clarkia (Salicylic) AF133053
Antirrhinum (Benzoic) AF198492
Arabidopsis (Indole acetic) At5g55250
Alfalfa (HCA) M63853
Populus (HCA) X62096
Liquidambar (HCA) AF139533
Chrysosplenium (3'/5'-Flavonol) U16794
Chrysosplenium (HCA) U16793
Arabidopsis (3'-Flavonol) U70424
Clarkia (Isoeugenol) U86760
Lolium (HCA) AF010291
Zea (HCA) M73235
Zinnia (HCA) U19911
Ammi (Furanocoumarin) AY443006
Peppermint (3'-Flavonoid) AY337460
Periwinkle (HCA) AY028439
Ammi (HCA) AY443007
Thalictrum (DHP/HCA/BIQA) AF064696
Thalictrum (DHP/HCA/BIQA) AF064693
Thalictrum (DHP/HCA) AF064695
Thalictrum (DHP/HCA) AF064694
Tobacco (DHP) X71430
Arabidopsis **AF344315**
Arabidopsis **AF344316**
Ice plant (m-Inositol) M87340
Rose (Phloroglucinol) AB121046
Alfalfa (2'-Chalcone) L10211
Coptis (9-Scoulerine) D29809
Barley (HCA) U54767
Barley (7-Flavonoid) X77467
Rye (8-Coumarin) AY177404
Pisum (3-Pterocarpan) U69554
Glycyrrhiza (4'-Isoflavanone) AB091684
Lotus (4'-Isoflavanone) AB091686
Alfalfa (7-Isoflavone) U97125
Glycyrrhiza (7-Isoflavone) AB091685
Ocimum (Chavicol) AF435007
Ocimum (Eugenol) AF435008
Rose (Orcinol) AF502433
Rose (Methyl Orcinol)AF502434
Peppermint (8-Flavone) AY337459
Periwinkle (3'&5'-Flavonol) AY127568
Periwinkle (4'-Flavonol) AY343489
Coptis (6-Norcoclaurine) D29811
Papaver (6-Norcoclaurine) AY268894
Coptis (4'-Methylcoclaurine) D29812
Coptis (Columbamine) AB073908
Pinus (CCoA/HCA) U39301
Papaver (7-Reticuline) AY268893
Peppermint (7-Flavonoid) AY337457
Peppermint (7-Flavonoid) AY337458
Peppermint (4'-Flavonoid) AY337461
Human U11098
Bos A42106
Streptomyces U10405

98 100 76 51 62 100 100 100 100 98 100 82 100 78 82 100 58 100 100 100 52 100 100 91 86 87 100 97

Figure 24: Strict consensus of eight parsimonious trees of the 60 biochemically characterized OMTs together with the *Arabidopsis* clones AF344315 and AF344316. Numbers below branches represent bootstrap values.

105

The sequences of three putative OMT clones of *S. tinctoria* were analyzed in the same way with the 60 biochemically characterized OMT clones utilizing the neighbor-joining method. Likewise, Figure 25 reveals a tree with an overall topology similar to the three trees previously reconstructed. The three *S. tinctoria* OMT-like clones, StOMT1, StOMT2 and StOMT3, grouped closely with a subclade (92% bootstrap support) in the B lineage that includes *Zinnia* COMT (U19911), *Ammi majus* COMT (AY443007), and a hydroxycinnamic acid and flavanoids utilizing OMT of *Periwinkle* (AY028439). Though phylogenetic analyses suggest that they utilize hydroxycinnamic acids as preferred substrates, these clones await further molecular and biochemical characterization.

Vitis (CCoA) Z54233

Populus (CCoA) AJ224894

Tobacco (CCoA) U38612
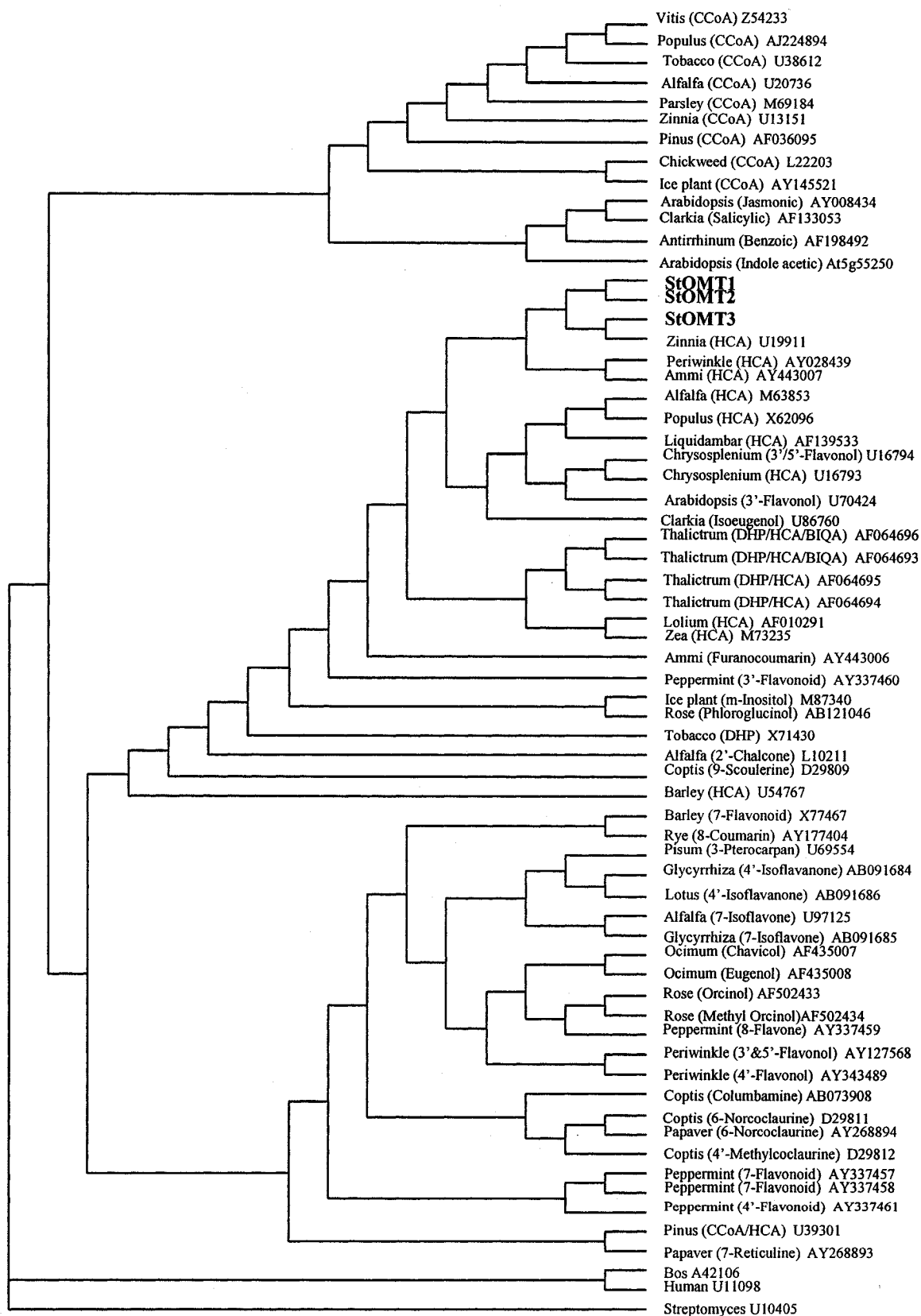
Alfalfa (CCoA) U20736

Parsley (CCoA) M69184

Zinnia (CCoA) U13151

Pinus (CCoA) AF036095

Chickweed (CCoA) L22203

Ice plant (CCoA) AY145521

Arabidopsis (Jasmonic) AY008434

Clarkia (Salicylic) AF133053

Antirrhinum (Benzoic) AF198492

Arabidopsis (Indole acetic) At5g55250

StOMT1

StOMT2

StOMT3

Zinnia (HCA) U19911

Periwinkle (HCA) AY028439

Ammi (HCA) AY443007

Alfalfa (HCA) M63853

Populus (HCA) X62096

Liquidambar (HCA) AF139533

Chrysosplenium (3'/5'-Flavonol) U16794

Chrysosplenium (HCA) U16793

Arabidopsis (3'-Flavonol) U70424

Clarkia (Isoeugenol) U86760

Thalictrum (DHP/HCA/BIQA) AF064696

Thalictrum (DHP/HCA/BIQA) AF064693

Thalictrum (DHP/HCA) AF064695

Thalictrum (DHP/HCA) AF064694

Lolium (HCA) AF010291

Zea (HCA) M73235

Ammi (Furanocoumarin) AY443006

Peppermint (3'-Flavonoid) AY337460

Ice plant (m-Inositol) M87340

Rose (Phloroglucinol) AB121046

Tobacco (DHP) X71430

Alfalfa (2'-Chalcone) L10211

Coptis (9-Scoulerine) D29809

Barley (HCA) U54767

Barley (7-Flavonoid) X77467

Rye (8-Coumarin) AY177404

Pisum (3-Pterocarpan) U69554

Glycyrrhiza (4'-Isoflavanone) AB091684

Lotus (4'-Isoflavanone) AB091686

Alfalfa (7-Isoflavone) U97125

Glycyrrhiza (7-Isoflavone) AB091685

Ocimum (Chavicol) AF435007

Ocimum (Eugenol) AF435008

Rose (Orcinol) AF502433

Rose (Methyl Orcinol)AF502434

Peppermint (8-Flavone) AY337459

Periwinkle (3'&5'-Flavonol) AY127568

Periwinkle (4'-Flavonol) AY343489

Coptis (Columbamine) AB073908

Coptis (6-Norcoclaurine) D29811

Papaver (6-Norcoclaurine) AY268894

Coptis (4'-Methylcoclaurine) D29812

Peppermint (7-Flavonoid) AY337457

Peppermint (7-Flavonoid) AY337458

Peppermint (4'-Flavonoid) AY337461

Pinus (CCoA/HCA) U39301

Papaver (7-Reticuline) AY268893

Bos A42106

Human U11098

Streptomyces U10405

Figure 25: Neighbor-joining tree the 60 biochemically characterized OMTs together with the _S. tinctoria_ OMT-like clones StOMT1, StOMT2 and StOMT3.

107

H. DISCUSSION

H.1   Molecular and Biochemical Characterization of *Arabidopsis* OMTs

The first plant methyltransferase for which a cDNA was isolated was caffeoyl-CoA 3-*O*-methyltransferase, an enzyme involved in the disease resistance response in parsley cell cultures (Schmitt *et al.*, 1991). Since then, the genes encoding various plant OMTs from diverse plant species have been identified (reviewed in Ibrahim *et al.*, 1998). Comparison of the amino acid sequences derived from the translated nucleotide sequences revealed up to five conserved regions near the C-termini of the polypeptides (Ibrahim, 1997). These consensus sequences may aid in the identification and direct cloning of methyltransferase-encoding these plant cDNAs.

The international effort to sequence the complete genome of *Arabidopsis* has created a remarkable opportunity to study the evolution and molecular mechanisms leading to the diversity of the OMT superfamily, as well as provided an expeditious way for the characterization and cloning of putative OMT cDNAs in *Arabidopsis*. This will enable the elucidation of the role of OMTs in various biosynthetic pathways, an important prerequisite for subsequenct genetic engineering efforts, as previously discussed.

In our attempt to identify putative OMTs in the *Arabidopsis* database (TAIR) using the flavonol 3'/5'-OMT (Gauthier *et al.*, 1996) as the query sequence, we came across two putative full-length OMT cDNA clones. Upon initial sequence analysis, these two clones, termed C00113 and C00114 (Accession # AF344315 and AF344316, respectively), showed characteristics of Group II OMTs; i.e. the five signature motifs,

molecular weight of ca. 40 kDa, and a comparable sequence identity to other biochemically characterized OMTs from various plant species. Therefore, we proceeded to heterologously express the two clones in a prokaryotic expression system to further characterize the encoded proteins. Upon expression in *E. coli*, the two recombinant proteins did not accept any of the substrates tested, regardless of the conditions being optimized for enzyme assays, as well as the purity of the proteins being used. The substrates used in these studies belonged to a wide range of compounds from different families within the phenylpropanoid and flavonoid groups of secondary metabolites (Table 2). It is unknown whether the appropriate substrates have yet to be identified or that the recombinant proteins are catalytically inactive.

In order to investigate the possibility that the expressed recombinant proteins in a prokaryotic expression system may not have been catalytically active due to improper folding and lack of post-translational modifications, we proceeded to express the C00113 clone in the eukaryotic system, *P. pastoris*. However, this recombinant protein failed to show any activity with any of the substrates tested, regardless of the level of purity of the proteins used or the optimized conditions for enzyme assays.

It may be possible that C00113 and C00114 encode OMTs that are not involved in the phenylpropanoid and flavonoid pathways, or are highly specific for substrates that are yet to be identified. However, even if the putative substrate was not used, an active OMT may still be expected to exhibit some residual background activity when utilizing structurally similar compounds to its preferred substrate, as has been the case with many OMTs (Frick and Kutchan, 1999; Huang *et al.*, 2004). The fact that no such activity was detected with either of the recombinant proteins, despite their expression in both

prokaryotic and eukaryotic systems, suggests that the two proteins may, in fact, be catalytically inactive. Interestingly, both Cacace *et al.* (2003) and Willits *et al.* (2004), among other researchers, while attempting to isolate OMT cDNA clones also discovered that some putatively positive OMT cDNA clones, although showing strong sequence identity to functional OMTs, encode proteins that display no OMT activity regardless of the substrates tested.

Pseudogenes have traditionally been described as non-functional sequences of genomic DNA (junk DNA) derived from functional genes, but display such degenerative traits as frameshift mutations and premature stop codons that disable their expression or block transcription if such defects occur in the promoter region (Mighell *et al.*, 2000; Balakirev and Ayala, 2003). As a result, if translation takes place at all, pseudogenes often lead to the production of proteins that do not have the same functional properties as the proteins encoded by their "normal" paralogous counterparts. Since pseudogenes do not have any apparent functional role, it has been assumed that they are generally released from selective pressure and, consequently, pseudogene mutations are selectively neutral and have high probability of becoming fixed in the genome (Balakirev and Ayala, 2003).

However, recent studies have demonstrated that pseudogenes are, more often than not, evolutionarily conserved as well as being transcriptionally and translationally active (Balakirev and Ayala, 2003). There is substantial evidence indicating that pseudogenes often exhibit functional roles in gene expression, gene regulation, and generation of genetic diversity. For example, studies on killer-cell immunoglobulin-like receptors (KIR) revealed that an originally non-functional KIR pseudogene, KIR3DP1, is brought

back to life in certain individuals via non-reciprocal recombination between different KIR haplotypes (Uhrberg, 2005). It follows that pseudogenes also display reduced nucleotide variability, excess synonymous over non-synonymous nucleotide polymorphism, and other features that are expected in genes or DNA sequences that possess functional roles (Balakirev and Ayala, 2003 and references therein).

Given that both C00113 and C00114 were obtained as full-length cDNA clones, they are transcriptionally active genes. They possess comparable sequence identity to various functional OMTs and produce seemingly normal proteins, but have no apparent enzymatic activity. Could these two *Arabidopsis* putative OMTs be pseudogenes? One of the characteristics of pseudogenes is that they are thought to arise by tandem duplication of the functional genes from which they are derived. In turn, they are likely to be inserted adjacent to their paralogous functional gene, but translocation to a different chromosome is also possible (Mighell *et al.*, 2000). A closer examination of the genomic location of C00113 and C00114 revealed that they originated from chromosome 1, and are indeed in tandem repeats with other putative OMTs. In fact, there are five such OMT-like genes in tandem repeats in this region (T22I11) of the chromosome, with C00113 (gene name AT1G21100.1) and C00114 (AT1G21120.1) in the first and third position, respectively. However, no further information was available from TAIR (The *Arabidopsis* Information Resource) to predict their substrate preferences.

In light of the possibility that C00113 and C00114 could be pseudogenes, it can be envisaged that these two genes might be non-functional due to nucleotide mutations in the protein-coding parts of their respective sequences, whereas their control elements in the promoter regions are maintained, thus producing pseudogene transcripts. These

transcripts could potentially serve a regulatory function such as anti-sense RNA that hybridizes with the sense RNA from homologous functional OMT gene(s) forming stable RNA/RNA duplex, thereby inhibiting their expression. Moreover, other means of repression could include competitive interaction of OMT pseudogenes with their homologous functional OMT gene(s) for positively acting transcription factor(s), thus affecting transcription. OMT pseudogenes could also represent a repertoire of sequences available for functional evolution with the potential of becoming genes whose products have acquired a new substrate preference(s), as pseudogenes are known to acquire functional roles (Uhrberg, 2005). This could, perhaps, be exemplified by the (+)6a-hydroxymaackiain (a pterocarpan) 3-OMT of *Pisum sativum* (Wu *et al.*, 1997), whereby out of the eight OMTs that were originally isolated, only two showed considerably higher enzyme activity than the others. The six OMTs that exhibited lower activity could have been pseudogenes that have salvaged 3-OMT activity with lower substrate specificity or regiospecificity, but might show higher activity with structurally similar substrates. Whether C00113 and C00114 are pseudogenes thus producing inactive proteins remains to be further investigated.

Arabidopsis thaliana represents an experimental model system for many plant species. Though the characterization of C00113 and C00114 is as yet to be determined, the future cloning and characterization of other functional OMTs in *Arabidopsis* would be invaluable in the elucidation of certain metabolic pathways, as well as enable the extension of this knowledge to other species, thus making feasible the application of genetic engineering. This is especially true if enhanced bioinformatics tools, such as software programs capable of annotating genes or detecting pseudogenes, could be

integrated into public databases. This would certainly improve the efficacy of gene searches and increase the possibility of identifying functional genes.

H.2   Screening of *Serratula tinctoria* cDNA Library

3-*O*-Methylquercetin is known for its roles as an anti-inflammatory and antiviral agent (Middleton and Kandaswami, 1994; Malhotra *et al.*, 1996), a specific inhibitor of viral RNA replication (Castrillo and Carrasco, 1987), and as a tobacco phytoanticipin that accumulates in response to wounding and insect herbivory (Roda *et al.*, 2003). Thus, the characterization of the OMT involved in its biosynthesis (Huang *et al.*, 2004) is of great economic interest with the ultimate goal of cloning and expressing the gene in transgenic plants.

O-Methylation at position 3 constitutes the first committed step in the multiple methylation sequence as has been demonstrated in a number of stepwise *O*-methylations of quercetin by distinct OMTs in various plant species, such as apple (*Malus domestica*) cell cultures (Macheix and Ibrahim, 1984), spinach (*Spinacia oleracea*) leaves (Thresh and Ibrahim, 1985), and *C. americanum* shoots (Ibrahim *et al.*, 1987). Thus, 3-*O*-methyl flavonols rarely accumulate in significant amounts in plants, since they serve as intermediates in the biosynthetic pathway of partially/highly methylated flavonoids (Ibrahim *et al.*, 1987; Macheix and Ibrahim, 1984; Thresh and Ibrahim, 1985). *S. tinctoria*, however, accumulates mainly 3,3'-dimethylquercetin and small amounts of 3-methylquercetin as an intermediate, suggesting the existence in this plant of both quercetin 3-OMT and 3-methylquercetin 3'-OMT enzymes (Huang *et al.*, 2004).

Recently, the native quercetin 3-OMT of *S. tinctoria* was partially purified, biochemically characterized and internal tryptic peptide information made available (Huang *et al.*, 2004). As a next logical step, we have since attempted to clone the gene encoding the flavonol 3-OMT employing degenerate primers based on the internal peptide sequences. Library screening was carried out utilizing a PCR-based screening strategy and standard colony lift hybridization methods. Following numerous rounds of screening with different primers, various primer combinations and PCR conditions for the PCR-based screening, amplification products containing sequence matches to the expected peptides of *S. tinctoria* 3-OMT native protein could not be identified. Colony lift screening with the TMAC protocol using terminally-labelled degenerate oligonucleotide probes was likewise unsuccessful in isolating the flavonol 3-OMT.

Given that 3-methylquercetin is an intermediate metabolite that rarely accumulates in plants grown under normal conditions, it is not unexpected that the enzyme catalyzing its methylation is under tight regulation both at the mRNA and protein levels, and hence its transcript may not be significantly expressed or accumulate to a degree that permits its detection. Therefore, it appears that the low expression of flavonol 3-OMT cDNA under normal conditions and/or the potential degradation of some unstable mRNA during extraction, could result in a lower level of transcript available for cDNA library construction and, as a result, leads to a lower copy number of flavonol 3-OMT clones in the overall pool. Perhaps, construction of a cDNA library derived from biotic- or abiotic-stressed plant tissues would improve the recovery of flavonol 3-OMT cDNA clones.

It has also been proposed that the 3-OMT in *C. americanum* is part of a multi-enzyme aggregate together with other OMTs catalyzing the stepwise *O*-methyaltion of quercetin (Ibrahim *et al.*, 1987). If such aggregation system exists in *C. americanum*, it would facilitate the formation of metabolic chains of intermediates and allows for increased catalytic activity. In such an efficient system, it can be envisaged that low levels of 3-OMT enzyme would be required for catalytic activity; however higher enzyme levels would be required to target substrates that are diffuse within the cytosol. Accordingly, assuming that *S. tinctoria* also possesses such a multi-enzyme aggregate for 3-*O*-methylation and subsequent stepwise substitution reactions, it is conceivable that the flavonol 3-OMT is under tight regulation and thus not expressed in great quantity.

Although the isolation of a full-length cDNA clone containing all the six tryptic peptides of the native flavonol 3-OMT was unsuccessful, three putative OMT-like full-length clones (StOMT1, StOMT2 and StOMT3) were isolated using *Arabidopsis* 3'-OMT cDNA clone as a heterologous probe. These three full-length clones, although lacking four of the tryptic peptides of the native flavonol 3-OMT, possess 100% match with peptide 3 and partial match with peptide 4 of the native protein. Upon heterologous expression of StOMT3 in *E. coli*, the resulting recombinant crude protein showed catalytic activity with quercetin and 5-hydroxyferulic acid, but not with 3-methylquercetin or caffeic acid. Given that the recombinant StOMT3 protein utilizes quercetin and not 3-methyl quercetion is significant, since the only difference between quercetin and 3-methylquercetin is the methyl-group at the 3 position. Therefore, it is possible that StOMT3 could very well be the gene encoding the native flavonol 3-OMT. Alternatively, StOMT3 could encode a COMT that also utilizes flavonol. Obviously,

more work lies ahead and immediate work includes the confirmation of the position of methylation and kinetic studies of the recombinant protein. Further work is currently in progress for the identification of these putative clones.

H.3   Phylogenomic Analysis of OMTs

The reconstruction of the maximum parsimony and neighbour-joining phylogenetic trees, based on protein sequence data of 60 biochemically characterized OMTs involved in the O-methylation of phenylpropanoids, flavonoids and alkaloids, allowed the development of a framework for predicting the putative function and substrate specificity of newly sequenced putative genes, and to infer the evolutionary history of OMTs in plant species.

The extraordinary diversity of plant secondary metabolites has been exploited by several researchers in chemosystematics and phylogenetics (Wink and Waterman, 1999 and references therein). In contrast, the paucity of reports on the biochemical and molecular characterization of the enzymes involved in secondary metabolite synthesis from different species prevents their extensive use in phylogenetic studies. However, in spite of the relatively small number of biochemically characterized OMTs, our investigation reveals that they catalyze the methylation of a variety of chemical structures, almost all of which are phenolic in nature. Functionally, these OMTs fall into two main categories; one involved in lignin biosynthesis and the other, in the biosynthesis of defence compounds, such as alkaloids and flavonoids.

Both phylogenetic trees obtained (Figures 22 and 23) demonstrate the segregation of the two OMT Groups into distinct, well supported lineages. The fact that Group I

OMTs are involved in the biosynthesis of the guaiacyl-type lignin that is considered primitive, since it occurs predominantly in pteridophytes and gymnosperms, makes it reasonable to assume that CCoAOMTs have appeared in parallel with the evolution of terrestrial vascular plants. They may have arisen by a selective, piece-by-piece assembly of pre-existing functional polypeptide subunits (Crow and Dove, 1989). Further fine-tuning of these assemblies, by selective amino acid substitutions have brought forth the functional diversity of the Group II OMTs with attenuated substrate- and regiospecificities to various methyl acceptor molecules (Ibrahim and Muzac, 2000).

The segregation of different Group II OMTs into distinct subclasses seems to follow the structure of their preferred substrates, which include carboxylic acids, various phenylpropanoids including phenylpropenes, (iso)flavonoids and alkaloids. The biosynthetic origin of these molecules may be considered as an important factor in the phylogeny of these OMTs, since except for the benzylisoquinoline alkaloids which are derived from L-tyrosine, all of the phenylpropanoid and flavonoid compounds are derived from L-phenylalanine. The two phylogenetic trees, nevertheless, could also be analyzed based not only on substrate utilization, but also on the physiological conditions of the taxa involved (i.e. stress induced or not), species relatedness, and other character states.

As shown in Figures 24 and 25, the phylogenetic trees can also be utilized as a powerful tool in substrate or functional predictions. The two *Arabidopsis* putative OMTs (AF344315 and AF344316) seem to cluster with ice plant *myo*-inositol and rose phloroglucinol-utilizing OMTs, suggesting that they could likewise utilize di-/trihydric phenols or simple sugar-like compounds such as inositol, as their preferred substrates (Figure 24). However, the fact that these substrates were not accepted further indicates

the possibility of the two clones being pseudogenes. In the case of StOMT1, StOMT2, and StOMT3, it is predicted that they utilize hydroxycinnamic acids as their preferred substrates, thus corroborating their activity with 5-hydroxyferulic acid and quercetin as substrates which are structurally related. In deed, this prediction has been substantiated with recombinant StOMT3 crude protein, which showed catalytic activity with quercetin and 5-hydroxyferulic acid.

These results demonstrate that the clustering of OMTs on the phylogenetic tree can be correlated with functional characteristics, and analyses of this nature allow for functional prediction of newly identified putative OMTs genes. This phylogenomic approach, therefore, is of great value in guiding future research endeavors. Although beyond the scope of this project, it would be valuable to compare members of the OMT family/superfamily for their 3-D structure (protein fold)-based phylogenetic trees to refine evolutionary relationships and explain the discrete differences among various OMTs.

I. REFERENCES

Akashi T, Sawada Y, Shimada N, Sakurai N, Aoki T, Ayabe S (2003) cDNA cloning and biochemical characterization of S-adenosyl-L-methionine: 2,7,4'-trihydroxyisoflavanone 4'-O-methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway. Plant Cell Physiol **44:** 103-112

Anzellotti D, Ibrahim RK (2004) Molecular characterization and functional expression of flavonol 6-hydroxylase. BMC Plant Biol **4:** 20

Atanassova R, Favet N, Martz F, Chabbert B, Tollier M, Monties B, Fritig B, Legrand M (1995) Altered lignin composition in transgenic tobacco expressing O-methyltransferase sequences in sense and antisense orientation. Plant Journal **8:** 465-477

Balakirev ES, Ayala FJ (2003) Pseudogenes: are they "junk" or functional DNA? Annu Rev Genet **37:** 123-151

Barriere Y, Argillier O, Chabbert B, Tollier MT, Monties B (1994) Breeding silage maize with brown-midrib genes. Feeding value and biochemical characterictics. Agronomie **14:** 15-25

Bohm BA (1998) Introduction to Flavonoids. Chemistry and Biochemistry of Organic Natural Products, Vol 2. Harwood Academic Publishers, Amsterdam

Boudet AM (1998) A new view of lignification. Trends Plant Science **3:** 67-71

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

Brierley RA, Davis GR, Holtz GC (1994) Production of insulin-like growth factor-1 in methylotrophic yeast cells. *In,* United States Patent 5,324,639

Brouillard R, Dangles O (1994) Flavonoids and flower colour. *In* JB Harborne, ed, The Flavonoids: Advances in Research Since 1986. Chapman and Hall, London, pp 565-586

Brunet G, Ibrahim R (1980) O-Methylation of flavonoids by cell-free extracts of Calamondin orange. Phytochemistry **19:** 741-746

Bugos RC, Chiang VL, Campbell WH (1991) cDNA cloning, sequence analysis and seasonal expression of lignin-bispecific caffeic acid/5-hydroxyferulic acid O-methyltransferase of aspen. Plant Mol Biol **17:** 1203-1215

Busam G, Junghanns KT, Kneusel RE, Kassemeyer HH, Matern U (1997) Characterization and expression of caffeoyl-coenzyme A 3-*O*-methyltransferase proposed for the induced resistance response of *Vitis vinifera* L. Plant Physiol **115**: 1039-1048

Cacace S, Schröder G, Wehinger E, Strack D, Schmidt J, Schröder J (2003) A flavonol *O*-methyltransferase from *Catharanthus roseus* performing two sequential methylations. Phytochemistry **62**: 127-137

Castrillo JL, Carrasco L (1987) Action of 3-methylquercetin on poliovirus RNA replication. J Virol **61**: 3319-3321

Chiron H, Drouet A, Claudot AC, Eckerskorn C, Trost M, Heller W, Ernst D, Sandermann H, Jr. (2000) Molecular cloning and functional expression of a stress-induced multifunctional *O*-methyltransferase with pinosylvin methyltransferase activity from Scots pine (*Pinus sylvestris* L.). Plant Mol Biol **44**: 733-745

Christensen AB, Gregersen PL, Olsen CE, Collinge DB (1998) A flavonoid 7-*O*-methyltransferase is expressed in barley leaves in response to pathogen attack. Plant Mol Biol **36**: 219-227

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci U S A **81**: 1991-1995

Clare JJ, Romanos MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA (1991) Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. Gene **105**: 205-212

Collazo P, Montoliu L, Puigdomenech P, Rigau J (1992) Structure and expression of the lignin *O*-methyltransferase gene from *Zea mays* L. Plant Mol Biol **20**: 857-867

Collins FW, De Luca V, Ibrahim R, Voirin B, Jay M (1981) Polymethylated flavonols in *Chrysosplenium americanum*: Identification and enzymatic synthesis. Z Naturforsch **36C**: 730-736

Croteau R, Kutchan T, Lewis NG (2000) Natural Products (Secondary Metabolites). *In* B Buchanan, W Gruissem, R Jones, eds, Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, MA pp 1250-1318

Crow JF, Dove WF (1989) Anecdotal, historical and critical commentaries on genetics. Genetics **122**: 1-6

Curir P, Lanzotti V, Dolci M, Dolci P, Pasini C, Tollin G (2003) Purification and properties of a new *S*-adenosyl-L-methionine:flavonoid 4'-*O*-methyltransferase from carnation (*Dianthus caryophyllus* L.). Eur J Biochem **270:** 3422-3431

Dakora FD, Phillips DA (1996) Diverse functions of isoflavonoids in legumes transcend anti-microbial definitions of phytoalexins. Physiol Mol Plant Pathol **49:** 1-20

De Carolis E, Ibrahim RK (1989) Purification and kinetics of phenylpropanoid *O*-methyltransferase activities from *Brassica oleracea*. Biochem Cell Biol **67:** 763-769

De Luca V, Ibrahim R (1985a) Enzymatic synthesis of polymethylated flavonols in *Chrysosplenium americanum*. I. Partial purification and some properties of *S*-adenosyl-L-methionine: flavonol 3-, 6-, 7-, and 4'-*O*-methyltransferases. Arch Biochem Biophys **238:** 596-605

De Luca V, Ibrahim R (1985b) Enzymatic synthesis of polymethylated flavonols in *Chrysosplenium americanum*. II. Substrate interaction and product inhibition studies of flavonol 3-, 6-, and 4'-*O*-methyltransferases. Arch Biochem Biophys **238:** 606-618

Dean JFD, Eriksson KEL (1992) Biotechnological modification of lignin structure and composition in forest trees. Holzforschung **46:** 135-147

Dewick PM, Martin M (1979) Biosynthesis of pterocarpan and isoflavan phytoalexins in *Medicargo sativa*: The biochemical interconversion of pterocarpan and 2'-hydroxyisoflavans. Phytochemistry **18:** 591-596

Dixon RA (1999) Isoflavonoids: Biochemistry, molecular biology, and biological functions. *In* U Sankawa, ed, Comprehensive Natural Products Chemistry, Vol 1. Elsevier, Amsterdam, pp 773-823

Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a metabolic grid, or independent pathways to guaiacyl and syringyl units? Phytochemistry **57:** 1069-1084

Dixon RA, Steele CL (1999) Flavonoids and isoflavonoids - a gold mine for metabolic engineering. Trends Plant Sci **4:** 394-400

Duby A, Jacobs KA, Celeste A (2000) Using synthetic oligonucleotides as probes. *In* Mea Frederick, ed, Current Protocols in Molecular Biology. John Wiley and Sons, Inc, New York

Dudareva N, Murfitt LM, Mann CJ, Gorenstein N, Kolosova N, Kish CM, Bonham C, Wood K (2000) Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. Plant Cell **12:** 949-961

Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res **8:** 163-167

Eisen JA, Kaiser D, Myers RM (1997) Gastrogenomic delights: a movable feast. Nat Med **3:** 1076-1078

Eisen JA, Wu M (2002) Phylogenetic analysis and gene functional predictions: phylogenomics in action. Theor Popul Biol **61:** 481-487

Faulkner IJ, Rubery PH (1992) Flavonoids and flavonoid sulphates as probes of auxin-transport regulation in *Curcurbita pepo* hypocotyl segments and vesicles. Planta **186:** 618-625

Forkmann G, Heller W (1999) Biosynthesis of flavonoids. *In* U Sankawa, ed, Comprehensive Natural Products Chemistry, Vol 1. Elsevier, Amsterdam, pp 713-748

Frick S, Kutchan TM (1999) Molecular cloning and functional expression of *O*-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. Plant J **17:** 329-339

Gang DR, Lavid N, Zubieta C, Chen F, Beuerle T, Lewinsohn E, Noel JP, Pichersky E (2002) Characterization of phenylpropene *O*-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. Plant Cell **14:** 505-519

Gauthier A, Gulick PJ, Ibrahim RK (1996) cDNA cloning and characterization of a 3'/5'-*O*-methyltransferase for partially methylated flavonols from *Chrysosplenium americanum*. Plant Mol Biol **32:** 1163-1169

Gauthier A, Gulick PJ, Ibrahim RK (1998) Characterization of two cDNA clones which encode *O*-methyltransferases for the methylation of both flavonoid and phenylpropanoid compounds. Arch Biochem Biophys **351:** 243-249

Goodman M, Czelusniak J, Moore GW, Romero-Herrera AE, Matsuda G (1979) Fitting the gene lineage into its species lineage: a parsimony strategy illustrated by cladograms constructed from globin sequences. System Zool **28:** 132-168

Gowri G, Paiva NL, Dixon RA (1991) Stress responses in alfalfa (*Medicago sativa* L.) 12. Sequence analysis of phenylalanine ammonia-lyase (PAL) cDNA clones and appearance of PAL transcripts in elicitor-treated cell cultures and developing plants. Plant Mol Biol **17:** 415-429

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell **13**: 73-88

Hain R, Grimmig B (2000) Modification of plant secondary metabolism by genetic engineering. *In* R Verpoorte, AW Alfermann, eds, Metabolic Engineering of Plant Secondary Metabolism. Kluwer Academic Publishers, Boston, pp 217-231

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 Mol Biol **36**: 43-54

He XZ, Dixon RA (1996) Affinity chromatography, substrate/product specificity, and amino acid sequence analysis of an isoflavone *O*-methyltransferase from alfalfa (*Medicago sativa* L.). Arch Biochem Biophys **336**: 121-129

He XZ, Dixon RA (2000) Genetic manipulation of isoflavone 7-*O*-methyltransferase enhances biosynthesis of 4'-*O*-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. Plant Cell **12**: 1689-1702

Hehmann M, Lukacin R, Ekiert H, Matern U (2004) Furanocoumarin biosynthesis in *Ammi majus* L. Cloning of bergaptol *O*-methyltransferase. Eur J Biochem **271**: 932-940

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406**: 477-483

Heller W, Forkmann G (1994) Biosynthesis of flavonoids. *In* JB Harborne, ed, The Flavonoids: Advances in Research Since 1986. Chapman and Hall, New York, pp 499-536

Huang TS, Anzellotti D, Dedaldechamp F, Ibrahim RK (2004) Partial purification, kinetic analysis, and amino acid sequence information of a flavonol 3-*O*-methyltransferase from *Serratula tinctoria*. Plant Physiol **134**: 1366-1376

Ibdah M, Zhang XH, Schmidt J, Vogt T (2003) A novel $Mg2^+$-dependent *O*-methyltransferase in the phenylpropanoid metabolism of *Mesembryanthemum crystallinum*. J Biol Chem **278**: 43961-43972

Ibrahim R (1997) Plant *O*-methyltransferase signatures. Trends Plant Sci **2**: 249-250

Ibrahim R, Muzac I (2000) The methyltransferase gene superfamily: a tree with multiple branches. *In* JT Romeo, R Ibrahim, L Varin, V De Luca, eds, Evolution of Metabolic Pathways. Elsevier Science, Amsterdam, pp 349-384

Ibrahim RK, Bruneau A, Bantignies B (1998) Plant *O*-methyltransferases: molecular analysis, common signature and classification. Plant Mol Biol **36**: 1-10

Ibrahim RK, De Luca V, Khouri H, Latchinian L, Brisson L, Charest PM (1987) Enzymology and compartmentation of polymethylated flavonol glucosides in *Chrysosplenium americanum*. Phytochemistry **26**: 1237-1245

Ibrahim RK, Anzellotti D (2003) Enzymatic basis of flavonoid diversity. *In* JT Romeo, ed, Integrative Phytochemistry: From Ethnobotany to Molecular Ecology. Plenum Press, New York, pp 1-36

Inoue K, Sewalt VJ, Murray GB, Ni W, Sturzer C, Dixon RA (1998) Developmental expression and substrate specificities of alfalfa caffeic acid 3-*O*-methyltransferase and caffeoyl coenzyme A 3-*O*-methyltransferase in relation to lignification. Plant Physiol **117**: 761-770

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res **15**: 6643-6653

Joshi CP, Chiang VL (1998) Conserved sequence motifs in plant *S*-adenosyl-L-methionine-dependent methyltransferases. Plant Mol Biol **37**: 663-674

Kagan RM, Clarke S (1994) Widespread occurrence of three sequence motifs in diverse *S*-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch Biochem Biophys **310**: 417-427

Karamenderes C, Apaydin S (2003) Antispasmodic effect of *Achillea nobilis* L. subsp. sipylea (O. Schwarz) Bassler on the rat isolated duodenum. J Ethnopharmacol **84**: 175-179

Khouri HE, Ishikura N, Ibrahim R (1986) FPLC purification and some properties of a partially methylated flavonol glucoside 2'/5'-*O*-methyltransferase from *Chrysosplenium americanum*. Phytochemistry **25**: 2475-2479

Ko WC, Chen MC, Wang SH, Lai YH, Chen JH, Lin CN (2003) 3-*O*-methylquercetin more selectively inhibits phosphodiesterase subtype 3. Planta Med **69**: 310-315

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685

Laks PE, Pruner MS (1989) Flavonoids biocides: structure-activity relations of flavonoid phytoalexin analogues. Phytochemistry **28:** 87-91

Latchinian L, Khouri H, Ibrahim R (1987) Fast protein affinity chromatography of two flavonoid glucosyltransferases. J Chromatog **388:** 235-242

Latchinian L, Ibrahim RK (1989) Characterization of a monoclonal antibody specific to a flavonol 2'-*O*-glucosyltransferase. Biochem Cell Biol **67:** 210-213

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 Physiol **129:** 1899-1907

Lee JE, Vogt T, Hause B, Lobler M (1997) Methyl jasmonate induces an *O*-methyltransferase in barley. Plant Cell Physiol **38:** 851-862

Lewis NG, Davin LB, Sarkanen S (1999) The nature and function of lignins. *In* M Pinto, ed, Comprehensive Natural Products Chemistry, Vol 3. Elsevier Science, Amsterdam, pp 617-745

Li L, Osakabe Y, Joshi CP, Chiang VL (1999) Secondary xylem-specific expression of caffeoyl-coenzyme A 3-*O*-methyltransferase plays an important role in the methylation pathway associated with lignin biosynthesis in loblolly pine. Plant Mol Biol **40:** 555-565

Li L, Popko JL, Zhang XH, Osakabe K, Tsai CJ, Joshi CP, Chiang VL (1997) A novel multifunctional *O*-methyltransferase implicated in a dual methylation pathway associated with lignin biosynthesis in loblolly pine. Proc Natl Acad Sci U S A **94:** 5461-5466

Liu CJ, Dixon RA (2001) Elicitor-induced association of isoflavone *O*-methyltransferase with endomembranes prevents the formation and 7-*O*-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. Plant Cell **13:** 2643-2658

Macheix JJ, Ibrahim R (1984) The *O*-methyltransferase system of apple fruit cell culture. Biochem Physiol Pflanz **179:** 659-664

Maddison DR, Maddison WP (2001) McClade, Version 4.03. Sinauer Associates, Sunderland, Massachusetts

Malhotra B, Onnyilagha JC, Bohm BA, Towers GHN, James D, Harborne JB, French CJ (1996) Inhibition of tomato ring-spot virus by flavonoids. Phytochemistry **43:** 1271-1276

Martz F, Maury S, Pincon G, Legrand M (1998) cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-*O*-methyltransferase, a lignin biosynthetic enzyme. Plant Mol Biol **36:** 427-437

Maxwell CA, Harrison MJ, Dixon RA (1993) Molecular characterization and expression of alfalfa isoliquiritigenin 2'-*O*-methyltransferase, an enzyme specifically involved in the biosynthesis of an inducer of *Rhizobium meliloti* nodulation genes. Plant J **4:** 971-981

Maxwell CA, Hartwig UA, Joseph CM, Phillips DA (1989) A chalcone and two related flavonoids released from alfalfa roots induce *nod* genes of *Rhizobium meliloti*. Plant Physiol **91:** 842-847

McAlister FM, Jenkins CLD, Watson JM (1998) Sequence and expression of a stem-abundant caffeic acid *O*-methyltransferase cDNA from perennial ryegrass (*Lolium perenne*). Aust J Plant Physiol **25:** 225-235

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 C, Burggraeve 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. J Biol Chem **275:** 36899-36909

Middleton E, Kandaswami C (1994) The impact of plant flavonoids on mammalian biology: implications for immunity, imflammation and cancer. *In* JB Harborne, ed, The Flavonoids: Advances in Research Since 1986. Chapman and Hall, London, pp 619-652

Mighell AJ, Smith NR, Robinson PA, Markham AF (2000) Vertebrate pseudogenes. FEBS Lett **468:** 109-114

Morishige T, Dubouzet E, Choi KB, Yazaki K, Sato F (2002) Molecular cloning of columbamine *O*-methyltransferase from cultured *Coptis japonica* cells. Eur J Biochem **269:** 5659-5667

Morishige T, Tsujita T, Yamada Y, Sato F (2000) Molecular characterization of the *S*-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-*O*-methyltransferase involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. J Biol Chem **275:** 23398-23405

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. Arch Biochem Biophys **375:** 385-388

NDong C, Anzellotti D, Ibrahim RK, Huner NP, Sarhan F (2003) Daphnetin methylation by a novel *O*-methyltransferase is associated with cold acclimation and photosystem II excitation pressure in rye. J Biol Chem **278**: 6854-6861

Osakabe K, Tsao CC, Li L, Popko JL, Umezawa T, Carraway DT, Smeltzer RH, Joshi CP, Chiang VL (1999) Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. Proc Natl Acad Sci U S A **96**: 8955-8960

Ounaroon A, Decker G, Schmidt J, Lottspeich F, Kutchan TM (2003) (R,S)-Reticuline 7-*O*-methyltransferase and (*R,S*)-norcoclaurine 6-*O*-methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. Plant J **36**: 808-819

Page RD (2000) Extracting species trees from complex gene trees: reconciled trees and vertebrate phylogeny. Mol Phylogenet Evol **14**: 89-106

Page RD, Holmes EC (1998) Molecular Evolution, A Phylogenetic Approach. Blackwell Science Ltd, Oxford

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 Physiol **103**: 509-517

Peterson G, Barnes S (1991) Genistein inhibition of the growth of human breast cancer cells: independence from estrogen receptors and the multi-drug resistance gene. Biochem Biophys Res Commun **179**: 661-667

Pincon G, Chabannes M, Lapierre C, Pollet B, Ruel K, Joseleau JP, Boudet AM, Legrand M (2001a) Simultaneous down-regulation of caffeic/5-hydroxy ferulic acid-*O*-methyltransferase I and cinnamoyl-coenzyme A reductase in the progeny from a cross between tobacco lines homozygous for each transgene. Consequences for plant development and lignin synthesis. Plant Physiol **126**: 145-155

Pincon G, Maury S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2001b) Repression of *O*-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth. Phytochemistry **57**: 1167-1176

Poulton JE (1981) Transmethylation and demethylation reactions in the metabolism of secondary plant products. *In* PK Strump, EE Conn, eds, The Biochemistry of Plants: A Comprehensive Treatise, Vol 7. Academic Press, London, pp 667-723

Preisig CL, Matthews DE, VanEtten HD (1989) Purification and characterization of *S*-adenoyl-L-methionine: 6a-hydroxymaackiain 3-*O*-methyltransferase from *Pisum sativum*. Plant Physiol **91:** 559-566

Qiu YL, Palmer JD (1999) Phylogeny of early land plants: insights from genes and genomes. Trends Plant Sci **4:** 26-30

Rakwal R, Agrawal GK, Yonekura M, Kodama O (2000) Naringenin 7-*O*-methyltransferase involved in the biosynthesis of the flavanone phytoalexin sakuranetin from rice (*Oryza sativa* L.). Plant Sci **155:** 213-221

Rausher MD, Miller RE, Tiffin P (1999) Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol **16:** 266-274

Rod AL, Oldham NJ, Svatos A, Baldwin IT (2003) Allometric analysis of the induced flavonols on the leaf surface of wild tobacco (*Nicotiana attenuata*). Phytochemistry **62:** 527-536

Ross JR, Nam KH, D'Auria JC, Pichersky E (1999) *S*-Adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. Arch Biochem Biophys **367:** 9-16

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:** 406-425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: a Laboratory Manual, Ed 3. Cold Spring Harbor Press, Cold Spring Harbor, ME

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:** 5463-5467

Sarkar FH, Li Y (2003) Soy isoflavones and cancer prevention. Cancer Invest **21:** 744-757

Sato F, Hashimoto T, Hachiya A, Tamura K, Choi KB, Morishige T, Fujimoto H, Yamada Y (2001) Metabolic engineering of plant alkaloid biosynthesis. Proc Natl Acad Sci U S A **98:** 367-372

Sato F, Tsujita T, Katagiri Y, Yoshida S, Yamada Y (1994) Purification and characterization of *S*-adenosyl-L-methionine: norcoclaurine 6-*O*-methyltransferase from cultured *Coptis japonica* cells. Eur J Biochem **225:** 125-131

Schmitt D, Pakusch AE, Matern U (1991) Molecular cloning, induction and taxonomic distribution of caffeoyl-CoA 3-*O*-methyltransferase, an enzyme involved in disease resistance. J Biol Chem **266:** 17416-17423

Schröder G, Wehinger E, Lukacin R, Wellmann F, Seefelder W, Schwab W, Schröder J (2004) Flavonoid methylation: a novel 4'-*O*-methyltransferase from *Catharanthus roseus*, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases. Phytochemistry **65:** 1085-1094

Schröder G, Wehinger E, Schröder J (2002) Predicting the substrates of cloned plant *O*-methyltransferases. Phytochemistry **59:** 1-8

Selman-Housein G, López MA, Hernández D, Civardi L, Miranda F, Rigau J, Puigdomènech P (1999) Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification. Plant Sci **143:** 163-172

Seo HS, Song JT, Cheong JJ, Lee YH, Lee YW, Hwang I, Lee JS, Choi YD (2001) Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. Proc Natl Acad Sci U S A **98:** 4788-4793

Slowinski JB, Page RD (1999) How should species phylogenies be inferred from sequence data? Syst Biol **48:** 814-825

Swofford DL (2001) PAUP. Sinauer Associate, Sunderland, Massachusetts

Takeshita N, Fujiwara H, Mimura H, Fitchen JH, Yamada Y, Sato F (1995) Molecular cloning and characterization of *S*-adenosyl-L-methionine:scoulerine-9-*O*-methyltransferase from cultured cells of *Coptis japonica*. Plant Cell Physiol **36:** 29-36

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673-4680

Thresh K, Ibrahim R (1985) Are spinach chloroplasts involved in flavonoid *O*-methylation? Z Naturforsch **40C:** 331-335

Uhrberg M (2005) The KIR gene family: life in the fast lane of evolution. Eur J Immunol **35:** 10-15

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

Vernon DM, Bohnert HJ (1992) A novel methyl transferase induced by osmotic stress in the facultative halophyte *Mesembryanthemum crystallinum*. Embo J **11**: 2077-2085

Verpoorte R (2000) Secondary metabolism. *In* R Verpoorte, AW Alfermann, eds, Metabolic Engineering of Plant Secondary Metabolism. Kluwer Academic Publishers, Boston, pp 1-29

Vidgren J, Svensson LA, Liljas A (1994) Crystal structure of catechol O-methyltransferase. Nature **368**: 354-358

Wang J, Dudareva N, Bhakta S, Raguso RA, Pichersky E (1997) Floral scent production in *Clarkia breweri* (Onagraceae). II. Localization and developmental modulation of the enzyme S-adenosyl-L-methionine:(iso)eugenol O-methyltransferase and phenylpropenoid emission. Plant Physiol **114**: 213-221

Whetten RW, MacKay JJ, Sederoff RR (1998) Recent advances in understanding Lignin Biosynthesis. Annu Rev Plant Physiol Plant Mol Biol **49**: 585-609

Willits MG, Giovanni M, Prata RT, Kramer CM, De Luca V, Steffens JC, Graser G (2004) Bio-fermentation of modified flavonoids: an example of in vivo diversification of secondary metabolites. Phytochemistry **65**: 31-41

Wink M, Waterman P (1999) Chemotaxonomy in relation to molecular phylogeny of plants. *In* M Wink, ed, Biochemistry of Plant Secondary Metabolism, Annual Plant Reviews. Sheffield Academic Press and CRC Press, Boca Raton, pp 300-341

Wu Q, Preisig CL, VanEtten HD (1997) Isolation of the cDNAs encoding (+)6a-hydroxymaackiain 3-O-methyltransferase, the terminal step for the synthesis of the phytoalexin pisatin in *Pisum sativum*. Plant Mol Biol **35**: 551-560

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 Physiol **135**: 95-102

Ye ZH, Kneusel RE, Matern U, Varner JE (1994) An alternative methylation pathway in lignin biosynthesis in *Zinnia*. Plant Cell **6**: 1427-1439

Ye ZH, Zhong R, Morrison WH, 3rd, Himmelsbach DS (2001) Caffeoyl coenzyme A
O-methyltransferase and lignin biosynthesis. Phytochemistry 57: 1177-1185

Ye ZH, Varner JE (1995) Differential expression of two O-methyltransferases in lignin
biosynthesis in *Zinnia elegans*. Plant Physiol 108: 459-467

Ylstra B, Touraev A, Moreno B, Stoger E, van Tunen AJ, Vincente O, Mol JNM,
Heberie-Brors E (1992) Flavonols stimulate development, germination and tube
growth of tobacco pollen. Plant Physiol 100: 902-907

Zhang H, Wang J, Goodman HM (1997) An *Arabidopsis* gene encoding a putative 14-
3-3-interacting protein, caffeic acid/5-hydroxyferulic acid O-methyltransferase.
Biochim Biophys Acta 1353: 199-202

Zhang XH, Chinnappa CC (1997) Molecular characterization of a cDNA encoding
caffeoyl-coenzyme A 3-O-methyltransferase of *Stellaria longipes*. J Biosci 22:
161-175

Zhong R, Iii WH, Negrel J, Ye ZH (1998) Dual methylation pathways in lignin
biosynthesis. Plant Cell 10: 2033-2046

Zubieta C, He XZ, Dixon RA, Noel JP (2001) Structures of two natural product
methyltransferases reveal the basis for substrate specificity in plant O-
methyltransferases. Nat Struct Biol 8: 271-279

Zubieta C, Kota P, Ferrer JL, Dixon RA, Noel JP (2002) Structural basis for the
modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid
3/5-O-methyltransferase. Plant Cell 14: 1265-1277

Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP (2003) Structural
basis for substrate recognition in the salicylic acid carboxyl methyltransferase
family. Plant Cell 15: 1704-1716