Tropical Plant Biol.
DOI 10.1007/s12042-008-9022-1

Comparison of Cytochrome P450 Genes from Six Plant Genomes

David R. Nelson · Ray Ming · Maqsudul Alam · Mary A. Schuler

Received: 2 November 2008 / Accepted: 14 November 2008

Abstract Plants depend on cytochrome P450 (CYP) enzymes for nearly every aspect of their biology. In several sequenced angiosperms, CYP genes constitute up to 1% of the protein coding genes. The angiosperm sequence diversity is encapsulated by 59 CYP families, of which 52 families form a widely distributed core set. In the 20 years since the first plant P450 was sequenced, 3,387 P450 sequences have been identified and annotated in plant databases. As no new angiosperm CYP families have been discovered since 2004, it is now apparent that the sampling of CYP diversity is beginning to plateau. This review presents a comparison of 1,415 cytochrome P450 sequences from the six sequenced genomes of *Vitis vinifera* (grape), *Carica papaya* (papaya), *Populus trichocarpa* (poplar), *Oryza sativa* (rice), *Arabidopsis thaliana* (Arabidopsis or mouse ear's cress) and *Physcomitrella patens* (moss). An evolutionary analysis is presented that tracks land plant P450 innovation over time from the most ancient and conserved sequences to the newest dicot-specific families. The earliest or oldest P450 families are devoted to the essential biochemistries of sterol and carotenoid synthesis. The next evolutionary radiation of P450 families appears to mediate crucial adaptations to a land environment. And, the newest CYP families appear to have driven the diversity of angiosperms in mediating the synthesis of pigments, odorants, flavors and order-/genus-specific secondary metabolites. Family-by-family comparisons allow the visualization of plant genome plasticity by whole genome duplications and massive gene family expansions via tandem duplications. Molecular evidence of human domestication is quite apparent in the repeated P450 gene duplications occurring in the grape genome.

Keywords Cytochrome P450 · Evolution · P450 superfamily · CYP · Papaya · Grape · Comparative genomics

Electronic supplementary material The online version of this article (doi:10.1007/s12042-008-9022-1) contains supplementary material, which is available to authorized users.

D. R. Nelson (✉)
Department of Molecular Sciences,
University of Tennessee Health Sciences Center,
858 Madison Ave., Suite G01,
Memphis, TN 38163, USA
e-mail: drnelson1@gmail.com

R. Ming
Department of Plant Biology,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA

M. Alam
Department of Microbiology, Advance Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii at Manoa,
2565 McCarthy Mall Keller 301,
Honolulu, HI 96822, USA

M. A. Schuler
Departments of Cell and Developmental Biology, Biochemistry and Plant Biology, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA

Introduction

As genome sequencing projects were undertaken over the past 15 years, leaders of these projects realized quite quickly that individual genomes were hard to annotate without comparison to closely related genomes. The subsequent sequencing of closely related pairs of organisms has provided the annotation bootstrapping necessary for more accurate assignment of open reading frames and intron positions. In animals, this strategy has been carried forward to some extremes with sequenced genomes now

Published online: 02 December 2008

available for 12 species of *Drosophila*, dozens of mammals, five fish and pairs of more unusual creatures, such as nematodes and sea squirts. Although plant genomes have not been deeply sampled at present, six land plant genomes (*Vitis vinifera* (grape), *Carica papaya* (papaya), *Populus trichocarpa* (poplar), *Oryza sativa* (rice), *Arabidopsis thaliana* (Arabidopsis or mouse ear's cress) and *Physcomitrella patens* (moss)) and several algal genomes (*Chlamydomonas reinhardtii* and *Volvox carteri* (green algae), *Ostreococcus tauri* and *O. lucimarinus* (microalgae), *Cyanidioschyzon merolae* and *Galdiera sulphuraria* (red algae)) have been fully sequenced and more plant species, including *Zea mays* (corn) and *Medicago truncatula* (barrel medic or barrel clover), are in the sequencing pipeline. Among these, the land plant genomes are close enough in sequence to aid in gene identification but far enough apart to provide insights into the evolutionary differences between these quite diverse species.

In this review, we focus on similarities and differences that exist in the cytochrome P450 (CYP) monooxygenase superfamily that in these six land plant genomes collectively contains 1,415 full-length and putatively functional genes. In this large number of genes, there exist only 73 families as defined in the standardized nomenclature developed for cytochrome P450 sequences [1] (Fig. 1, Supplementary Table 1). Sixteen of these families are moss-specific and the remaining 57 families encompass the P450 diversity existing in the five angiosperms. Four additional CYP families are not included in this collection because they are of limited taxonomic range. Of these, CYP719 is found only in Ranunculales and Aristolochiales, which is one of the ANITA (Amborella-Nymphaeales-Illiciales-Trimeniales-Aristolochiales) basal angiosperms [2]. CYP725 is only found in Taxus but has some overlap with the more common CYP716 family and may be an offshoot of that family. CYP726 is found only in some Euphorbia species producing vernolic acid [3] and it clusters within the CYP71 family suggesting that it is part of a CYP71 subfamily. CYP750 is present only in gymnosperms, but is somewhat related to the CYP75 and CYP92 families. The most recent review comparing Arabidopsis and rice CYP genes is [4] and the most recent review comparing moss and angiosperm CYP genes is [1].

The recently sequenced genomes of grapevine and papaya have added 457 P450 genes to the expanding collection of plant sequences. With both being fruiting woody plants and important crops, grapevine CYP genes are especially interesting since this plant has been cultivated for thousands of years and selected for flavors and colors generated by many P450-mediated biosynthetic steps. As noted in the French-Italian Consortium paper on the inbred PN40024 grape genome [5], the grape genome contains 43 stilbene synthases and 89 terpene synthases important in the

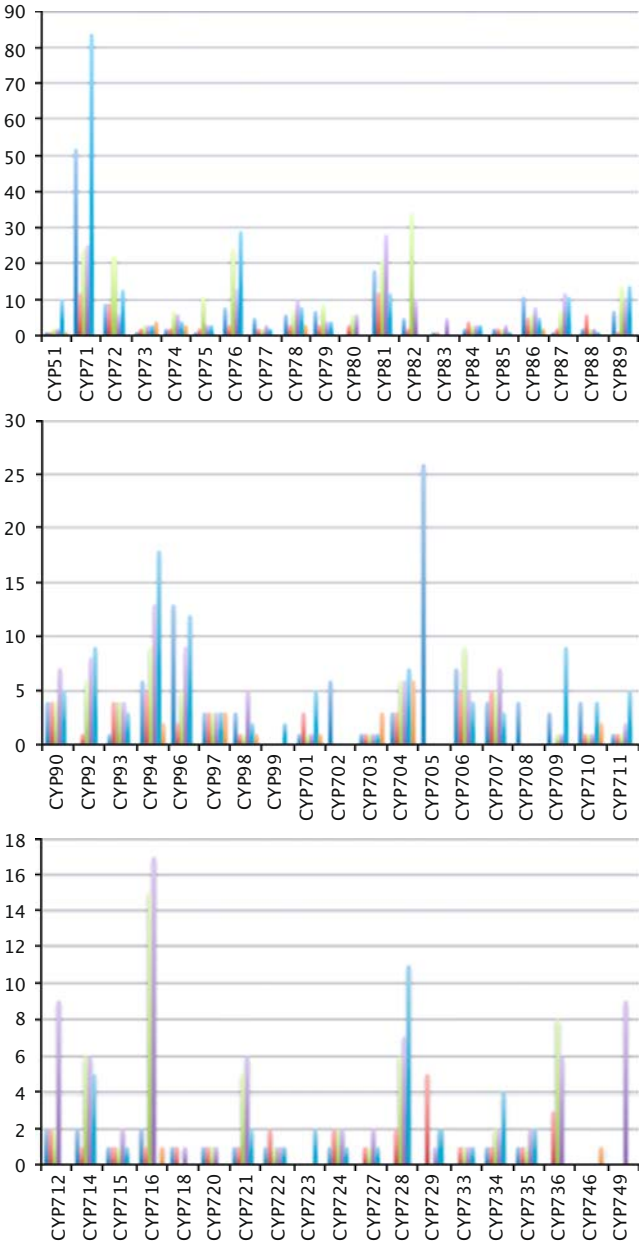

Fig. 1 Cytochrome P450 genes in each family among the six plant genomes. The order of the species in each family block is Arabidopsis, papaya, grape, poplar, rice and moss. The numbers are given in Supplementary Table 1. The number of genes tends to decrease at higher CYP family numbers that were historically discovered and named later

synthesis of resveratrol and monoterpenes that contribute to the aroma of grapes, wine and foods derived from them. Despite the obvious expansion of its secondary metabolic pathways, no mention of the multiplicity of grape P450 sequences has been included in this analysis except in their Supplemental Table S9 [5]. More recent analysis of a heterozygous Pinot Noir ENTAV 115 strain [6] identified 21 stilbene synthases and 35 terpene synthases in this genome with the large differences in gene counts for these

two protein families suggesting that the initial grape genome paper counted alleles for these loci as separate entities. The discussion of CYP genes in [6] focused on five enzymes in the phenolic and terpenoid pathways, including cinnamate 4-hydroxylase (C4H/CYP73, three genes in Supplemental Table S4), flavonoid 3'-hydroxylase (F3'H/CYP75B, one gene), flavonoid 3',5'-hydroxylase (F3'5'H/CYP75A, ten genes) and two unidentified P450s in Fig. 4b. Other than cursory asides on CYP genes in the text, Velasco et al. Supplemental Table S5 lists five CYP707A genes, and orthologues of CYP97A3, CYP701A1/GA3 and CYP711A1 [6]. To better appreciate the divergence of the plant CYP superfamily, more detailed analyses are clearly justified.

An Evolutionary Tour of Plant CYP Genes from most Ancient to most Recent

CYP Gene Families Found in all Six Plant Genomes, the Core CYPs of Land Plants

The first of several CYP gene categories to be discussed are those highly conserved orthologs that exist in a clear 1:1:1:1:1:1 pattern across all six species (Fig. 1, Supplemental Tables S1 and S2). The only three genes in this most highly conserved group are CYP97A [7] and CYP97C [8], which are carotenoid hydroxylases, and the closely related CYP97B, whose function has not yet been identified. The oldest of these three genes is CYP97B as evidenced by its higher phylogenetic identity with the CYP97E and CYP97F sequences of diatoms and other protists (Fig. 2 and Supplemental Fig. 1).

The next category contains 11 CYP families that exist in all six species, but not in 1:1 relationships and not necessarily with the same subfamilies across all six species (Fig. 1, Supplemental Table 2). The oldest of these families are CYP51 and CYP710 that are also found in non-plant species including trypanosomes and euglena (Euglenozoa). The CYP51 family, whose members are required for obtusifoliol 14 α-demethylation in plant sterol biosynthesis [9, 10], may represent the oldest eukaryotic CYP families. It appears to have undergone relatively little expansion with papaya and moss having only one CYP51G subfamily gene and Arabidopsis having one functional gene with a second one being the expressed CYP51G2P pseudogene that contains a single nucleotide deletion in exon 1 [10]. With poplar, rice and grape each having two CYP51G genes, synteny analysis has shown that CYP51G5 is part of a recent genome duplication in poplar. The grape CYP51G6 is 90% identical to grape CYP51G1, but the surrounding gene neighborhoods are very different, ruling out any large segmental duplication. The rice CYP51G3 is bracketed by

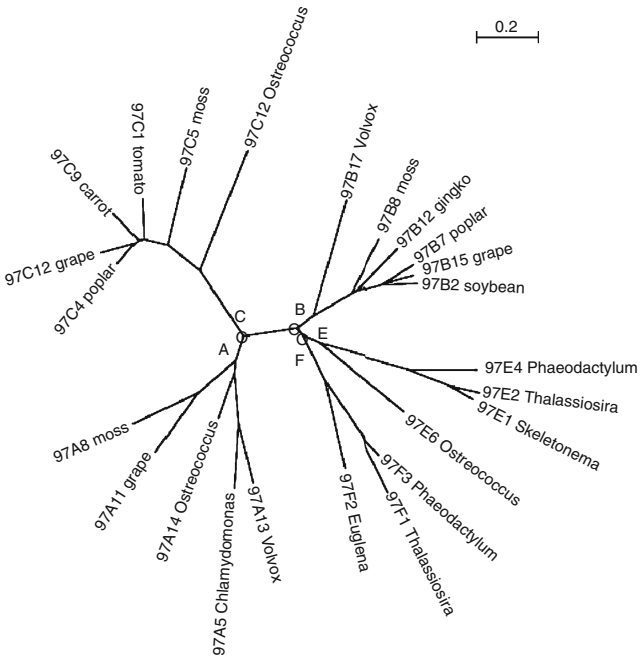

Fig. 2 An unrooted Neighbor-joining tree of CYP97 plant sequences. Sequences were aligned using the program ClustalW and the NJ tree was computed using the Phylip package

hypothetical proteins that are not syntenic with the genes around rice CYP51G1 and the percent sequence identity with Arabidopsis CYP51G1 is only 55% compared to 76% for the ortholog CYP51G1 pair. In addition to these uncharacterized CYP51G subfamily members, rice is unusual in having a divergent CYP51H subfamily with eight intact genes. A paralog of these, oat (*Avena strigosa*) CYP51H10 is synonomous with *SAD2* (saponin-deficient 2) whose protein mediates the second step in the synthesis of avenacin (an antifungal triterpene glycoside) and saponin in root tips [11]. The identification of this divergent function indicates that the CYP51 family has branched out from the exclusive synthesis of sterols to include plant defense compounds.

The CYP710 family contains sterol C22-desaturases that appear to be equivalent to CYP61 in fungi [12]. Rice and Arabidopsis each have four CYP710 genes and those of Arabidopsis are differentially regulated [12]. Poplar, papaya and grape each have only one gene while moss has two, indicating that duplication and specialization are possible but not essential in the evolution of this CYP family.

Interestingly, these three conserved families (CYP97, CYP51 and CYP710) are categorized in CYP clans each containing only one family (Fig. 3). Another single family clan conserved across all six species, CYP74, includes allene oxide synthases (AOS/CYP74A and some CYP74C sequences), hydroperoxide lyases (HPL/CYP74B) and divinyl ether synthases (DES/CYP74) with less distinct subfamily divisions. The most conserved of these subfami-

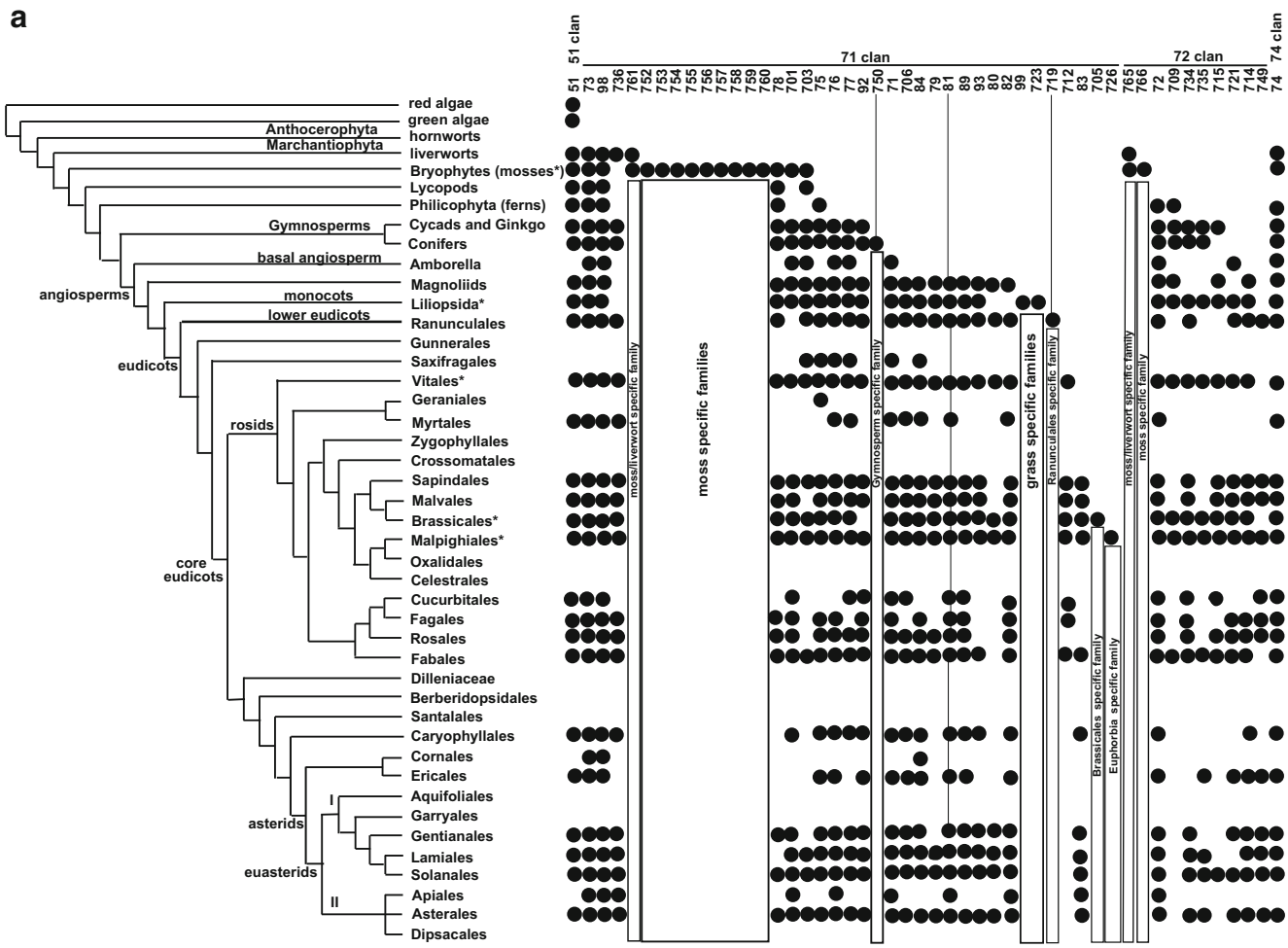

Fig. 3 Distribution of 94 cytochrome P450 families in plant phylogeny. Cytochrome P450 family data from the Cytochrome P450 Homepage was mapped onto a plant phylogeny based on [134, 135], with the placement of Vitis based on the chloroplast phylogeny of Jansen et al., 2006 [131]. The branches only represent branching order and they are not drawn to scale. The P450 families are sorted by P450 clan, with the oldest families to the left and the most limited (youngest) families to the right in each clan. The stair-step appearance of the uppermost dots illustrates the first appearance in evolution of each P450 family. Labeled boxes indicate taxon-specific families. Panel **a** covers clans 51, 71, 72 and 74. Panel **b** covers the remaining P450 plant clans and some unplaced CYP families from green algae

lies (AOS/CYP74A) has a single gene in Arabidopsis and papaya, two genes in rice, poplar and moss and six genes in grape.

The very recent publications of crystal structures for *Arabidopsis* CYP74A1 and its ortholog in *Parthenium argentatum* (guayule) [13, 14] have now provided the first insights into the mechanism by which allene oxide synthase rearranges its acyl hydroperoxide substrate to an epoxide during oxylipin biosynthesis. Unlike the more classical plant P450s that incorporate molecular oxygen into their substrates, the substrates for CYP74 enzymes provide their own oxygen in the form of hydroperoxide. The release of its oxygen binding constraints has led to structural alterations in the I-helix and heme-binding region in all CYP74 enzymes. Specifically, conserved I-helix motifs are missing from CYP74 sequences and the heme-binding region harbors a unique nine amino acid insert that is present in all CYP74 sequences in the six genomes compared here. Crystallographic analysis indicates that a highly conserved Lys in the I-helix pairs with the acyl chain carboxyl and another conserved Asn hydrogen bonds with the hydroperoxide near the heme iron. Interestingly, the CYP74 family has been identified recently in a number of animal species but not yet in fungi or vertebrates [13].

Hydroperoxide lyase activities, functionally assigned to the CYP74B subfamily in Arabidopsis [15] and *Medicago sativa* (alfalfa) [16], have also been demonstrated for the more distantly related CYP74E and CYP74F subfamilies in rice (CYP74E2/OsHPL1, CYP74E1/OsHPL2 and CYP74F1/OsHPL3) [17, 18]. These activities have also been assigned to moss CYP74G1 (AJ316567) [19] and claimed for *Z. mays* CYP74F2 (AY540745, US patent

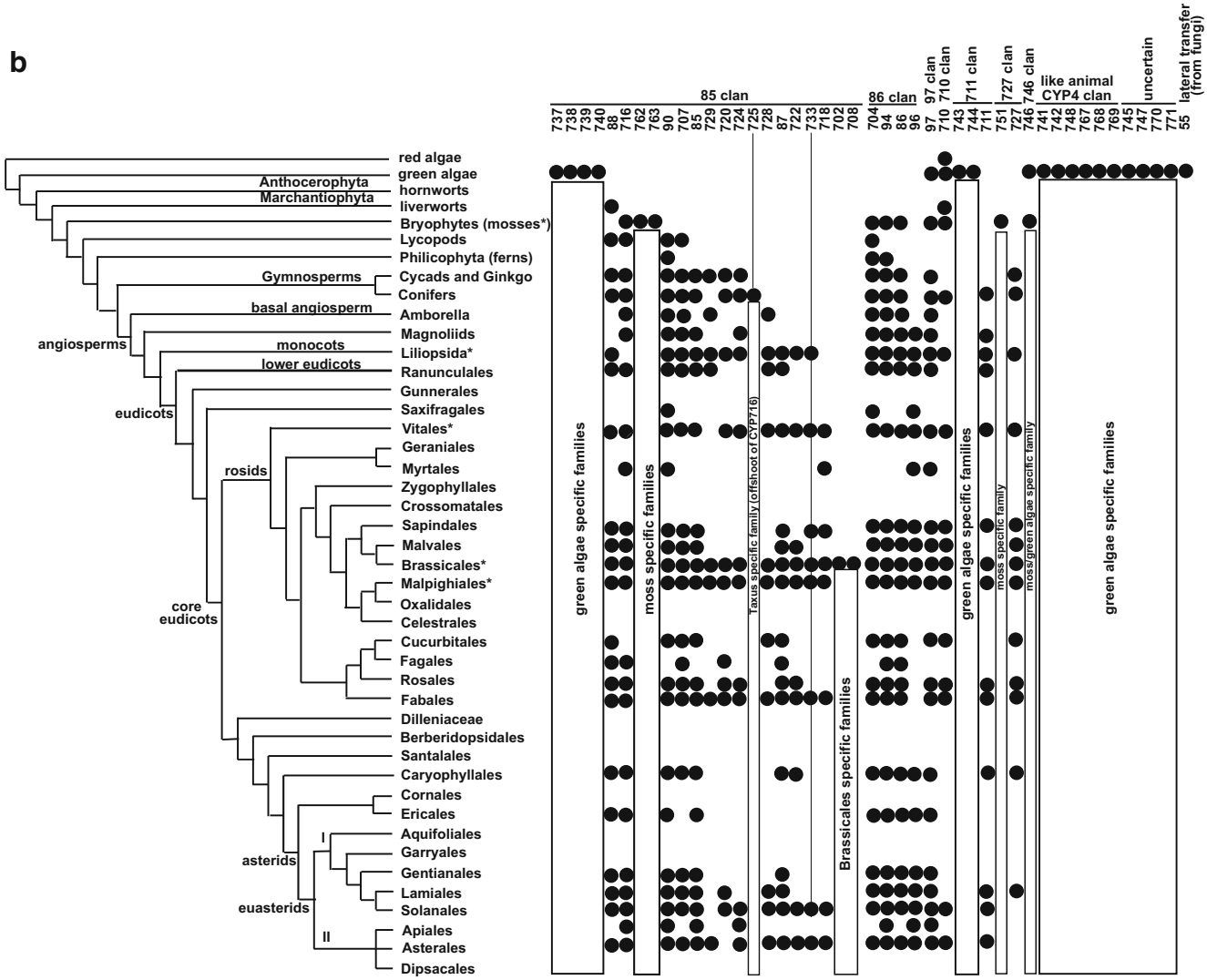

Fig. 3 (continued)

6444874 [20]), *Hordeum vulgare* CYP74F3 (AJ318870) [21] sequences. CYP74F1 is 73% identical to CYP74F2 and 71% identical to CYP74F3. Thus, HPL activities are encoded by a variety of subfamilies (CYP74B, CYP74E, CYP74F, CYP74G) existing in each of the six plant genomes included in our comparison. Less clear is the evolutionary relationship between these sequences and those annotated as divinyl ether synthases in a more limited number of Solanales species and garlic (a monocot). Comparisons of the tomato CYP74D1 (AF317515), potato CYP74D2 (AJ309541), tobacco CYP74D3 (AF070976), bell pepper CYP74D4 (DQ832721) and garlic CYP74H1 (AJ867809) proteins against CYP74A and CYP74B sequences show that the DES sequences are ~44% identical to CYP74A and ~36% identical to CYP74B sequences from Arabidopsis. The spotty distribution of DES in different CYP74 subfamilies [22] suggests that the monocot and dicot DES genes could have evolved independently from other CYP74 sequences, possibly AOS.

Two other plant CYP clans have families conserved in all land plants: the CYP86 clan that includes the CYP86, CYP94 and CYP704 families and the CYP71 clan that includes the CYP73, CYP78, CYP98, CYP701 and CYP703 families (Fig. 3). In the CYP86 clan, multiple CYP86 members mediate fatty acid ω-hydroxylations [23–27] and CYP96A15 mediates mid-chain alkane hydroxylations [28]. CYP86A8 (*LACERATA*), one of the first characterized in Arabidopsis, is a fatty acid ω-hydroxylase with well defined activities against short and medium-chain unsaturated fatty acids that are important for the development of epidermal cutin [24]. CYP86 sequences also contribute to the synthesis of suberins that are components of the underground plant cuticle [29, 30]. In contrast to moss that has only two CYP86 genes, the five

angiosperms compared here each have 5–11 CYP86 genes with those in dicots designated in the CYP86A, CYP86B and CYP86C subfamilies and those in rice designated in the CYP86A and CYP86B subfamilies (Fig. 1, Supplemental Table 2). In addition, rice has a CYP86E1 gene that may represent an orthologue of the CYP86C subfamily in dicots; it is 51% identical to Arabidopsis CYP86C1 and 55% identical to grape CYP86C8 and CYP86C10. The large number of genes in this family probably reflects the varying needs for cuticle on different plant surfaces and at different times in development. Among other CYP86 clan members, the CYP94 family that mediates hydroxylations on oxygenated fatty acids (Duan, H., Civjan, N. and Schuler, M.A. unpublished data) varies from two genes in moss, six genes in papaya up to 18 genes in rice and currently includes eight subfamilies. The CYP704 family has three to seven genes in each of the six species. Among its subfamilies, the CYP704B subfamily is conserved in all six species, while the other CYP704 subfamilies are not.

The CYP71 clan is the largest CYP clan conserved in the land plants with a current total of 808 members in these six species (Fig. 1, Supplemental Table 1 with clan membership given in Fig. 3). One potential explanation for the amplification of this clan is that emergence of plants onto land and formation of multicellular structures presented special challenges in the development of structural components including molecules for protection from UV radiation and desiccation, for production of new signaling pathways and survival of gametes in dry environments. The phenylpropanoid pathway, that originates in moss (and possibly liverworts, see Fig. 3a), provides many of the new molecules for these roles by functionalizing the aromatic phenylalanine ring. C4H/CYP73A and coumaryl shikimic acid 3'-hydroxylase (C3H/CYP98A) represent early enzymes in the core pathway and its lignin synthetic branch. While the angiosperms have from one to three CYP73 genes, moss has four genes indicating that expansion in this CYP family is not related to the complexity of these plants. Gene numbers in the CYP98 family are also not consistent with the complexity of these plants: papaya, grape and moss have single CYP98 genes, while Arabidopsis, rice and poplar have three to five genes.

In the CYP71 clan, the CYP701 family begins the pathway for gibberellin synthesis with CYP701A (*ent*-kaurene oxidase) mediating the first three reactions in the pathway and CYP88A (kaurenoic acid oxidase) mediating three subsequent steps [31, 32], that have appeared only later in plant evolution (CYP88 is not found in moss). Whereas moss, grape, poplar and Arabidopsis each have single CYP701A genes, rice has five genes and papaya has three genes. The expansions in this subfamily in these last two species seem to be nearly random. Contrasting with this, rice has only one gene in the CYP88 family, Arabidopsis has two genes and papaya has six genes (Fig. 1, Supplemental Table 2). The fact that Arabidopsis and papaya are in the same plant order (Brassicales) indicates that the trigger for gene expansion in this family is not taxon-specific.

The CYP71 clan also contains a CYP gene associated with gamete protection. Functionally characterized in Arabidopsis, this CYP703A2 gene encodes a lauric acid 7-hydroxylase that allows the product to be coupled to phenylpropanoid units in the formation of sporopollenin, a tough fluorescent polymer in the pollen wall [33]; *cyp703a2* knockout mutants lack this fluorescent layer. CYP703A subfamily members are widely conserved with most plants having just one CYP703 gene and, for unknown reasons, moss having three genes. This species-specific expansion is similar to the previously mentioned expansion in the CYP704B subfamily: all five angiosperms have one CYP704B gene whereas moss has three CYP704B genes plus additional CYP704D, CYP704E and CYP704F genes.

In contrast with other families in the CYP71 clan, the CYP78 family contains 3–10 genes in all six species analyzed here. Experiments in Arabidopsis have shown that one of its members (CYP78A5) controls organ size with *cyp78a5* knockout mutants having smaller organs and CYP78A5 overexpressing mutants having larger organs [34]. These phenotypic effects have suggested that CYP78A5 generates a mobile growth signal that may be an ω-hydroxy fatty acid as *Z. mays* CYP78A1 has been shown to mediate lauric acid ω-hydroxylations [35]. Experiments in rice have shown that CYP78A11 (*PLASTOCHRON1)* is a locus that affects the time elapsed between formation of leaf primordia in the developing plant [36]. The high degree of identity between rice CYP78A11 and Arabidopsis CYP78A5 and CYP78A7 suggests that these last two are orthologs of *PLASTOCHRON1* and that depletion of CYP78A5 shortens the plastochron period. Recent analysis has indicated that double *cyp78a5*/*cyp78a7* knockouts die as embryos without growth between the leaf primordia [37]. The existence of the CYP78 family in moss (Fig. 3a) suggests that the genes involved in this signaling system represent ancient innovations in the evolution of land plants.

In the CYP85 clan, CYP716 is found in two mosses (*P. patens* and *Selaginella mollendorffii* (gemmiferous spike moss)) and many dicots but not in rice (Fig 3b). Since the absence from rice was unexpected, a closer look at the rice sequences has indicated that its CYP728 sequences are closely related to CYP716 sequences. The fact that several species (*Amborella trichopoda* (amborella), poplar, grape and *Nicotiana tabacum* (tobacco)) have both CYP728 and CYP716 genes suggests that rice and probably other grasses have lost the CYP716 family or diverged it so extensively

that it is now designated in another CYP family. While no function is known for CYP716 or CYP728, the number of CYP716 genes is highly variable with papaya having a single CYP716 gene and grape having 14 genes (Fig. 1, Supplemental Table 2). Comparisons among a broader range of plants indicate that, in addition to the previously mentioned CYP728/CYP716 overlap, there is some overlap in the CYP716 and CYP725 families that provides evidence of the extensive divergence occurring within this subset of genes in the CYP85 clan. Specifically, CYP716B1 of *Picea sitchensis* (Sitka spruce) is 44% identical to CYP725A4 (taxadiene 5-α hydroxylase) in taxol biosynthesis in *Taxus canadensis* (Canadian yew) and some other CYP725A sequences [38, 39]. The CYP716B sequence, which was obtained after the CYP725 family was named, now appears to bridge these two families. The 44% sequence identity may suggest some limited similarity in substrates between CYP716B sequences of unknown function and CYP725A sequences that act on taxane diterpenoids.

Making the Leap to Angiosperms, Emerging CYPs from Moss to Flowering Plants

The plant genomes available today do not include transitional stages in plant evolution between moss and the angiosperms. No lycopod, fern or gymnosperm genomes exist except for that of *S. mollendorffii*, whose sequence is not yet published. With a large increase in complexity occurring in the evolution from moss to flowering plants, the number of new CYP families is correspondingly great even though the number of genes within these families may be low. Not including the eudicot-specific families and more taxon-limited families described later, there are 31 CYP families found only in angiosperms (and not in moss) with many of these representing extensions of old pathways and creations of novel pathways. Among these, the CYP720 family is the only CYP family that originated in a common ancestor of gymnosperms and angiosperms and continued as a recognizable family into the angiosperms. In phylogenetic trees, the CYP720 family is most like the CYP90 family and probably derived from it. *Pinus taeda* (loblolly pine) CYP720B1 has been shown to function as an abietadienol and abietadienal oxidase (PtAO) in diterpene resin acid biosynthesis [40]. But, given their significant divergence, it is unlikely that other angiosperm CYP720A proteins will have the same function.

The CYP736 family may also have had its origin in the early seed plants. While the original member of this family, CYP736A1 from pear, was only 41% identical to the original CYP750A1 from *Pinus*, the addition of more sequences to our current databases has revealed closer matches between the CYP736 and CYP750 families. The pine CYP750A1 is 46% identical to grape CYP736A25 and the pear CYP736A1 sequence is 47% identical to pine CYP750A2. The original separation of these families has faded and branched in such a way that CYP736 and CYP750 sequences should probably be treated as members of one family with representatives in both angiosperms and gymnosperms. The fact that the CYP750 family has not been found outside of gymnosperms and that the newest Sitka spruce sequences (EF085383 and EF086310) are more like CYP736 sequences than CYP750 sequences suggest that the CYP750 family may be derived from a CYP736 precursor. Additional sequences from a *Marchantia polmorpha* (liverwort) EST (BJ866141) showing 54% amino acid identity to grape CYP736A25 may push the origin of the CYP736 family back to the earliest land plants even though no CYP736 sequence has been found in moss. The complete sequence of this interesting liverwort CYP gene will soon be available [41] allowing us to clarify the relationship between these angiosperm and primordial sequences.

Building Pathways with CYPs

Approximately 1% of the genes in sequenced plant genomes encode full-length CYP proteins. In many cases, these gene products appear frequently in biochemical pathways with some mediating sequential steps in a single pathway, as is the case with the six CYPs existing in the brassinosteroid synthetic grid. In the evolution of pathways that have multiple CYP proteins organized in a more linear fashion with sequential steps modifying similar substrates, duplication followed by divergence of the second copy is likely to account for the incorporation of additional CYP functions in these growing pathways. Two examples in support of this are CYP85A1 and CYP85A2 in the Arabidopsis brassinosteroid pathway and CYP90B1, CYP90A1 and CYP90D1 in the same pathway. After these enzymes sequentially modify their brassinosteroid substrates and the product is subsequently reduced, the closely related CYP90C1 continues the pathway and produces substrates for the CYP85A proteins. One could suppose that two CYP85 genes and four CYP90 genes would be sufficient for brassinosteroid synthetic reactions, but recent analysis has demonstrated that CYP724B2 of tomato mediates the same C-22 hydroxylation steps in this pathway as the CYP90B proteins [42]. Two final examples of gene duplication followed by divergence in function are Arabidopsis CYP97A3 and CYP97C1 that act in sequential steps of this pathway from α-carotene to zeinoxanthin to lutein [7, 8].

Contrasting with this mode of pathway evolution, entirely different CYP gene families may be responsible for generating successive intermediates in a pathway. In gibberellin synthesis, CYP701A (in the CYP71 clan)

performs three sequential oxidations on *ent*-kaurene and CYP88A (in the CYP85 clan) oxidizes the substrate three additional times. In this instance, CYP88 was recruited into the pathway as an enzyme that could recognize the kaurenoic acid product of CYP701 in spite of its obviously different evolutionary history. Beyond these two sets of P450 reactions, the gibberellin pathway continues in a network of reactions that involve more oxidations that utilize 2-oxoglutarate-dependent dioxygenases [43] rather than any P450s. Even though logic might suggest CYP gene duplication as an option for these reactions, another type of oxidase was apparently recruited and then its gene was subjected to repeated duplication and divergence for later steps in the pathway. Interestingly, the fungus *Gibberella fujikuroi* synthesizes the same complex end product (GA_{12}) by an unrelated series of P450 reactions [44] and uses this hormone mimic in its attack on rice plants. In abscisic acid (ABA) synthesis, the grape fungus *Botrytis cinerea* generates this compound by a pathway different from plants using two CYP proteins designated CYP527D1/bcaba1 [45] and CYP563A2/bcaba2 [46]. One of the most striking examples of divergent CYP families mediating similar reactions is that of morphine synthesis in opium poppies and human neuroblastoma cells where both pathways use approximately 19 steps that include several unrelated P450 families [47]. These examples clearly indicate that CYP sequences can be recruited into pathways by gene duplication events as well as usurping unrelated sequences from different clans. They also provide evidence that organisms in different kingdoms have evolved the ability to create the same complex end products by recruiting divergent CYP proteins into novel pathways.

Plant Hormone Pathways

As mentioned earlier, CYP701A genes in gibberellin biosynthesis are present in moss and all angiosperms while CYP88A genes mediating the next three steps in this pathway are absent in moss. The presence of CYP88A in all angiosperms and *Ginkgo biloba* (maidenhair tree) (CYP88A12) suggests that this family originated in the species between bryophytes and the earliest seed plants whose genomes are not yet sequenced. However, a curious C-terminal EST (BJ852874) from liverwort *M. polymorpha* is 47% identical to Arabidopsis CYP88A4. This sequence may suggest an earlier origin, with gene loss in moss. The presence of six CYP88A genes in papaya but only one to two in other angiosperms suggests that, if the reaction catalyzed by all six genes (four in a tandem array and two separated from it) is the same, some spatial or temporal specificity requirements may have driven this expansion. An analysis of substitutions in the SRS regions of CYP88 genes suggests that the two solo CYP88 genes in papaya have diverged significantly (see "Notable features of the papaya CYP sequences" below).

The brassinosteroid pathway, which contains the previously mentioned duplicated CYP85 and CYP90 genes [48–51] is found in all angiosperms but not in moss (Fig. 3b). Moss also lacks the CYP724 family that has recently been identified as a C22-hydroxylase in brassinosteroid synthesis of tomato [42]. The sequences in moss closest to these are the CYP763 family members whose function has not yet been defined. BLAST searches of available EST databases indicate that CYP85A1, which acts in the terminal parts of the brassinsoteroid pathway, is present in gymnosperms (*Picea glauca* (white spruce), Sitka spruce, loblolly pine and ginkgo). The assembly of CYP90A18, which acts in the early part of the pathway, from white spruce ESTs and existence of CYP90A fragments in *Cycas rumphii* (cycad, DR061491) and *Adiantum capillus-veneris* (maidenhair fern, BP916923.1) suggests that the early stages of brassinosteroid synthesis were already present in these species. Additional CYP90C- and CYP90D-like sequences have been found in *Cryptomeria japonica* (Japanese cedar), white spruce, Sitka spruce and loblolly pine and a complete CYP90E1 sequence has been assembled from trace file sequences of *S. mollendorffii* (see [52] for sequences and accession numbers). Therefore the brassinosteroid pathway appears functional in gymnosperms with some evidence for early stages of the pathway in lycopods and ferns. Figure 3b summarizes the phylogenetic distribution of the CYP85 clan members.

Angiosperms synthesize the sesquiterpene hormone ABA from carotenoid precursors [48, 53]. The rate limiting enzyme 9-cis-epoxycarotenoid dioxygenase (NCED) is found in the gymnosperms *Taxodium distichum* (AB211838.1) and *Cryptomeria japonica* (AB211645.1) as well as *Picea* and *Pinus* (many ESTs). ABA is catabolized by 8'-hydroxylation mediated by members of the CYP707A subfamily [54, 55]. CYP707A EST sequences have been assembled from *Pinus* and a CYP707A20 sequence has been assembled from trace files of *Selaginella.* Although the CYP707 family is missing in *P. patens*, moss has two genes (XM_001784544.1, XM_001765944.1) that are 56% and 57% identical to *Arachis hypogaea* (peanut) NCED1 (AHY574819) suggesting that the first part of the ABA pathway might be present in moss and that the whole pathway is present in lycopods. Figure 1 and Supplemental Table S2 show that 3–7 CYP707 genes exist in each of these five angiosperm genomes.

P450 proteins are involved in the catabolism of several other plant hormones. Gibberellin GA_4 accumulates to high levels in the rice *eui1* (*ELONGATED UPPERMOST INTERNODE 1*) mutant [56]. The EUI1 gene encodes CYP714D1 that epoxidates GA_4 in the process of inactivating it [57, 58] and represents just one of six genes in three

CYP714 subfamilies existing in rice. Substrates for these other rice CYP714 proteins have not been defined, which is not surprising given that 107 different GA compounds have already been identified in plants with 12 different GA derivatives in rice, 18 in corn and 24 in Arabidopsis [32, 43, 59]. While the diversity of their structures suggests that multiple inactivating enzymes may be required to control levels of GA derivatives, CYP714 gene numbers in our five angiosperms range from one in papaya to two in Arabidopsis and six each in rice, poplar and grape. In some of these genomes having limited CYP714 numbers, breakdown of the various GAs may be mediated by alternate CYP families, as is the case for the catabolism of brassinolide, castasterone and other brassinosteroids by CYP734A1/BAS1 and CYP72C1/SOB7 in Arabidopsis [60–63]. Arabidopsis CYP734A1 has clear homologues, such as tomato CYP734A7 that mediates the same 26-hydroxylation reaction [64] and four CYP734A proteins that have not yet been biochemically characterized. Arabidopsis CYP72C1 has only two potential homologues in papaya (CYP72C2, CYP72C3) and neither of these have been functionally defined.

As plants became vascularized and developed branches, a need arose for hormone control of the branching pattern. A carotenoid-derived branch-inhibiting hormone made by the P450 CYP711A1/MAX1 is a part of the MAX pathway involving at least four genes designated MAX1-4 [65–67]. CYP711 represents another single family plant P450 clan that is found in angiosperms and conifers but not in *P. patens* indicating that it originated after the branching of the Tracheophyta (vascular plants) from the Bryophyta (mosses) (Fig 3b). Most vascular plants have a single CYP711 gene, poplar has two genes resulting from a recent genome duplication and rice has five genes. The unbranched phenotype characteristic of rice and other grasses may be evolutionarily connected to the repeated duplication of these CYP711A genes that make a branch-inhibiting hormone. Phylogenetic analysis of the CYP711 genes shows that the five rice genes appear on three distinct branches. The existence of three orthologs of these in corn (CYP711A13, CYP711A18, CYP711A19) suggests that the evolution of these three subtypes of CYP711A genes may be conserved in the grasses and they may have specialized functions (data not shown).

Analysis of the CYP711A variations within their substrate recognition sequences (designated SRS by Gotoh [68]) that contribute to defining the catalytic site cavity indicates varying degree of divergence. From the available crystal structures of bacterial and mammalian P450 proteins, SRS1 (in a series of six SRS) corresponds to the loop between the B- and C-helices positioned over the heme, SRS2 and SRS3 correspond to the F-and G-helices comprising part of the substrate access channel, SRS4 corresponds to the catalytically important I-helix extending through the back of the catalytic site, SRS5 and SRS6 correspond to the N-terminus of β-strand 1–4 and the β-turn at the end of β-sheet 4 that both protrude into the catalytic site. Using the single copy CYP711A1 sequence in Arabidopsis as the standard for comparison, the single-copy CYP711A sequences in papaya and grape maintain extremely high identity to the Arabidopsis CYP711A1 in all six SRS regions (between none and two amino acid differences) (Supplemental Fig. 2). The duplicated CYP711A sequences in poplar retain this same high degree of identity to the Arabidopsis sequence and vary from one another at only one or two positions in five of six SRS (not SRS3). The repeatedly duplicated CYP711A sequences in rice has yielded one gene (CYP711A3) sharing very high identity to Arabidopsis CYP711A1 in SRS3 (7/8 identities) and SRS4 (17/18 identities), lower identity in SRS5 (6/9 amino acid identity) and SRS1 (8/15 identities) and much lower identity in SRS6 (3/9 identities) and SRS2 (3/7 identities). Two of the other four genes in rice share slightly lower identity to Arabidopsis CYP711A1 in SRS4, SRS1, SRS2, SRS3 and SRS5 and equal identity in SRS5. Two of the remaining genes in rice include multiple divergent SRS making it likely that they mediate activities different from other CYP711A subfamily members.

Cytokinins represent another class of plant hormones made from 5'AMP with either an aromatic or an isoprene unit attached to N^6 and modified in various ways [48, 69]. Arabidopsis CYP735A1 and CYP735A2 hydroxylate cytokinin on the isoprene sidechain [70]. The five angiosperm genomes compared here have one or two CYP735A genes sharing 53–91% amino acid identity and no other CYP735 subfamilies. The limited gene numbers and lack of diversity in this family suggest that it is dedicated to this one synthetic function. CYP735A partial sequences have also been found in loblolly pine and a complete CYP735A9 has been assembled from ESTs of the cycad *C. rumphii* but the CYP735 family is not present in moss.

Phenylpropanoid Pathway Leading to Lignins and Flavonoids

CYP proteins are also used to make phenylpropanoids, flavanoid and anthocyanin pigments and plant volatiles. The phenylpropanoid pathway mentioned earlier in connection with CYP73 and CYP98 has undergone much elaboration in gymnosperms and angiosperms. In contrast, CYP84 (F5H) is present in angiosperms but not in moss or gymnosperms (Fig. 3a), which is consistent with the fact that they do not make syringyl lignin monomers. In *Selaginella*, it appears that convergent evolution has allowed this species to make syringyl lignin monomers via a different P450 family (CYP788A1, EU032589) than that used in angiosperms

[71]. While this sequence did not cluster in a Bayesian tree with F5H/CYP84 sequences, evidence supporting this functional assignment exists in the fact that it complements the Arabidopsis *fah1* knockout mutant and has substrate preferences that match angiosperm F5H.

After coumaryl-CoA the phenylpropanoid pathway branches, with both branches leading to flavonoids among other products. The CYP75 family that plays an important role in this pathway includes CYP75A (F3'5'H) and CYP75B (F3'H) that participate in the production of delphinidin (blue) and cyanidin (red-orange) anthocyanin pigments [72] that are important for wine color in grapes and flower/fruit color in other plants. The CYP75A genes so important for the wine industry have repeatedly duplicated in grape to a total of nine genes where papaya has only one CYP75A gene (Fig. 1, Supplemental Table S2). Other industrial applications associated with the CYP75A subfamily have included genetic engineering of them into carnations, chrysanthemums and roses in an attempt to impart blue hues in their flowers [73].

Isoflavonoids and Plant Defense Molecules

Isoflavonoids branch from the phenylpropanoid pathway after production and isomerization of chalcone to form flavanones [74, 75]. CYP proteins involved in flavonoid metabolism include isoflavone 2'-hydroxylase CYP81E1 in *Glycyrrhiza echinata* (licorice) [76], CYP81E3 in *Cicer arietinum* (chickpea) [77] and CYP81E7 in barrel medic [78], isoflavone 3'-hydroxylase CYP81E9 in barrel medic [78], dihydroxypterocarpan 6α-hydroxylase CYP93A1 in *Glycine max* (soybean) [79], the flavone synthase II CYP93B subfamily [80] and the isoflavone synthase CYP93C subfamily in legumes [81–84], the triterpene hydroxylase CYP93E subfamily in legumes [85] and flavone 6-hydroxylase CYP71D9 in soybean [86]. Note that many of these represent highly specialized P450s are represented only in legumes.

Sesquiterpenes derived from farnesyl pyrophosphate in the isoprenoid pathway are modified by P450s in the CYP706 family. *Gossypium arboreum* (cotton) CYP706B1 is a (+)-delta-cadinene hydroxylase in the pathway to synthesizing gossypol, a sesquiterpene aldehyde [87]. Related CYP proteins, products of the CYP706B3, CYP706C5, CYP706C6, CYP706D1 and CYP706D2 genes in poplar may act on poplar-specific sesquiterpenes but heterologous expression of these is needed to confirm this. The monoterpene limonene is hydroxylated at various positions by CYP71D13, CYP71D15 (*Mentha* x *piperita*, peppermint) and CYP71D18 (*Mentha spicata*, spearmint) with the regiospecificities of these so tightly controlled that single amino acid changes in their active sites are capable of altering 6-hydroxylase functions to 3-hydroxylase functions [88, 89]. The fact that the related CYP71D9 mentioned earlier was a 6-hydroxylase on flavones highlights the substrate differences that can occur even within a single subfamily.

Other groups of plant defense compounds synthesized by P450-dependent reactions include the glucosinolates that are cleaved by myrosinases into bioactive toxins [90]. Two classes of these S-containing compounds are derived from IAA via parallel branched pathways that synthesize indole glucosinolates via CYP79B and CYP83B subfamily members [91–95] and aliphatic glucosinolates via CYP79F and CYP83A subfamily members [93, 95–99]. The papaya CYP83B1 gene is 67% identical to Arabidopsis CYP83B1 suggesting they are orthologs. Global phylogenetic analysis of the CYP71 clan members has suggested that the CYP83 family might be better classified as a subfamily of the CYP71 family, but it has been kept as an independent family for historical reasons [4]. Related cyanogenic glucosides, which are found in a much broader range of plants, are synthesized via CYP79A and CYP71E subfamily members [100, 101].

The remaining 14 angiosperm-specific families CYP72, 76, 77, 87, 89, 92, 709, 715, 721, 722, 727, 728, 729, 733 that we have not specifically discussed have few functions assigned to them at present. The large size of some of these families and subfamilies suggest that their functions will not be limited to one particular reaction or pathway and that, as defense toxin pathways duplicated, many of these evolved new functions for individual angiosperms. Among the few functions assigned to these families, CYP72A1 (*C. roseus*, Madagascar periwinkle) is identified as secologanin synthase that converts loganin into secologanin in vinca indole alkaloid biosynthesis [102]. Two additional P450s in this *C. roseus* pathway are tabersonine 16-hydroxylase CYP71D12 [103] and geraniol 10-hydroxylase CYP76B6 [104]. CYP76A4 and CYP76B9 (*Petunia hybrida*, petunia) hydroxylate lauric acid at its ω- or ω-1 positions [105, 106]. CYP71D20 (*Nicotiana tabacum*, tobacco), the fourth CYP71D subfamily sequence with a function assigned, acts to hydroxylate 5-epi-aristolochene and 1-deoxycapsidiol to form the defense compound capsidiol, a bicyclic, dihydroxylated sesquiterpene [107]. CYP709C1 (*Triticum aestivum*, wheat) hydroxylates fatty acids at the ω-1 and ω-2 positions and is induced by methyl jasmonate, possibly as a defense response [108].

Continued Specialization: Eudicot Specific CYP Gene Families

Nine CYP families appear to have arisen in dicots (CYP80, 82, 83, 702, 708, 705/712, 718, 749) after their split from monocots. Among these, the CYP80 family was initially found only in Ranunculales and later in other angiosperms

like poplar, papaya and grape. CYP80B1 is required for the pathway to benzophenanthridine alkaloids in *Eschscholzia californica* (California poppy), CYP80B3 is required in the pathway to morphinan alkaloids in *Papaver somniferum* (opium poppy) [109, 110] and the CYP80A, CYP80F and CYP80G subfamilies are also involved in alkaloid biosynthesis. CYP80A1 mediates C-O phenol coupling in the bisbenzylisoquinoline alkaloid synthesis in *Berberis stolonifera* (barberry) [111]. CYP80F1 is involved in the rearrangement of littorine to hyoscyamine in *Hyoscyamus niger* (black henbane) [112]. CYP80G2 catalyzes an intramolecular C-C phenol coupling of (S)-reticuline in magnoflorine biosynthesis in *Coptis japonica* (Japanese goldthread) [113]. Several of these have been used in semi-synthetic strategies to produce alkaloids using *E. coli* and yeast expression systems [114]. The role of additional CYP80 subfamilies in grape, poplar, papaya and *Sesamum indicum* (sesame) are not known. Our phylogenetic comparisons suggest that the CYP80 family is probably derived from the CYP76 family (data not shown).

Other CYP sequences mediating plant-specific alkaloid pathways are found in the CYP82 family that includes tobacco CYP82E5, a nicotine demethylase involved in conversion of nicotine to nornicotine [115, 116] and Arabidopsis CYP82C2 and CYP82C4, 8-methoxypsoralen hydroxylases mediating modifications on toxic furanocoumarins [117]. In comparison with the two full-length CYP82 genes (CYP82C17, CYP82L1) in the papaya genome, the grape genome has16 CYP82D sequences, six CYP82H sequences and 12 CYP82S sequences. In addition to this 17-fold increase in the number of full-length CYP82 genes in grape, this genome also has 34 CYP82 pseudogenes corresponding to identifiable exon fragments or prematurely truncated open reading frames. The presence of this large number of pseudogenes is suggestive of actively evolving gene clusters as seen in the human CYP2ABFGST cluster and several rodent clusters [118].

The CYP705/CYP712 families are like the CYP83/CYP71 families where early classifications placed them in distinct families and later discoveries challenged this separation. In the case of CYP705/CYP712 sequences, the identification of CYP712D subfamily sequences in grape and soybean that are about 45% identical to both families indicated that CYP705/CYP712 should be merged to reflect the closer evolutionary relationships of these two groups; nevertheless, the CYP705A genes of Arabidopsis and Brassica have been kept separate for historical reasons. The CYP705/CYP712 family has been derived from the CYP93 family as is most apparent from the fact that the Arabidopsis CYP712B1 gene occurs adjacent to the CYP93D1 gene in this species. And, in poplar, CYP93A4, CYP93A7P and CYP712C1 are adjacent in an 8 kbp region of LG_XIV, with CYP712A5 165 kbp upstream from this block. In a second conserved region, CYP712A3v1, CYP712C2, CYP93A6 and CYP93A5P are adjacent in a 13 kbp block on scaffold_152. Likewise, in papaya, CYP712A8, CYP712A9, CYP93A10, CYP93A11 and CYP93A12P are located within a 27 kbp region on supercontig_85. We have concluded that synteny between the CYP93 and CYP712 gene families has been preserved in three genomes from the eurosids, but not in the grape genome at the base of the rosids.

With 25 CYP705A subfamily members existing throughout the Arabidopsis genome, a function has only been assigned to CYP705A5, the thalian-diol desaturase in the thalianol pathway [11]. Although the CYP705A gene subfamily diverged from the CYP712 family, seen in all four dicots compared here, CYP705A is limited to Brassicales and is presumably central to certain Brassicales specific pathways. One key observation of Field and Osbourn [11] is that CYP705A5 is part of a four gene metabolite cluster involving CYP708A2 (thalianol hydroxylase) and a triterpene synthase (thalianol synthase). They suggest that this linkage of P450 genes and a triterpene synthase gene may be exploited in other Arabidopsis pathways through duplication and divergence of gene blocks that preserve a metabolite gene cluster. Consistent with this, CYP708A3, which is one of three other CYP708 genes in Arabidopsis, is adjacent to the At1g78480 locus, which is similar to a pentacyclic triterpene synthase and resembles the thalianol synthase that occurs upstream of CYP708A2 in the thalianol synthesis pathway. This juxtaposition suggests a similar functional relationship between CYP708A3 and its adjacent pentacyclic triterpene synthase. The fact that the CYP705 and CYP708 genes are in different CYP clans indicates that they have not been brought together by tandem gene duplication but rather by significant rearrangements within the ancestral Arabidopsis genome.

The two P450 and triterpene synthase linkages noted above establish a precedent and suggest that additional pairings may exist. Further searching for P450s adjacent to terpene synthases uncovered a large cluster on Arabidopsis chromosome 4 with eight P450s and two pentacyclic triterpene synthases. The gene order from At4g15310 to At4g15396 is: CYP702A2, CYP702A3, CYP705A1, pentacyclic triterpene synthase (ATPEN1), CYP702A4P, CYP705A2, CYP705A3, baruol synthase (BARS1), CYP705A4, CYP702A5, CYP702A6. The BARS1 gene has been reported to make baruol and 22 additional triterpene products [119], some of which may be substrates for these P450s, but the ATPEN1 product is not known [120]. Even so, it is noteworthy that the CYP702 and CYP705 genes are intermingled even though they are in different CYP clans and do not share a close evolutionary history. The CYP705A, CYP702 and CYP708 genes are all

unique to Arabidopsis and its closest relatives, such as *Brassica napus*, making these the only known Brassicales-specific CYP proteins (Fig. 3).

The CYP718 family appears in P450 gene collections of asterids and rosids but infrequently in EST and cDNA collections. The CYP718 gene is present in potato (*Solanum tuberosum*), tomato (*Lycopersicon esculentum*), poplar, papaya, grape, barrel medic and Arabidopsis. One full-length CYP718 cDNA has been identified in Arabidopsis and seven CYP718 ESTs have been identified in Viridiplantae (in a collection of 18,215,441 ESTs) but no ESTs are found in 232,980 ESTs from potato, 259,990 ESTs from tomato, 89,943 ESTs from poplar, 352,984 ESTs from grapevine, 77,158 ESTs from papaya, 259,969 ESTs from barrel medic or 1,526,133 ESTs from Arabidopsis. This suggests that CYP718 transcripts are expressed only transiently or at extremely low abundance. One of the seven CYP718 ESTs is from *Thellungiella salsuginea,* an Arabidopsis-like extremophile tolerant of cold, drought and salinity. At the amino acid level, CYP718 is most similar to the older CYP716 family that has one gene in papaya, two genes in Arabidopsis, 17 genes in poplar (not counting individual variants or pesudogenes for the different loci) and 14 genes in grape.

The CYP749 family is the last of the dicot-specific families. This family, which is in the CYP72 clan, is first observed in Ranunculales (Aquilegia hybrid columbine, DR921551, Fig. 3a) in the lower eudicots. Full-length CYP749 sequences are present in the poplar and tobacco genomes and as EST contigs from Aquilegia and cotton. The absence of CYP749 in the monocots and magnoliids and its appearance in Ranunculales suggest that it represents an early CYP branch in dicot evolution. Nothing is known about the biochemical function of this family.

At the beginning of this section, nine CYP families were listed as dicot-specific. We have now reduced that number to seven by viewing CYP83 as a part of the CYP71 family, by merging CYP705/CYP712. At this level of phylogenetic analysis, it appears that innovation in angiosperm P450s has been concentrated primarily at the subfamily level or lower with few new families emerging in the angiosperms.

Evolution in Recent Time: Highly Limited CYP Families

Among these many CYP families exist some that are only found in one or a few closely related species. The CYP726 family is only known from *Euphorbia lagascae* and *Euphorbia esula* where it mediates synthesis of vernolic acid (an epoxy fatty acid) [3]. Phylogenetic trees place CYP726 inside the CYP71 family suggesting that it should be viewed as a subfamily rather than an independent family. The previously mentioned CYP702 and CYP708 families are found in Arabidopsis and closely related Brassica species, but absent from papaya that is also in the Brassicales. At least some of these CYP proteins are involved in modifying triterpenes, including the CYP708A2 thalian-diol-desaturase [11] that was mentioned in the previous section (Eudicot-specific gene families).

The CYP99 and CYP723 families exist in grasses and not in any other sequenced genomes. Once again, CYP99 falls inside the CYP71 family as the boundaries of this largest plant CYP family continue to expand with more sequences becoming known. In rice, the CYP99A2 and CYP99A3 loci are found in a five-gene biosynthetic gene cluster on chromosome 4 mediating the synthesis of momilactones [121]; these P450s have been shown to act on a diterpenoid precursor generated by the diterpene cyclases OsCyc1 and OsKS4. Another gene cluster existing on rice chromosome 2 includes OsCyc2, OSDTC1 and multiple CYP genes and is implicated in the biosynthesis of phytocassanes (novel phytoalexin cassane-type diterpenes) [122]. Examination of Japonica strain contig AP008208.1 indicates that a ten-gene cluster includes CYP76M5, CYP76M8, CYP76M7, OSDTC1, CYP71Z6, CYP71Z7, OsCyc2, OsKS5, OsKS6, CYP76M6 (proteins BAF09097 to BAF09106 in Entrez Protein). The four diterpene cyclases and six P450s in this cluster are reminiscent of those that exist in the thalianol and triterpenoid pathways in Arabidopsis. The many ring and sidechain modifications that occur in terpene biochemistry may account for a considerable portion of the P450 enzymes in plants and also for the metabolic gene clusters in fungi [123, 124].

The CYP723 family clusters inside the CYP89 family and might be considered a CYP89 subfamily (see the top of Fig. 5 in [4]). The larger CYP89 family itself may be an offshoot of the closely related CYP77 family. According to the data in Fig. 3a for the CYP71 clan, the CYP77 family appears to be more ancient with members in the cycads and ginkgo, while CYP89 is first observed in magnoliids. In rice, CYP723A3 is adjacent to CYP723A1 on chromosome 4, and it is 41% identical to CYP89C1 on chromosome 6 that has no known function. While the CYP89 family and most CYP77 family members contain no introns, the rice CYP77A9 gene has a single intron.

Comparison of the Slowly Evolving Papaya Genome to the Rapidly Evolving Grape Genome

Notable Features of the Papaya CYP Sequences: Stasis in a Conservative CYP Genome

Of the five angiosperm genomes considered here, papaya has the fewest full-length CYP genes (142) and considerably less than the numbers of CYP genes found in poplar (310), grapevine (315), rice (332) and Arabidopsis (245)

(Supplemental Table S1). Even with this small CYP count, there are only two of the mainstream CYP families (CYP709, CYP749) (not including the taxon-limited families CYP99, CYP702, CYP705, CYP708, CYP719, CYP723, CYP725, CYP726 and CYP750) that appear to be lost in papaya. Papaya does have all of the other CYP72 clan families of vascular plants CYP72, CYP714, CYP715, CYP721, CYP734 and CYP735. The presence of CYP709 and CYP749 in other rosids and eurosids argues for their loss in Brassicales (Fig 3a, CYP72 clan), a fairly common event in the evolution of P450 sequences. Examples of this can be found in Arabidopsis that has lost a significant number of mainstream families including CYP80, CYP92, CYP724, CYP727, CYP728, CYP729, CYP733, CYP736 and CYP749 as well as grape and rice that have lost CYP749.

Papaya is nearly minimalist in its CYP collection with no single subfamily having more than seven genes and 42 subfamilies having only one gene. In contrast, Arabidopsis, that is also in the order Brassicales contains 36 CYP71B subfamily genes and 25 CYP705A subfamily genes. It appears that Arabidopsis has nearly doubled its CYP gene numbers even though it lacks eight CYP families existing in papaya. Comparisons among these plant genomes indicate that P450 evolution has repeatedly occurred by retaining the core CYP genes needed for critical functions and increasing the number of CYP genes and the diversity of CYP subfamilies. With its lower number of CYP genes, papaya provides a snapshot of an angiosperm genome that has retained much of its early character with only twice the number of CYP genes found in the moss genome, but displaying all the complexities of a woody, fruiting plant.

There are additional unusual features of CYP gene evolution in papaya compared to the other dicots included in our comparison. The first of these is that the pathway for gibberellin synthesis has expanded substantially. Grape, poplar and Arabidopsis each have one CYP701 gene (*ent*-kaurene oxidase mediating the first three steps in GA_{12} synthesis) while papaya has three genes. Grape, poplar and Arabidopsis each have two CYP88 genes (kaurenoic acid oxidase mediating the next three steps) while papaya has six genes. For both sets of loci, this represents a three-fold increase in the papaya CYP sequences responsible for early reactions leading to GA_{12}. In Arabidopsis, gibberellins derived from GA_{12} are sorted into three types: early 3-hydroxylated products, early 13-hydroxylated products and non-early-3,13-hydroxylated products [32, 43]. The early 3-hydroxylation pathway depends on gibberellin 3β-hydroxylase (*GA4*) that is a 2-oxoglutarate Fe^{2+} dioxygenase and not a P450. The early 13-hydroxylation pathway depends on a 13-hydroxylase that has not yet been identified. The non-early-3,13-hydroxylation pathway depends on gibberellin 20-oxidase (*GA5*) that is another 2-oxoglutarate Fe^{2+} dioxygenase. The three-fold reiteration of CYP701 and CYP88 genes in papaya compared with other dicots suggests that there may be specialization in papaya to allow for the production of different gibberellins requiring unique multi-enzyme complexes or metabolons for the three different pathways.

Comparisons of the SRS regions in the CYP701 and CYP88 genes of papaya and other dicots provide support for this suggestion (Supplemental Figs. 3 and 4). The CYP701A genes in grape and poplar are among those most closely related to Arabidopsis CYP701A3 with two differences in SRS1, two or four differences in SRS2 and SRS5, five or six differences in SRS3, four differences in SRS4 and two or three differences in SRS6 (Supplemental Fig. 3). The duplicated CYP701A genes in papaya have diverged further from Arabidopsis CYP701A3 with six to eight differences in SRS1, two differences in SRS2, four differences in SRS3, five to seven differences in SRS4, one or three differences in SRS5, one or five differences in SRS6. Three of the five CYP701A genes in rice are more closely related to Arabidopsis CYP701A3 in SRS1 than either of these papaya sequences, equivalently related in SRS2 and more divergent in SRS3 and SRS5; for these three rice genes, differences in SRS4 and SRS6 are intermediate between the two papaya CYP701A sequences. The remaining two CYP701A3 genes in rice exhibit substantially further divergence from Arabidopsis CYP701A3 and the other rice CYP701A sequences.

In comparison, the duplicated CYP88A genes in Arabidopsis contain one difference in SRS2, SRS3 and SRS4, three differences in SRS1 and SRS6 and no differences in SRS5 (Supplemental Fig. 4). The two CYP88A sequences in grape are slightly more divergent from Arabidopsis CYP88A3 with one difference in SRS1 (restricted to CYP88A23), SRS2 and SRS3, two to three differences in SRS4 and SRS5 and two differences in SRS6. Similarly, the two poplar CYP88A sequences show low variation from Arabidopsis CYP88A3 with one difference in SRS2 and SRS3, one to two differences in SRS4, SRS5 and SRS6 and three differences in SRS1. Of the six CYP88A genes in papaya, one (CYP88A17) has SRS regions more closely related to the Arabidopsis CYP88A3 sequences than its two closest relatives (CYP88A18, CYP88A19) and two others that are separated from the main cluster (CYP88A20, CYP88A21) have SRS regions that have diverged significantly from the Arabidopsis, grape and poplar sequences. Of these, CYP88A17 contains six differences in SRS1, two differences in SRS4, four differences in SRS5 and three differences in SRS6 but none in SRS2 or SRS3. CYP88A18 and CYP88A19 contain six or seven differences in SRS1, one or two differences in SRS2 and SRS3, one or three differences in SRS4, three or five differences in SRS5, two or three differences in SRS6. CYP88A20 and CYP88A21 contain

seven differences in SRS1, three or five differences in SRS2, one or two differences in SRS3, six differences in SRS4 and SRS5 and two differences in SRS6. In contrast to these last two papaya CYP88A genes, the single-copy CYP88A gene in rice maintains substantial identity to Arabidopsis CYP88A3 with one difference in SRS2, two differences in SRS3 and SRS6, three differences in SRS4 and six differences in SRS1 and SRS5.

The second unusual feature of CYP gene evolution in papaya is that the pathway for lignin synthesis has expanded in papaya compared to Arabidopsis probably because of the increased need for phenolics in wood production. Papaya contains two CYP73A genes (cinnamate 4-hydroxylase in the core phenylpropanoid pathway) and four CYP84A genes (ferulate 5-hydroxylase in the lignin branch pathway) compared to one CYP73A and two CYP84A genes in Arabidopsis. Among the sinapoylglucose, sinapoylmalate, and sinapoylcholine that are produced in Arabidopsis as the result of CYP84A1 action [125–127], sinapoylcholine may serve as a seed reservoir of choline for phospatidylcholine synthesis and sinapoylmalate protects leaves against UV-B radiation. Papaya may use its additional CYP84A genes to further modify these compounds producing the multiply hydroxylated compounds needed in wood.

With the CYP84A subfamily sharing high overall identity (48–97%) in these five species, the duplicated CYP84A sequences in Arabidopsis contain multiple differences in SRS1 (13/15 identities), SRS3 (6/8 identities) and SRS6 (6/9 identities) and single differences in the other three SRS regions (Supplemental Fig. 5). The single-copy CYP84A31 gene in grape is identical in its SRS regions to Arabidopsis CYP84A1 involved in lignin production except for two differences in SRS3 and one difference in SRS6. Of the three CYP84A genes in poplar, two share extremely high identity to Arabidopsis CYP84A1 with one or two differences existing in SRS1 and SRS6, three differences existing in SRS3 and no differences existing in the other three SRS regions. The remaining CYP84A12 gene in poplar contains an additional variation in SRS4, three additional variations in SRS1 and fewer variations in SRS3 but these SRS sequences remain more related to Arabidopsis CYP84A1 than the divergent Arabidopsis CYP84A4. Of the four CYP84A genes in papaya, two again share extremely high identity to Arabidopsis CYP84A1 with only two or three differences in SRS3, one difference in SRS2, one difference in SRS6 (restricted to CYP84A26) and no differences in the other three SRS regions. In the remaining two CYP84A28 and CYP84A29 genes, two to five differences from Arabidopsis CYP84A1 exist in SRS1, three differences exist in SRS3, one difference exists in SRS4, SRS6 and SRS5 (restricted to CYP84A29). Probably because of the universally important nature of the CYP84A activities in lignin production, variations in these SRS regions are limited even in the evolutionarily distant rice CYP84A sequences. Of the three CYP84A genes in rice, two again are more closely related to Arabidopsis CYP84A1 with two differences in SRS1, SRS3, SRS4 (restricted to CYP84A6) and one difference in SRS6 and SRS5 (restricted to CYP84A5). The remaining CYP84A7 gene is significantly more divergent with two to seven differences in SRS1, SRS2, SRS3 and SRS5 and an insertion and three differences in SRS4; in surprising contrast, its SRS6 is identical to that found in Arabidopsis CYP84A1. The extensive SRS identity found in the CYP84A1 subfamily contrast significantly with variations in the other CYP subfamilies whose SRS sequences are summarized in this paper.

The third unusual feature of papaya compared to Arabidopsis is the large size of its fruit. With several CYP71 family members known to participate in ripening and production of fruit flavors and aromas in avocados (CYP71A1), bananas (CYP71N1) and strawberries (CYP71AR1) [128–130], papaya contains 12 CYP71 genes in four subfamilies. Seven of these genes are in the CYP71B subfamily that has 35 members in Arabidopsis and the other five genes are in new subfamilies not shared with Arabidopsis. Three of these five, CYP71BF1, CYP71BF2 and CYP71BE2 are candidates for fruit-related hydroxylations with CYP71BF1 being 52% identical to strawberry CYP71AR1 (10α-pinene hydroxylase) [130] and CYP71F2 being 54% identical to tomato CYP71AU2 and most similar to the Arabidopsis CYP71A subfamily. CYP71BE2 has been assembled from papaya ESTs EX262791 and EX271814 isolated from a mixed tissue library and CYP71BE1 has been identified in grape, another fruiting woody plant. The existence of eight CYP71BE subfamily genes and nine pseudogenes in the grape genome suggests that active evolution is occurring in that subfamily, possibly due to intense selection for fruit characteristics.

Notable Features of the Grape CYPs: Evolution Via Human Selection

In phylogenetic analysis, grape has been very difficult to place in evolutionary history and was only recently assigned on the basis of complete chloroplast genome sequences to the base of the rosids [131]. Arabidopsis and papaya are both in Brassicales making them evolutionarily closer to each other than they are to grape. However, BLAST searches with grape CYP sequences have found the best hits to be with many papaya CYP sequences. This closer degree of identity between papaya and grape suggests that papaya has evolved more slowly than Arabidopsis after the two diverged and is in accord with

the smaller numbers of CYP sequences in papaya. Sequencing of the grape genome has revealed hexaploidization of this genome prior to the last common ancestor of Arabidopsis and grape and before the radiation of the eurosids. Two additional whole genome duplications took place in the lineage leading to Arabidopsis such that, as noted by Jaillon et al. [5] and Lyons et al. [132], most regions in the grape genome are reiterated four times in the Arabidopsis genome. Sequencing of the papaya genome found that 30 of 121 co-linear segments matched four Arabidopsis segments implying that papaya diverged from Arabidopsis before these two duplication events took place. Since comparison between papaya and grape CYP sequences does not involve any large-scale genome duplications since their last common ancestor, the increase in grape CYP genes must be attributed to small-scale duplication events, such as tandem duplications, but not to whole genome duplication events.

Analysis of the CYP differences between grape and papaya has potential to reveal P450 functions critical to the evolution of fruit characteristics in grape, a process that has been heavily influenced by human selection. Large expansions in some families such as the CYP75 family are likely to be important in this regard. Within this family, the CYP75A subfamily has expanded from one gene in papaya to nine genes in grape with CYP75A28 (DQ298201-DQ298203) and CYP75A36 (DQ298204-DQ298205) being flavonoid 3'5'-hydroxylase genes important for production of berry color [72]. While this 2006 paper did not yet know the full extent of CYP75A genes in the grape genome, special attention was given to these genes in [6] after sequencing of the Pinot Noir strain ENTAV 115 genome. Other interesting gene expansions occur in the CYP76 family where papaya has three genes and grape has 24 full-length CYP76 genes and 31 pseudogenes (Fig. 1, Supplemental Table 2).

The largest expansion between grape and papaya is in the CYP82 family with two genes in papaya and 34 genes in grape. Of these, papaya CYP82C17 is adjacent to a transducin-like gene just as grape CYP82D19 is adjacent to a transducin-like gene suggesting that these are probable orthologs. Grape CYP82D19 also represents a solo P450 that is not part of a long set of tandem duplications seen in the main block of CYP82D sequences occurring from 9,581 kb to 9,745 kb on chromosome 18. The fact that grape CYP74A1 is located 45 kb downstream from this cluster just as papaya CYP74A12 is located 38 kb downstream of CYP82C17 suggests that the CYP82D loci share either left or right side synteny with the papaya CYP82C17 locus. It also suggests that the CYP82D19 gene and the transducin-like neighbor were split off from the left edge (next to CYP82D14) of the main grape CYP82D gene cluster. We have concluded from this data that grape expanded a single CYP82 gene ancestor shared with papaya to a 16 gene plus 5 pseudogene cluster and that this cluster then split at one end to generate a new CYP82D19 locus.

In contrast to this pattern of duplication and rearrangement, the papaya CYP72 family has all except two of its nine genes and five pseudogenes in one large cluster of about 100 kb (supercontig_77:897282-995688). The two additional genes are on small supercontigs that may join at the ends of the cluster on supercontig_77. The grape CYP72 family has all of its 22 genes and 20 pseudogenes in seven different clusters and at four solo gene locations. This scattered and expanded state of the CYP72 family in grape demonstrates that some species tend to reposition CYP genes to new locations and then expand gene copy numbers at these sites by tandem duplication. A second example of this evolutionary pattern occurs in the CYP81B subfamily with the grape CYP81B genes occurring in two clusters on chromosomes 9 and 18 and the papaya CYP81B19 to CYP81B22 and CYP81K3 occurring in one 36 kb cluster. This location of the CYP81K3 gene suggests that it is derived from one of the CYP81B genes. Three additional examples of these repeated duplications occur in the CYP89A subfamily with 1 gene in papaya and 14 genes in grape, in the CYP71 family with 12 genes in papaya and 25 genes in grape and in the CYP79 family with three genes in papaya and nine genes in grape. The expanded gene families and subfamilies in grape may track evolutionary changes occurring in our historical time frame rather than more prolonged geological time frames. Where visible human-selected changes are apparent in many animal-breeding programs, the CYP duplications and subsequent amino acid variations may open a special window on molecular evolution in domesticated plants. Proof of the effects of domestication on evolution may come from the future sequencing of an uncultivated grape genome such as *Vitis vinifera subsp. Silvestris*, if this truly represents one not tainted by human selection [133].

In contrast to these examples of families and subfamilies that have expanded, 21 subfamilies (CYP71A, 74B, 76G, 77A, 77B, 86B, 90A, 90B, 90C, 90D, 97A, 97B, 97C, 98A, 703A, 704B, 710A, 711A, 720A, 727A, 735A) exist in 1:1 copy numbers in papaya and grape and three subfamilies (CYP85A, 94B, 724A) exist in 2:2 copy numbers. The CYP94C subfamily exists in a 3:3 ratio, the CYP93A subfamily exists in a 4:4 ratio and the CY707A subfamily exists in a 5:5 ratio (Fig. 1, Supplemental Table 2). Many of these have been discussed and are presumed to perform fundamental processes not associated with evolutionary novelty in the rosids.

Conclusions

The sequences of six plant genomes offer many unique opportunities for comparative genomic analysis. Among the

many gene families that might be evaluated, the P450 superfamily presents a special case in that it has been used extensively by plants in their adaptation to the land environment and their subsequent evolution of complex tissues and organs. The existence of CYP proteins mediating many aspects of plant biochemistry and physiology including the formation of lignified structures, signaling compounds, flower and fruit colors, odors and defense compounds provides for many different levels of divergence depending on the extent to which each group of molecules is essential for growth and development.

Four of the six genomes whose P450s have been compared here are in the rosids. Differences between Arabidopsis and papaya highlight the rather broad changes that have taken place in cruciferous plants with two whole genome duplications and biochemical specializations requiring 100 extra CYP sequences in Arabidopsis compared to papaya. The evidence from genome comparisons of grape, papaya and Arabidopsis supports the idea that the 142 CYP genes of papaya have retained many of the ancestral characteristics of rosids that can act as a baseline for evaluation of the grape genome. The expansion of the CYP sequences in grape is particularly clear for those subfamilies that have many tandem duplications evolving from a single gene or a few.

A careful synteny analysis between the CYP subfamilies in grape and papaya has the potential to uncover how time's arrow has expanded the CYPs in grape. Synteny can unambiguously determine the history of gene expansion if flanking genes can be found in both species as in the case of CYP76A11 and CYP76G7 or the CYP82C17 genes in papaya. The story is complicated by duplication to other sites in the genome with loss of synteny as is apparent in the CYP72 family. Comparison of papaya and grape can uncover many of the molecular changes in the grape genome, but to read the true impact of human selection it will be necessary to find and sequence a wild grape genome. Such a project would move molecular evolution of the grape from a geological time scale to a historical time scale.

References

1. Nelson DR (2006) Cytochrome P450 nomenclature 2004. In: Phillips IR, Shephard EA (eds) Methods in Molecular Biology Vol 320, Cytochrome P450 Protocols, 2nd edn. Humana Press Inc., Totowa, NJ, pp 1–10
2. Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of Acorus calamus chloroplast genome and its phylogenetic implications. Mol Biol Evol 22:1813–1822. doi:10.1093/molbev/msi173
3. Cahoon EB, Ripp KG, Hall SE, McGonigle B (2002) Transgenic production of epoxy fatty acids by expression of a cytochrome P450 enzyme from Euphorbia lagascae seed. Plant Physiol 128:615–624. doi:10.1104/pp.010768
4. Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004a) Comparative genomics of *Oryza sativa* and *Arabidopsis thaliana*. Analysis of 727 Cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol 135:756–772. doi:10.1104/pp.104.039826
5. Jaillon O et al (2007) French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467. doi:10.1038/nature06148
6. Velasco R et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2 (12):e1326. doi:10.1371/journal.pone.0001326
7. Kim J, DellaPenna D (2006) Defining the primary route for lutein synthesis in plants: the role of *Arabidopsis* carotenoid beta-ring hydroxylase CYP97A3. Proc Natl Acad Sci USA 103:3474–3479. doi:10.1073/pnas.0511207103
8. Tian L, Musetti V, Kim J, Magallanes-Lundback M, DellaPenna D (2004) The *Arabidopsis* LUT1 locus encodes a member of the cytochrome p450 family that is required for carotenoid epsilon-ring hydroxylation activity. Proc Natl Acad Sci USA 101:402–407. doi:10.1073/pnas.2237237100
9. Kushiro M, Nakano T, Sato K, Yamagishi K, Asami T, Nakano A, Takatsuto S, Fujioka S, Ebizuka Y, Yoshida S (2001) Obtusifoliol 14alpha-demethylase (CYP51) antisense *Arabidopsis* shows slow growth and long life. Biochem Biophys Res Commun 285:98–104. doi:10.1006/bbrc.2001.5122
10. Kim HB, Schaller H, Goh CH, Kwon M, Choe S, An CS, Durst F, Feldmann KA, Feyereisen R (2005) *Arabidopsis* cyp51 mutant shows postembryonic seedling lethality associated with lack of membrane integrity. Plant Physiol 138:2033–2047. doi:10.1104/pp.105.061598
11. Field B, Osbourn AE (2008) Metabolic diversification—independent assembly of operon-like gene clusters in different plants. Science 320:543–547. doi:10.1126/science.1154990
12. Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, Saito S, Oikawa A, Suzuki H, Sakurai N, Shibata D, Wadano A, Sakata K, Ohta D (2006) Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in *Arabidopsis* and tomato. Plant Cell 18:1008–1022. doi:10.1105/tpc.105.036012
13. Lee DS, Nioche P, Hamberg M, Raman CS (2008) Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. Nature 455:363–368. doi:10.1038/nature07307
14. Li L, Chang Z, Pan Z, Fu ZQ, Wang X (2008) Structural insights into the evolutionary paths of oxylipin biosynthetic enzymes. Proc Natl Acad Sci USA 105:13883–13888. doi:10.1073/pnas.0804099105
15. Bate NJ, Sivasankar S, Moxon C, Riley JM, Thompson JE, Rothstein SJ (1998) Molecular characterization of an *Arabidopsis* gene encoding hydroperoxide lyase, a cytochrome P-450 that is wound inducible. Plant Physiol 117:1393–1400. doi:10.1104/pp.117.4.1393
16. Noordermeer MA, Van Dijken AJ, Smeekens SC, Veldink GA, Vliegenthart JF (2000) Characterization of three cloned and expressed 13-hydroperoxide lyase isoenzymes from alfalfa with unusual N-terminal sequences and different enzyme kinetics. Eur J Biochem 267:2473–2482. doi:10.1046/j.1432-1327.2000.01283.x
17. Kuroda H, Oshima T, Kaneda H, Takashio M (2005) Identification and functional analyses of two cDNAs that encode fatty acid 9-/13-hydroperoxide lyase (CYP74C) in rice. Biosci Biotechnol Biochem 69:1545–1554. doi:10.1271/bbb.69.1545
18. Chehab EW, Raman G, Walley JW, Perea JV, Banu G, Theg S, Dehesh K (2006) Rice HYDROPEROXIDE LYASES with unique expression patterns generate distinct aldehyde signatures in *Arabidopsis*. Plant Physiol 141:121–134. doi:10.1104/pp.106.078592

19. Stumpe M, Bode J, Göbel C, Wichard T, Schaaf A, Frank W, Frank M, Reski R, Pohnert G, Feussner I (2006) Biosynthesis of C9-aldehydes in the moss Physcomitrella patens. Biochim Biophys Acta 1761:301–312
20. US patent 6444874: http://www.freepatentsonline.com/6444874.html
21. Koeduka T, Stumpe M, Matsui K, Kajiwara T, Feussner I (2003) Kinetics of barley FA hydroperoxide lyase are modulated by salts and detergents. Lipids 38:1167–1172. doi:10.1007/s11745-003-1175-9
22. Grechkin AN (2002) Hydroperoxide lyase and divinyl ether synthase. Prostaglandins Other Lipid Mediat 68–69:457–470. doi:10.1016/S0090-6980(02)00048-5
23. Benveniste I, Tijet N, Adas F, Philipps G, Salaün JP, Durst F (1998) CYP86A1 from *Arabidopsis* thaliana encodes a cytochrome P450-dependent fatty acid omega-hydroxylase. Biochem Biophys Res Commun 243:688–693. doi:10.1006/bbrc.1998.8156
24. Wellesen K, Durst F, Pinot F, Benveniste I, Nettesheim K, Wisman E, Steiner-Lange S, Saedler H, Yephremov A (2001) Functional analysis of the LACERATA gene of *Arabidopsis* provides evidence for different roles of fatty acid omega -hydroxylation in development. Proc Natl Acad Sci USA 98:9694–9699. doi:10.1073/pnas.171285998
25. Xiao F, Goodwin SM, Xiao Y, Sun Z, Baker D, Tang X, Jenks MA, Zhou JM (2004) *Arabidopsis* CYP86A2 represses Pseudomonas syringae type III genes and is required for cuticle development. EMBO J 23:2903–2913. doi:10.1038/sj.emboj.7600290
26. Duan H, Schuler MA (2005) Differential expression and evolution of the *Arabidopsis* CYP86A subfamily. Plant Physiol 137:1067–1081. doi:10.1104/pp.104.055715
27. Rupasinghe SG, Duan H, Schuler MA (2007) Molecular definitions of fatty acid hydroxylases in *Arabidopsis thaliana*. Proteins 68:279–293. doi:10.1002/prot.21335
28. Greer S, Wen M, Bird D, Wu X, Samuels L, Kunst L, Jetter R (2007) The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of *Arabidopsis*. Plant Physiol 145:653–667. doi:10.1104/pp.107.107300
29. Li Y, Beisson F, Koo AJ, Molina I, Pollard M, Ohlrogge J (2007) Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. Proc Natl Acad Sci USA 104:18339–18344. doi:10.1073/pnas.0706984104
30. Höfer R, Briesen I, Beck M, Pinot F, Schreiber L, Franke R (2008) The *Arabidopsis* cytochrome P450 CYP86A1 encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis. J Exp Bot 59:2347–2360. doi:10.1093/jxb/ern101
31. Olszewski N, Sun TP, Gubler F (2002) Gibberellin signaling: biosynthesis, catabolism, and response pathways. Plant Cell 14 (Suppl):S61–S80
32. Yamaguchi S (2008) Gibberellin metabolism and its regulation. Annu Rev Plant Biol 59:225–251. doi:10.1146/annurev.arplant.59.032607.092804
33. Morant M, Jørgensen K, Schaller H, Pinot F, Møller BL, Werck-Reichhart D, Bak S (2007) CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. Plant Cell 19:1473–1487. doi:10.1105/tpc.106.045948
34. Anastasiou E, Kenz S, Gerstung M, MacLean D, Timmer J, Fleck C, Lenhard M (2007) Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Dev Cell 13:843–856. doi:10.1016/j.devcel.2007.10.001
35. Imaishi H, Matsuo S, Swai E, Ohkawa H (2000) CYP78A1 preferentially expressed in developing inflorescences of Zea mays encoded a cytochrome P450-dependent lauric acid 12-monooxygenase. Biosci Biotechnol Biochem 64:1696–1701. doi:10.1271/bbb.64.1696
36. Miyoshi K, Ahn BO, Kawakatsu T, Ito Y, Itoh J, Nagato Y, Kurata N (2004) PLASTOCHRON1, a timekeeper of leaf initiation in rice, encodes cytochrome P450. Proc Natl Acad Sci USA 101:875–880. doi:10.1073/pnas.2636936100
37. Wang JW, Schwab R, Czech B, Mica E, Weigel D (2008) Dual Effects of miR156-Targeted SPL Genes and CYP78A5/KLUH on Plastochron Length and Organ Size in *Arabidopsis thaliana*. Plant Cell 20:1231–1243. doi:10.1105/tpc.108.058180
38. Jennewein S, Park H, DeJong JM, Long RM, Bollon AP, Croteau RB (2005) Coexpression in yeast of Taxus cytochrome P450 reductase with cytochrome P450 oxygenases involved in Taxol biosynthesis. Biotechnol Bioeng 89:588–598. doi:10.1002/bit.20390
39. Rontein D, Onillon S, Herbette G, Lesot A, Werck-Reichhart D, Sallaud C, Tissier A (2008) CYP725A4 from yew catalyzes complex structural rearrangement of taxa-4(5),11(12)-diene into the cyclic ether 5(12)-oxa-3(11)-cyclotaxane. J Biol Chem 283:6067–6075. doi:10.1074/jbc.M708950200
40. Ro DK, Arimura G, Lau SY, Piers E, Bohlmann J (2005) Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. Proc Natl Acad Sci USA 102:8060–8065. doi:10.1073/pnas.0500825102
41. JGI Why sequence a liverwort? http://www.jgi.doe.gov/sequencing/why/99191.html
42. Ohnishi T, Watanabe B, Sakata K, Mizutani M (2006a) CYP724B2 and CYP90B3 function in the early C-22 hydroxylation steps of brassinosteroid biosynthetic pathway in tomato. Biosci Biotechnol Biochem 70:2071–2080. doi:10.1271/bbb.60034
43. Talon M, Koornneef M, Zeevaart JA (1990) Endogenous gibberellins in *Arabidopsis* thaliana and possible steps blocked in the biosynthetic pathways of the semidwarf ga4 and ga5 mutants. Proc Natl Acad Sci USA 87:7983–7987. doi:10.1073/pnas.87.20.7983
44. Malonek S, Bömke C, Bornberg-Bauer E, Rojas MC, Hedden P, Hopkins P, Tudzynski B (2005) Distribution of gibberellin biosynthetic genes and gibberellin production in the Gibberella fujikuroi species complex. Phytochemistry 66:1296–1311. doi:10.1016/j.phytochem.2005.04.012
45. Siewers V, Smedsgaard J, Tudzynski P (2004) The P450 Monooxygenase BcABA1 Is Essential for Abscisic Acid Biosynthesis in *Botrytis cinerea*. Appl Environ Microbiol 70:3868–3876. doi:10.1128/AEM.70.7.3868-3876.2004
46. Siewers V, Kokkelink L, Smedsgaard J, Tudzynski P (2006) Identification of an abscisic acid gene cluster in the grey mold *Botrytis cinerea*. Appl Environ Microbiol 72:4619–4626. doi:10.1128/AEM.02919-05
47. Boettcher C, Fellermeier M, Boettcher C, Dräger B, Zenk MH (2005) How human neuroblastoma cells make morphine. Proc Natl Acad Sci USA 102:8495–8500. doi:10.1073/pnas.0503244102
48. Crozier A, Yokota T, Bishop GJ, Kamiya Y (2000) Biosynthesis of hormones and elicitors molecules. In: Buchanan BB, Gruissem W, Jones RL (eds) Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, MD, pp 850–929
49. Bishop GJ, Yokota T (2001) Plants steroid hormones, brassinosteroids: current highlights of molecular aspects on their synthesis/metabolism, transport, perception and response. Plant Cell Physiol 42:114–120. doi:10.1093/pcp/pce018
50. Clouse SD (2002) Brassinosteroids: September 30, 2002. The Arabidopsis Book. Rockville, MD: American Society of Plant Biologists. doi: 10.1199/tab.0009 http://www.aspb.org/publications/arabidopsis/

51. Asami T, Nakano T, Fujioka S (2005) Plant brassinosteroid hormones. Vitam Horm 72:479–504. doi:10.1016/S0083-6729(05)72014-8
52. Cytochrome P450 Homepage: http://drnelson.utmem.edu/cytochromeP450.html
53. Finkelstein RR, Rock CD (2002) Brassinosteroids: September 30, 2002. The Arabidopsis Book. Rockville, MD: American Society of Plant Biologists. doi:10.1199/tab.0058 http://www.aspb.org/publications/arabidopsis/
54. Kushiro T, Okamoto M, Nakabayashi K, Yamagishi K, Kitamura S, Asami T, Hirai N, Koshiba T, Kamiya Y, Nambara E (2004) The *Arabidopsis* cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism. EMBO J 23:1647–1656. doi:10.1038/sj.emboj.7600121
55. Saito S, Hirai N, Matsumoto C, Ohigashi H, Ohta D, Sakata K, Mizutani M (2004) *Arabidopsis* CYP707As encode (+)-abscisic acid 8'-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. Plant Physiol 134:1439–1449. doi:10.1104/pp.103.037614
56. Ma H, Zhang S, Ji L, Zhu H, Yang S, Fang X, Yang R (2006) Fine mapping and in silico isolation of the EUI1 gene controlling upper internode elongation in rice. Plant Mol Biol 60:87–94. doi:10.1007/s11103-005-2762-5
57. Luo A, Qian Q, Yin H, Liu X, Yin C, Lan Y, Tang J, Tang Z, Cao S, Wang X, Xia K, Fu X, Luo D, Chu C (2006) EUI1, encoding a putative cytochrome P450 monooxygenase, regulates internode elongation by modulating gibberellin responses in rice. Plant Cell Physiol 47:181–191. doi:10.1093/pcp/pci233
58. Zhu Y, Nomura T, Xu Y, Zhang Y, Peng Y, Mao B, Hanada A, Zhou H, Wang R, Li P, Zhu X, Mander LN, Kamiya Y, Yamaguchi S, He Z (2006) ELONGATED UPPERMOST INTERNODE encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. Plant Cell 18:442–456. doi:10.1105/tpc.105.038455
59. Plant hormones http://www.plant-hormones.info/occurrence_of_gas_in_plants.htm
60. Neff MM, Nguyen SM, Malancharuvil EJ, Fujioka S, Noguchi T, Seto H, Tsubuki M, Honda T, Takatsuto S, Yoshida S, Chory J (1999) BAS1: A gene regulating brassinosteroid levels and light responsiveness in *Arabidopsis*. Proc Natl Acad Sci USA 96:15316–15323. doi:10.1073/pnas.96.26.15316
61. Nakamura M, Satoh T, Tanaka S, Mochizuki N, Yokota T, Nagatani A (2005) Activation of the cytochrome P450 gene, CYP72C1, reduces the levels of active brassinosteroids in vivo. J Exp Bot 56:833–840. doi:10.1093/jxb/eri073
62. Takahashi N, Nakazawa M, Shibata K, Yokota T, Ishikawa A, Suzuki K, Kawashima M, Ichikawa T, Shimada H, Matsui M (2005) shk1-D, a dwarf *Arabidopsis* mutant caused by activation of the CYP72C1 gene, has altered brassinosteroid levels. Plant J 42:13–22. doi:10.1111/j.1365-313X.2005.02357.x
63. Turk EM, Fujioka S, Seto H, Shimada Y, Takatsuto S, Yoshida S, Wang H, Torres QI, Ward JM, Murthy G, Zhang J, Walker JC, Neff MM (2005) BAS1 and SOB7 act redundantly to modulate *Arabidopsis* photomorphogenesis via unique brassinosteroid inactivation mechanisms. Plant J 42:23–34. doi:10.1111/j.1365-313X.2005.02358.x
64. Ohnishi T, Nomura T, Watanabe B, Ohta D, Yokota T, Miyagawa H, Sakata K, Mizutani M (2006b) Tomato cytochrome P450 CYP734A7 functions in brassinosteroid catabolism. Phytochemistry 67:1895–1906. doi:10.1016/j.phytochem.2006.05.042
65. Lazar G, Goodman HM (2006) MAX1, a regulator of the flavonoid pathway, controls vegetative axillary bud outgrowth in *Arabidopsis*. Proc Natl Acad Sci USA 103:472–476. doi:10.1073/pnas.0509463102
66. Bennett T, Sieberer T, Willett B, Booker J, Luschnig C, Leyser O (2006) The *Arabidopsis* MAX pathway controls shoot branching by regulating auxin transport. Curr Biol 16:553–563. doi:10.1016/j.cub.2006.01.058
67. Stirnberg P, Furner IJ, Ottoline Leyser HM (2007) MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. Plant J 50:80–94. doi:10.1111/j.1365-313X.2007.03032.x
68. Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analysis of amino acids and coding nucleotide sequences. J Biol Chem 267:83–90
69. Kieber JJ (2002) Cytokinins: March 27, 2002. The Arabidopsis Book. Rockville, MD: American Society of Plant Biologists. doi:10.1199/tab.0063 http://www.aspb.org/publications/arabidopsis/
70. Takei K, Yamaya T, Sakakibara H (2004) *Arabidopsis* CYP735A1 and CYP735A2 encode cytokinin hydroxylases that catalyze the biosynthesis of trans-Zeatin. J Biol Chem 279:41866–41872. doi:10.1074/jbc.M406337200
71. Weng JK, Li X, Stout J, Chapple C (2008) Independent origins of syringyl lignin in vascular plants. Proc Natl Acad Sci USA 105:7887–7892. doi:10.1073/pnas.0801696105
72. Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S, Adam-Blondon AF, Testolin R (2006) Colour variation in red grapevines (*Vitis vinifera L.*): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3',5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. BMC Genomics 7:12. doi:10.1186/1471-2164-7-12
73. Katsumoto Y, Fukuchi-Mizutani M, Fukui Y, Brugliera F, Holton TA, Karan M, Nakamura N, Yonekura-Sakakibara K, Togami J, Pigeaire A, Tao GQ, Nehra NS, Lu CY, Dyson BK, Tsuda S, Ashikari T, Kusumi T, Mason JG, Tanaka Y (2007) Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin. Plant Cell Physiol 48:1589–1600. doi:10.1093/pcp/pcm131
74. Dixon RA, Paiva N (1995) Stress-induced phenylpropanoid metabolism. Plant Cell 7:1085–1097
75. Croteau R, Kutchan TM, Lewis NG (2000) Natural products (secondary metabolites). In: Buchanan BB, Gruissem W, Jones RL (eds) Biochemistry & Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, MD, pp 1250–1318
76. Akashi T, Aoki T, Ayabe S (1998) CYP81E1, a cytochrome P450 cDNA of licorice (*Glycyrrhiza echinata L.),* encodes isoflavone 2'-hydroxylase. Biochem Biophys Res Commun 251:67–70. doi:10.1006/bbrc.1998.9414
77. Overkamp S, Hein F, Barz W (2000) Cloning and characterization of eight cytochrome P450 cDNAs from chickpea (*Cicer arietinum L.*) cell suspension cultures. Plant Sci 155:101–108. doi:10.1016/S0168-9452(00)00214-4
78. Liu CJ, Huhman D, Sumner LW, Dixon RA (2003) Regiospecific hydroxylation of isoflavones by cytochrome p450 81E enzymes from *Medicago truncatula*. Plant J 36:471–484. doi:10.1046/j.1365-313X.2003.01893.x
79. Schopfer CR, Kochs G, Lottspeich F, Ebel J (1998) Molecular characterization and functional expression of dihydroxypterocarpan 6a-hydroxylase, an enzyme specific for pterocarpanoid phytoalexin biosynthesis in soybean (Glycine max L.). FEBS Lett 432:182–186. doi:10.1016/S0014-5793(98)00866-7
80. Martens S, Forkmann G (1999) Cloning and expression of flavone synthase II from Gerbera hybrids. Plant J 20:611–618. doi:10.1046/j.1365-313X.1999.00636.x
81. Akashi T, Aoki T, Ayabe S (1999) Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. Plant Physiol 121:821–828. doi:10.1104/pp.121.3.821

82. Jung W, Yu O, Lau S-MC, O'Keefe DP, Odell J, Fader G, McGonigle B (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat Biotechnol 18:208–212. doi:10.1038/72671
83. Sawada Y, Kinoshita K, Akashi T, Aoki T, Ayabe SI (2002) Key amino acid residues required for aryl migration catalysed by the cyrochrome P450 2-hydroxyisoflavanone synthase. Plant J 31:555–564. doi:10.1046/j.1365-313X.2002.01378.x
84. Kim BG, Kim SY, Song HS, Lee C, Hur HG, Kim SI, Ahn JH (2003) Cloning and expression of the isoflavone synthase gene (IFS-Tp) from *Trifolium pratense*. Mol Cells 15:301–306
85. Shibuya M, Hoshino M, Katsube Y, Hayashi H, Kushiro T, Ebizuka Y (2006) Identification of beta-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J 273:948–959. doi:10.1111/j.1742-4658.2006.05120.x
86. Latunde-Dada AO, Cabello-Hurtado F, Czittrich N, Didierjean L, Schopfer C, Hertkorn N, Werck-Reichhart D, Ebel J (2001) Flavonoid 6-hydroxylase from soybean (Glycine max L.), a novel plant P-450 monooxygenase. J Biol Chem 276:1688–1695
87. Luo P, Wang YH, Wang GD, Essenberg M, Chen XY (2001) Molecular cloning and functional identification of (+)-delta-cadinene-8-hydroxylase, a cytochrome P450 mono-oxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. Plant J 28:95–104. doi:10.1046/j.1365-313X.2001.01133.x
88. Haudenschild C, Schalk M, Karp F, Croteau R (2000) Functional expression of regiospecific cytochrome P450 limonene hydroxylases from Mint (*Mentha spp*.) in *Escherichia coli* and *Saccharomyces cerevisiae*. Arch Biochem Biophys 379:127–136. doi:10.1006/abbi.2000.1864
89. Schalk M, Croteau R (2000) A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase. Proc Natl Acad Sci USA 97:11948–11953. doi:10.1073/pnas.97.22.11948
90. Halkier BA (1999) Glucosinolates. In: Ikan R (ed) Naturally Occurring Glucosides. Wiley & Sons Ltd., New York, pp 193–223
91. Hull AK, Vij R, Celenza JL (2000) *Arabidopsis* cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. Proc Natl Acad Sci USA 97:2379–2384. doi:10.1073/pnas.040569997
92. Mikkelsen MD, Hansen CH, Wittstock U, Halkier BA (2000) Cytochrome P450 CYP79B2 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. J Biol Chem 275:33712–33717. doi:10.1074/jbc.M001667200
93. Bak S, Feyereisen R (2001) The involvement of two p450 enzymes, CYP83B1 and CYP83A1, in auxin homeostasis and glucosinolate biosynthesis. Plant Physiol 127:108–118. doi:10.1104/pp.127.1.108
94. Bak S, Tax FE, Feldmann KA, Galbraith DW, Feyereisen R (2001) CYP83B1, a cytochrome P450 at the metabolic branch point in auxin and indole glucosinolate biosynthesis in *Arabidopsis*. Plant Cell 13:101–111
95. Naur P, Petersen BL, Mikkelsen MD, Bak S, Rasmussen H, Olsen CE, Halkier BA (2003) CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in *Arabidopsis*. Plant Physiol 133:63–72. doi:10.1104/pp.102.019240
96. Hansen CH, Wittstock U, Olsen CE, Hick AJ, Pickett JA, Halkier BA (2001) Cytochrome p450 CYP79F1 from *Arabidopsis* catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. J Biol Chem 276:11078–11085. doi:10.1074/jbc.M010123200
97. Reintanz B, Lehnen M, Reichelt M, Gershenzon J, Kowalczyk M, Sandberg G, Godde M, Uhl R, Palme K (2001) Bus, a bushy *Arabidopsis* CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. Plant Cell 13:351–367
98. Chen S, Glawischnig E, Jørgensen K, Naur P, Jørgensen B, Olsen CE, Hansen CH, Rasmussen H, Pickett JA, Halkier BA (2003) CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. Plant J 33:923–937. doi:10.1046/j.1365-313X.2003.01679.x
99. Hemm MR, Ruegger MO, Chapple C (2003) The *Arabidopsis* ref2 mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. Plant Cell 15:179–194. doi:10.1105/tpc.006544
100. Bak S, Paquette SM, Morant M, Morant AV, Saito S, Bjarnholt N, Zagrobelny M, Jørgensen K, Hamann T, Osmani S, Simonsen HT, Pérez RS, van Heeswijck TB, Jørgensen B, Møller BL (2006) Cyanogenic glucosides: a case study for evolution and application of cytochromes P450. Phytochem Rev 5:309–329. doi:10.1007/s11101-006-9033-1
101. Nielsen KA, Tattersall DB, Jones PR, Møller BL (2008) Metabolon formation in dhurrin biosynthesis. Phytochemistry 69:88–98. doi:10.1016/j.phytochem.2007.06.033
102. Irmler S, Schröder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, Strack D, Matern U, Schröder J (2000) Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. Plant J 24:797–804. doi:10.1046/j.1365-313x.2000.00922.x
103. Schröder G, Unterbusch E, Kaltenbach M, Schmidt J, Strack D, De Luca V, Schröder J (1999) Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. FEBS Lett 458:97–102. doi:10.1016/S0014-5793(99)01138-2
104. Collu G, Unver N, Peltenburg-Looman AM, van der Heijden R, Verpoorte R, Memelink J (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. FEBS Lett 508:215–220. doi:10.1016/S0014-5793(01)03045-9
105. Tamaki K, Imaishi H, Ohkawa H, Oono K, Sugimoto M (2005) Cloning, expression in yeast, and functional characterization of CYP76A4, a novel cytochrome P450 of petunia that catalyzes (omega-1)-hydroxylation of lauric acid. Biosci Biotechnol Biochem 69:406–409. doi:10.1271/bbb.69.406
106. Imaishi H, Petkova-Andonova M (2007) Molecular cloning of CYP76B9, a cytochrome P450 from Petunia hybrida, catalyzing the omega-hydroxylation of capric acid and lauric acid. Biosci Biotechnol Biochem 71:104–113. doi:10.1271/bbb.60396
107. Ralston L, Kwon ST, Schoenbeck M, Ralston J, Schenk DJ, Coates RM, Chappell J (2001) Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from tobacco (*Nicotiana tabacum*). Arch Biochem Biophys 393:222–235. doi:10.1006/abbi.2001.2483
108. Kandel S, Morant M, Benveniste I, Blée E, Werck-Reichhart D, Pinot F (2005) Cloning, functional expression, and characterization of CYP709C1, the first sub-terminal hydroxylase of long chain fatty acid in plants. Induction by chemicals and methyl jasmonate. J Biol Chem 280:35881–35889. doi:10.1074/jbc.M500918200
109. Pauli HH, Kutchan TM (1998) Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3'-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. Plant J 13:793–801. doi:10.1046/j.1365-313X.1998.00085.x
110. Apuya NR, Park JH, Zhang L, Ahyow M, Davidow P, Van Fleet J, Rarang JC, Hippley M, Johnson TW, Yoo HD, Trieu A, Krueger S, Wu CY, Lu YP, Flavell RB, Bobzin SC (2008) Enhancement of alkaloid production in opium and California poppy by transactivation using heterologous regulatory factors. Plant Biotechnol J 6:160–175. doi:10.1111/j.1467-7652.2007.00302.x

111. Kraus PF, Kutchan TM (1995) Molecular cloning and heterologous expression of a cDNA encoding berbamunine synthase, a C–O phenol-coupling cytochrome P450 from the higher plant *Berberis stolonifera*. Proc Natl Acad Sci USA 92:2071–2075. doi:10.1073/pnas.92.6.2071
112. Li R, Reed DW, Liu E, Nowak J, Pelcher LE, Page JE, Covello PS (2006) Functional genomic analysis of alkaloid biosynthesis in Hyoscyamus niger reveals a cytochrome P450 involved in littorine rearrangement. Chem Biol 13:513–520. doi:10.1016/j.chembiol.2006.03.005
113. Ikezawa N, Iwasa K, Sato F (2008) Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C-C phenol coupling of (S)-reticuline in magnoflorine biosynthesis, from cultured Coptis japonica cells. J Biol Chem 283:8810–8821. doi:10.1074/jbc.M705082200
114. Minami H, Kim JS, Ikezawa N, Takemura T, Katayama T, Kumagai H, Sato F (2008) Microbial production of plant benzylisoquinoline alkaloids. Proc Natl Acad Sci USA 105:7393–7398. doi:10.1073/pnas.0802981105
115. Gavilano LB, Siminszky B (2007) Isolation and characterization of the cytochrome P450 gene CYP82E5v2 that mediates nicotine to nornicotine conversion in the green leaves of tobacco. Plant Cell Physiol 48:1567–1574. doi:10.1093/pcp/pcm128
116. Chakrabarti M, Bowen SW, Coleman NP, Meekins KM, Dewey RE, Siminszky B (2008) CYP82E4-mediated nicotine to nornicotine conversion in tobacco is regulated by a senescence-specific signaling pathway. Plant Mol Biol 66:415–427. doi:10.1007/s11103-007-9280-6
117. Kruse T, Ho K, Yoo HD, Johnson T, Hippely M, Park JH, Flavell R, Bobzin S (2008) In planta biocatalysis screen of P450s identifies 8-methoxypsoralen as a substrate for the CYP82C subfamily, yielding original chemical structures. Chem Biol 15:149–156. doi:10.1016/j.chembiol.2008.01.008
118. Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, Wain HM, Nebert DW (2004b) Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics 14:1–18. doi:10.1097/00008571-200401000-00001
119. Lodeiro S, Xiong Q, Wilson WK, Kolesnikova MD, Onak CS, Matsuda SP (2007) An oxidosqualene cyclase makes numerous products by diverse mechanisms: a challenge to prevailing concepts of triterpene biosynthesis. J Am Chem Soc 129:11213–11222. doi:10.1021/ja073133u
120. Husselstein-Muller T, Schaller H, Benveniste P (2001) Molecular cloning and expression in yeast of 2,3-oxidosqualene-triterpenoid cyclases from *Arabidopsis thaliana*. Plant Mol Biol 45:75–92. doi:10.1023/A:1006476123930
121. Shimura K, Okada A, Okada K, Jikumaru Y, Ko KW, Toyomasu T, Sassa T, Hasegawa M, Kodama O, Shibuya N, Koga J, Nojiri H, Yamane H (2007) Identification of a biosynthetic gene cluster in rice for momilactones. J Biol Chem 282:34013–34018. doi:10.1074/jbc.M703344200
122. Otomo K, Kanno Y, Motegi A, Kenmoku H, Yamane H, Mitsuhashi W, Oikawa H, Toshima H, Itoh H, Matsuoka M, Sassa T, Toyomasu T (2004) Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and oryzalexins A-F in rice. Biosci Biotechnol Biochem 68:2001–2006. doi:10.1271/bbb.68.2001
123. Perrin RM, Fedorova ND, Bok JW, Cramer RA, Wortman JR, Kim HS, Nierman WC, Keller NP (2007) Transcriptional regulation of chemical diversity in Aspergillus fumigatus by LaeA. PLoS Pathog 3(4):e50. doi:10.1371/journal.ppat.0030050
124. Kelly DE, Kraševec N, Mullins J, Nelson DR (2008) The CYPome (Cytochrome P450 complement) of Aspergillus nidulans. Fungal Genet Biol 2008(Sep):12. Epub ahead of print.
125. Meyer K, Cusumano J, Somerville C, Chapple C (1996) Ferulate-5-hydroxylase from *Arabidopsis* thaliana defines a new family of cytochrome P450-dependent monoxygenases. Proc Natl Acad Sci USA 93:6869–6874. doi:10.1073/pnas.93.14.6869
126. Ruegger M, Meyer K, Cusumano JC, Chapple C (1999) Regulation of ferulate-5-hydroxylase expression in *Arabidopsis* in the context of sinapate ester biosynthesis. Plant Physiol 119:101–110. doi:10.1104/pp.119.1.101
127. Humphreys JM, Hemm MR, Chapple C (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. Proc Natl Acad Sci USA 96:10045–10050. doi:10.1073/pnas.96.18.10045
128. Bozak KR, Yu H, Sirevåg R, Christoffersen RE (1990) Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proc Natl Acad Sci USA 87:3904–3908. doi:10.1073/pnas.87.10.3904
129. Pua EC, Lee YC (2003) Expression of a ripening-related cytochrome P450 cDNA in Cavendish banana (*Musa acuminata cv. Williams*). Gene 305:133–140. doi:10.1016/S0378-1119(02)01237-4
130. Aharoni A, Giri AP, Verstappen FW, Bertea CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell 16:3110–3131. doi:10.1105/tpc.104.023895
131. Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H (2006) Phylogenetic analyses of Vitis (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol 6:32. doi:10.1186/1471-2148-6-32
132. Lyons E, Pedersen B, Kane J, Freeling M (2008) The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. Trop Plant Biol. doi:10.1007/s12042-008-9017-y.
133. This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. Trends Genet 22:511–519. doi:10.1016/j.tig.2006.07.008
134. Zhu XY, Chase MW, Qiu YL, Kong HZ, Dilcher DL, Li JH, Chen ZD (2007) Mitochondrial matR sequences help to resolve deep phylogenetic relationships in rosids. BMC Evol Biol 7:217. doi:10.1186/1471-2148-7-217
135. Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on matK sequence information. Am J Bot 90:1758–1776. doi:10.3732/ajb.90.12.1758