Plant Cytochrome P450s from Moss to Poplar

David R. Nelson

The University of Tennessee Dept. of Molecular Sciences

and the Center of Excellence for Genomics and Bioinformatics

Memphis, TN 38163

dnelson@utmem.edu

key words cytochrome P450, pathway, evolution, plants, moss, Populus

DOI 10.1007/s11101-006-9015-3

Summary

This review represents the first attempt to define the origins of the major P450-containing pathways in plants. Comparative genomics with five complete P450 gene sets from *Chlamydomonas reinhardtii* with 39 sequences, *Physcomitrella patens* (moss) with 71 sequences, rice with 356 sequences, Arabidopsis with 246 sequences and Populus with 312 sequences is used to estimate how old each gene family is and to identify the most ancient P450s and their pathways. The pathways included are the phenylpropanoid and lignin pathways, the gibberellin pathway, the oxylipin/jasmonate pathway, the basic flavonoid pathway, the brassinosteroid pathway, the abscisic acid pathway and the cutin synthesis pathway. An effort is made to identify at least some examples of P450s that have emerged at many different levels of the evolutionary bush, from the base to the tips.

Introduction

Genome sequencing projects are rapidly expanding outward from the first model organisms in each Kingdom of life. From the first bacterial genome in 1995, there are now more than 360 complete bacterial genomes and more than 25 archaeal genomes (1). The Joint Genome Institute sequenced 15 bacterial genomes in October 2000 (Microbial Month) to demonstrate a sequencing rate of one bacterial genome every working day and a half. The Venter Institute has announced plans to sequence 100 marine microbial genomes in one year (2) and this has been increased to 150 bacterial and 5 archaeal genomes (3). Fungal genomes are also accumulating with the goal of comparative genomics of highly similar genomes, as in six species of Saccharomyces: *S. cerevisiae, S. mikatae, S. kudriavzevii , S. bayanus, S. castellii,* and *S. kluyveri* (4,5). Additional disparate fungal genomes are being sequenced to cover a broader phylogenetic range. The National Human Genome Research Institute NHGRI is funding three fungal genome initiatives to sequence 18 species (6), of which 16 are already assembled and released as of June 2006 (7, 8). Numerous non-fungal eukaryotic microbes, mostly pathogens like *Giardia* (9), *Plasmodium* (10), *Trypanosoma* (11) and *Toxoplasma* (12) are completed or well underway.

Animals have larger genomes and are more difficult to assemble due to the numerous highly repeated sequences that they contain. There has been emphasis on mammals, with human leading the way, followed by mouse, rat, chimp, dog, cow and pig. Many other mammals will follow. Already (June 2006), there are 41 million sequence traces for the laboratory opossum *Monodelphis domestica,* 29 million for the duck-billed platypus *Ornithorhynchus anatinus*, 12 million for purple sea urchin

Strongylocentrotus purpuratus, 10 million for African savanna elephant *Loxodonta africana*, and 9 million for the nine-banded armadillo *Dasypus novemcinctus* at the NCBI Trace Archive (13) Rabbit, European common shrew and the Lesser hedgehog or tenrec have two-fold genome sequence coverage completed. Two species of pufferfish are sequenced and zebrafish assembly Zv5 contains 1.63 Gb of about 1.8 Gb in the genome (90%). Purple sea urchin has been sequenced and is assembled at the UCSC genome browser (14) and at Baylor College of Medicine (15). Two species of tunicate or sea squirts are now available as well as 12 species of Drosophila (16), two species of mosquitos, two species of roundworms, and simpler animals like hydra. The NHGRI maintains a web list of proposed sequencing projects, their white papers and current status (7).

Information on plant genomes has been more slowly collected. Arabidopsis and two strains of rice have been sequenced, with Arabidopsis and the japonica rice strain reaching a highly polished "gold standard" of completion and assembly (17,18). These genomes are now moving into the post-genomic or functional genomic era focusing on annotation of the 20-25 thousand genes that exist in these species (19-21). Recently, *Populus trichocarpa* (black cottonwood) became the first tree to have its genome sequenced (22). At the other end of the plant spectrum, the moss *Physcomitrella patens* is being sequenced with over 7 million traces in the Trace Archive (23). Sequence data on Gemmiferous spike moss (*Selaginella moellendorffii*) is now released to the Trace Archive with 1.9 million reads or about 10X coverage for this small plant genome (24), These genomes should make it possible to follow plant evolution by genome comparison from *Chlamydomonas reinhardtii* (25) to an angiosperm tree. Even so, there will be

significant gaps in this evolutionary progression, since some plant genomes are prohibitively large for complete sequencing. Ferns have a genome size of about 160 Gb compared to humans at a paltry 3 Gb. Even with these large genomes, EST sampling, as has already been done with pine and spruce (26-28), can provide some measure of the range of transcripts expressed *in vivo*. In fact, 20 of the top 50 organisms listed in the ESTdb at Genbank are plants (29).

Plants adaptation to land, the development of water barriers, vascular systems, and lignin for structural support, and their production of long range signaling mechanisms for defense responses, flower color and fragrance for attraction of pollinators, defense toxins to repel herbivores and pathogens have all required new biochemistries and novel pathways. Due to the usefulness of oxygen in building complex molecules, cytochrome P450 enzymes are often found in these pathways. In fact they are extremely abundant, comprising about 1% of the genes in some plant genomes such as Arabidopsis, rice and poplar. At least a part of plant evolution can be tied to the appearance of these novel pathways and in turn, to the emergence of P450s that are essential to these pathways. This article compares whole genome sets of cytochrome P450s or CYP genes from the moss to rice, Arabidopsis and poplar to begin the process of identifying the origins of the P450s and thus the pathways themselves. The Cytochrome P450 Homepage (30) has information on the complete gene sets.

Chlamydomonas as a beginning point

The green alga Chlamydomonas has as long an evolutionary history as any living member of the Viridiplantae or green plants. However, it has retained a single-celled,

aquatic lifestyle and is considered to be a good model of the clade founder. Analysis of the CYP genes in Chlamydomonas finds 39 genes (31) . These 39 genes are distributed in 21 families according to the CYP nomenclature. Only four of these families had been seen previously. CYP51, which is the sterol 14α-demethylase found in most eukaryotes, exists in the green alga as expected. CYP97 is found in Chlamydomonas, pine, rice, Arabidopsis and poplar with three orthologous subfamilies in each species. Notably, each subfamily has only one member suggesting a unique highly conserved function for each. Of the individual genes within this family CYP97C1 of Arabidopsis is a carotenoid ε-hydroxylase involved in the formation of lutein from α-carotene (32). CYP97A3 is a carotenoid β-hydroxylase (33). CYP97B also is likely to metabolize carotenoids, a function that was present in algae before plants came onto the land. Another subfamily CYP97E is found in the diatoms *Skeletonema costatum* and *Thalisossira pseudonana*. CYP710 exists in Arabidopsis and rice as multiple genes, and in poplar as a single gene in the CYP710A subfamily. Chlamydomonas has only one gene CYP710B1. The different subfamily designation is due to sequence divergence, but this is the probable ortholog to the CYP710A sequences. Comparison to other eukaryotic P450s shows that CYP710 belongs to the CYP61 clan found in the ergosterol pathway of all fungi. In fact some fungi only have two CYPs, CYP51 and CYP61, with CYP61 serving as a C-22 sterol desaturase that acts after CYP51 in ergosterol biosynthesis. CYP710A1 in Arabidopsis, is also a C-22 sterol desaturase acting on β-sitosterol (34). The last conserved family in Chlamydomonas is CYP55, a fungal P450 family. CYP55 is nitric oxide reductase in some fungi (35). CYP55 is unique among eukaryotic P450s in that it is a soluble P450 not associated with any membrane or subcellular structures. Sequence

comparisons to other CYPs show a high similarity to a bacterial family CYP105 from Streptomyces and this probably represents the origin of this family via lateral gene transfer to fungi. Since CYP55 is not seen in land plants, it is likely to have been acquired in algae from fungi after land plants diverged from their common ancestor. The algal CYP55B1 shares one intron-exon boundary with CYP55A6 of *Neurospora crassa*, so it had to share a common ancestor with the fungal gene and not the CYP105s of bacteria.

All plant CYP families except CYP51, CYP97, CYP710 and CYP746 are new in land plants.

Before the moss genome was sequenced, all known plant P450s fell into 63 families (36). Annotation of moss adds 16 new families that are not radically different from other plant CYPs. All 71 moss CYPs except one belong in the 10 existing plant CYP clans (Fig. 1). One sequence, which is ambiguous and sorts with the CYP72 clan or the CYP97 clan, has been given its own clan. This sequence, CYP746B1, is 51% identical to the Chlamydomonas sequence CYP746A1.

All plant clans except CYP711 are found in moss. The fact that several CYP711-like sequences are seen in Chlamydomonas (CYP743 and CYP744), suggest that these genes may have been lost in moss. This implies that all the major innovations of sequence at the CYP clan level were already found in moss but not in algae. There must have been a burst of evolution in the interval between emergence on land and arrival at the moss stage of complexity. During this period the CYP71, CYP74, CYP86 and CYP727 clans evolved. Among present-day angiosperms, the CYP71 clan is by far the largest, with

one-third to one-half of the plant P450s in any one genome. The CYP71 clan is also the largest clan in moss with 41 sequences. This fact suggests that CYP71 was one of the earliest plant CYP clans, aside from the solo-function CYPs like CYP51, CYP97A, CYP97B, CYP97C, and CYP710. Even these genes are showing recent signs of expansion, though it may only be for expression of the same function in different tissues as seen with CYP710A1, A2, A3 and A4 in Arabidopsis. These four highly similar CYP710 genes occur in two clusters of two genes each. According to the P450 tissue profiling data at the Functional Genomics of Arabidopsis P450s site (37), CYP710A1 is more highly expressed in root than in other tissues, while CYP710A2 is more evenly expressed in shoot, root and stem. However, Poplar has only one CYP710 gene, so a single function may still be the norm for CYP710, with a recent series of duplications in Arabidopsis . The existence of CYP51G1 (CYP51A2) and CYP51G2 (CYP51A1) in Arabidopsis suggested two CYP51 genes had evolved in this species, but recently, the CYP51G2 sequence was shown to be an expressed pseudogene and thus non-functional (38). Rice has eight full length CYP51H genes and these presumably have taken on some new functions. Arabidopsis has only one CYP711 gene, but Poplar has two (due to a very recent genome duplication) and rice has five. The single CYP711 gene in Arabidospsis suggests a single conserved function. Recently, CYP711 has been identified as MAX1, a gene making a carotenoid-derived branch-inhibiting hormone (39). Other than matches to CYP711 members, CYP711 has its best BLAST hits to CYP5, thromboxane A_2 synthase in vertebrates. CYP5 is a member of the CYP3 clan in animals that includes CYP3 and CYP5 in vertebrates and CYP6 and CYP9 in insects. CYP711 may share a common ancestor with the CYP3 clan.

Emergence of ancient P450-containing pathways.

Moss has 71 CYPs (Fig. 1) with more than half in the CYP71 clan, sometimes referred to as the plant group A P450s (center of Fig. 1). This expansion of the CYP71 clan sequences is noted in other plant genomes and seems to have begun early in land plants (40). However, most of the plant families in the CYP71 clan are not the same in moss and angiosperms or pine. The five CYP71 clan families that are clearly present in moss and other plants including Arabidopsis, rice and pine are CYP73, CYP78, CYP98, CYP701 and CYP703. Two of these CYPs have well defined roles in the phenylpropanoid pathway: CYP73 is cinnamate 4-hydroxylase (C4H) in the core pathway (41), and CYP98 is coumaryl shikimic acid 3'-hydroxylase (C3H), in the branch pathway forming G-lignols (42,43). Moss which is not a woody plant lacks CYP84 (F5H) for making S-lignol monomers in higher plants (44). The absence of this gene in the moss genome suggests that the evolution of the lignin synthetic pathway is only partly complete in the moss genome, or less likely, that other enzymes fulfill this role in moss. CYP701 is another interesting case. Another interesting example of the limited pathway evolution in moss is CYP701 that is the ent-kaurene oxidase enzyme at the beginning of the gibberellin biosynthetic pathway (45). In the complete pathway found in higher plants, there is a kaurenoic acid oxidase CYP88, that forms GA_{12} from ent-kaurenoic acid (46,47) but this enzyme is missing in moss. Again, moss is showing the first part of a pathway for molecules critical for higher plants.

Gibberellins have been shown to affect sex determination in some fern gametophytes in schizaeaceous taxa (Anemia, Lygodium) but not in non-schizaeaceous

taxa (Ceratopteris, Dryopteris, Polypodium) (48). GA_9 methyl ester has been found in the fern *Lygodium japonicum*, where it acts as an antheridiogen (49). The GA_9 structure is similar to the GA_{12} structure and probably derived from GA_{12}. That would mean that CYP701 and CYP88 would be present in Lygodium and probably other schizaeaceous fern taxa. Therefore, CYP88 must have been present in the last common ancestor of ferns and higher plants. The evolution of CYP88 is a prerequisite for gibberellin GA_{12} biosynthesis in green plants, but not by the independent pathway of *Gibberella fujikuroi* (50). As such CYP88 becomes a clade marker or synapomorphy for this important plant signaling precursor.

CYP78, also in the CYP71 clan, is not as well characterized except for CYP78A11 which is the Plastochron1 gene in rice that controls the timing of lateral organ formation from the apical meristem (51) and CYP78A1 in *Zea mays* which is a lauric acid ω-hydroxylase expressed in developing inflorescences (52). The pathway for these important developmental roles have not yet been determined, but the existence of related genes in moss suggests that it is an ancient plant pathway possibly involving fatty acid hydroxylations. Another family within this clan, CYP703, is also expressed in flowers and capable of hydroxylating fatty acids (53). The cellular function of CYP703 is not known, though it is plainly conserved in moss (Fig. 1).

The CYP72 clan has at least two members CYP734A1 and CYP72C1 that inactivate brassinolide (54). One might expect parallel evolution of these sequences with the CYP85 and CYP90 families that are involved in synthesizing brassinolide (55-57). Although there are no clear CYP72 or CYP734 family sequences in moss, three moss sequences in the CYP766 family cluster with CYP72A8. These three sequences are

placed in the CYP72 clan. Recently, a third CYP72 clan member, CYP714D1 has been identified as EUI1 (elongation of uppermost internode 1) in rice and suggested to have a role in inactivating GA signalling in the growing rice plant. A mutation in this gene led to excessive levels of an active form of GA in the uppermost internode (58). This result has linked the CYP72 clan to inactivation of both brassinosteroids and gibberellins. Blast searches in fern ESTs identified a sequence 45% identical to CYP72A18 over 276 amino acids (CV735871 Ceratopteris) showing that the CYP72 family probably exists in ferns. The CYP72 clan is diverse, with eight established families and two new families in moss, CYP765 and CYP766. Ferns have at least one more EST with 53% identity to CYP709C9 (BP914549 *Adiantum capillus-veneris*) indicating the presence of at least two known CYP72 clan families in ferns with probable precursors in the last common ancestor of ferns and moss.

The CYP74 family is important in modifying unsaturated fatty acid hydroperoxides derived from linoleic or α-linolenic acids and includes at least three classes of enzymes, allene oxide synthases, hydroperoxide lyases and divinyl ether synthases. The CYP74A subfamily contains the allene oxide synthases in the oxylipin pathway to jasmonic acid and methyl jasmonate, that act as signals in plant defense pathways. These products have been likened to the arachidonic acid metabolites found in animals (59). The CYP74A subfamily first appears in phylogenetic trees in moss and has related sequences in gymnosperms (*Picea glauca*), monocots and dicots. The CYP74B subfamily includes hydroperoxide lyases. However, as one moves farther back in phylogeny to rice and moss, the number of CYP74 subfamilies increases as the sequences diverge, and it becomes less clear which subfamilies encode hydroperoxide lyases and

which encode divinyl ether synthases. Of the three members in the CYP74 clan of moss, one has been identified as allene oxide synthase (CYP74A1 GenBank AJ316566 and potentially CYP74A8) and one as hydroperoxide lyase CYP74G1 GenBank AJ316567. The CYP74 genes do not use molecular oxygen in their mechanism, so they have lost conservation of the I-helix oxygen binding region. This and other changes gives them a very long branch on the usual phylogenetic tree of plant P450s (Fig. 1).

The CYP85 clan has several functions, one is the previously mentioned CYP88 kaurenoic acid oxidase (46, 47) that is absent from moss and acts after CYP701 in the gibberellin pathway in higher plants. Other CYP85 clan members are involved in brassinosteroid metabolism. CYP85 and CYP90 members are required to make these particular growth inducing compounds (55-57). While these two families are not recognized as separate families in moss, the CYP763 family is closely related and probably shares a common ancestral gene with CYP85 and CYP90. It is not yet clear if moss produces brassinosteroids. BLAST searches for CYP genes in the EST database show a strong CYP90A-related sequence in the fern *Adiantum capillus-veneris* (BP916923, 48% to 90A1, 37% to 90B1). This provides evidence that brassinosteroid synthesis may exist in ferns with emergence of a clear precursor to the CYP90A subfamily.

The CYP86 clan has three families, CYP86, CYP94 and CYP704, in moss that are seen in Arabidopsis, poplar, pine and rice (36). These are. The functions of these families appear to have been established very early in plant evolution with land plants needing to protect themselves against water loss. This is accomplished by forming a waxy water barrier referred to as the cuticle. The monomers of cutins that make up this

barrier are ω-hydroxylated fatty acids with several CYP86 and CYP94 enzymes mediating hydroxylations on these fatty acids at the ω position (60-62) . It is notable that CYP96, a fourth well known family in the CYP86 clan, is absent in moss but present in monocots and dicots. CYP96 seems to be a younger CYP family.

CYP727 was first identified in rice and has since been sequenced in barley, wheat and poplar. No CYP727 family in evident in moss, but there is a related CYP751A1 sequence that is a probable ortholog of this family. CYP727A1 is represented by a single gene in rice, two exist in poplar, as the result of a recent genome duplication, while none are found in Arabidopsis. These are probably solo-function genes as seen with CYP51G.

More recent P450-containing pathways appearing in gymnosperms or angiosperms.

Present evidence concerning P450s in ferns and gymnosperms is incomplete. EST projects in these organisms do provide some information, but the absence of a specific CYP gene in these species cannot yet be confirmed because complete gene sets are not defined. Rice, Arabidopsis and poplar do provide complete CYP gene sets from one monocot and two dicots. The predicted origin of a CYP family in angiosperms may be pushed farther back as more sequence data accumulates for gymnosperms, ferns and the gemmiferous spike moss (*Selaginella moellendorffii*).

Searches of coniferales in the nr and ESTdb sections of GenBank with CYP96 sequences show the best BLAST hits are to CYP86 or CYP94, but not CYP96. Additional searches in filicales (ferns) found best hits that matched the CYP86 or CYP704B sequences, but not CYP96. This suggests that CYP96 may be absent from

ferns and conifers because it is an angiosperm-specific family. This family is quite abundant in rice, Arabidopsis and poplar.

Abscisic acid is a growth regulator/stress signal in plants that is largely associated with the prevention of water loss (63). The catabolism of abscisic acid is a P450-dependent 8'-hydroxylation catalyzed by members of the CYP707A subfamily (64-66). The CYP707 family is not seen in moss and BLAST searches of the nr or ESTdb sections at GenBank do not show any strong hits for coniferales. The best matches in conifers are in the CYP716 family, but these are seen in Arabidopsis, poplar and soybeans that also have a CYP707. Therefore, one is forced to conclude that CYP707 is an angiosperm-specific P450 family present in both monocots and dicots. An EST (CV005308.1) having about 70% identity to rice CYP707A6 is found in *Persea americana* (avocado) that is in the Laurales group which predates the divergence of monocots and dicots.

Flavonoids are made in very complex pathways that utilize several P450s. Over 9000 flavonoid structures are known (67). CYP75A is a flavonoid 3', 5'-hydroxylase and CYP75B is a flavonoid 3'-hydroxylase. Because the flavonoid 3', 5'-hydroxylase is missing in Rosacaea roses and their relatives cannot make blue pigments (68). Both of these P450s use dihydroflavonol as substrate which is made from flavanone by flavanone 3β-hydroxylase. Flavanone is a substrate for two P450 proteins CYP93B (flavone synthase II, FSII) and CYP93C (isoflavone synthase, IFS). CYP93C produces isoflavone that can be modified by isoflavone-O-methyl transferase to make formononetin. Isoflavones made by IFS are almost exclusively found in legumes (69). Formononetin is then a substrate for CYP81E1 (isoflavone 2'-hydroxylase) (70a) and CYP81E9 (isoflavone 3'-hydroxylase) (70b). Both CYP81E3 and CYP81E7 act later in this

pathway on the way to maackiain and pisatin (71). None of these CYP75, CYP81 or CYP93 families are found in moss. However, CYP75A14 and CYP75B13 are found in pine. It is possible that some of these enzymes may have been lost in *Physocomitrella* but are still present in other mosses since some mosses do have flavonoids, but "more than 50% of bryophyte species examined do not contain detectable flavonoids." (72).

CYP93B and CYP93C are missing in rice, although there are some related sequences named CYP93F and CYP93G. It is not clear what their function is. The fact that the best BLAST hits in conifers to CYP93B are CYP75A and CYP75B sequences suggests that The CYP75 family is older and the CYP93 genes arose from CYP75 genes by duplication and divergence. It is likely that the CYP93B (FSII) and CYP93C (IFS) genes are dicot-specific genes since they do not appear to be in rice. In contrast to the limited distribution of isoflavones and CYP93C to the legumes, the CYP93B sequences are found in wider distribution in both rosids and asterids. CYP93B (FSII) appears to be an older subfamily and CYP93C may have derived from it.

Ah, but the strawberries!

Some plant P450s (and probably the majority) are unique to the specific requirements and properties of individual plant families or possibly even a species. I have written about P450s and the individuality of species before, referring to polymorphisms in animals (73), but plants have three to five times as many P450s as animals, with the potential for some acting as taxon-specific enzymes. There are certainly P450s such as members of the CYP725 family involved in the taxol biosynthesis pathway of yew trees that make rare chemicals in a limited taxonomic range.

There are at least 8 monooxygenation steps in the taxol pathway that are all probably encoded by CYP725 enzymes (74). A BLAST search in GenBank with CYP725A1 showed 14 hits to Taxus species with expect values of 1e-146 or better. The next best hit is to *Picea sitchensis* AY779542 CYP716B1 at 1e-113. Another case of a unique biochemical pathway involving a P450 is vernolic acid biosynthesis in Euphorbiaceae species which is catalyzed by the CYP726A1 enzyme. Vernolic acid has also been found in Asteraceae species, but in these species it is made by a different Δ(12)-oleic acid desaturase-like enzyme that is not a P450 (75), so the CYP726 family is apparently unique to Euphorbia.

Finally, the flavor and aroma of strawberry fruits are dependent on more than 300 compounds, many of which are volatile chemicals. A recent review on the biochemistry of plant volatiles states that "more than 1000 low M_r organic compounds have been reported to be emitted from plants" (76). The major difference between the inedible wild strawberry (a diploid) and cultivated varieties (octoploid) is the presence of different terpene compounds. In the wild strawberry *Fragaria vesca,* pinene synthase (FvPINS) produces pinene that is hydroxylated to myrtenol by the CYP71AR1 P450 and acetylated to make myrtenyl acetate. In cultivated strawberries (*Fragaria* x *ananassa*), the pinene synthase gene FaPINS is defective, eliminating the P450 substrate pinene and production of myrtenol and myrtenyl acetate. Instead, a different terpene synthase (FaNES1) produces nerolidol and linalool (77). CYP71AR1 represents only one of 303 P450s presently named in the CYP71 family. Although it is not known what most of these enzymes do, it is likely that many of them are like the strawberry CYP71AR1, dedicated to unique functions in complex processes such as the production of fruit flavors and

aromas. Interestingly, the very first cloned plant P450 was CYP71A1, a fruit ripening P450 from avocado (78).

Conclusions

Plant cytochrome P450s can be divided into ancient genes that predate the emergence of land plants (CYP51, CYP97, CYP710, CYP746), and moderately old genes that were early innovations in evolution (CYP73, CYP78, CYP98, CYP701 and CYP703). Genes in all of these families are easily recognized in moss. There are also clade-specific genes for each level of plant evolution. Some are found only in gymnosperms and angiosperms (CYP75), or in the older Laurales (CYP707), a distant angiosperm group, but not further back. Some are limited to dicots and are missing in rice (CYP93B), while others are in more limited phylogenetic clades like yews (CYP725), legumes (CYP93C) and strawberries (CYP71AR). Following the phylogenetic distribution of these enzymes, it is possible to determine P450 families that mediate the primary core biochemistry of plants. The recent identification of functions for CYP97A3 (33) , CYP97C1 (32), CYP710A1 (34) and CYP711A1 (39) continues the process of defining P450 roles in plants. Only the CYP727 and CYP746 clans remain without any known function. Considering the nature of evolution, only a few P450 genes will be in the conserved central pathways. The majority can be expected to be in specialized roles in secondary metabolism such as vinca alkaloid synthesis as in tabersonine 16-hydroxylase CYP71D12 (79), secologanin synthase CYP72A1 (80), and geraniol 10-hydroxylase CYP76B6 (81). The majority of members in the CYP71 family are probably engaged in secondary metabolism.

By definition, the eleven plant P450 clans have ancient beginnings because they are recognized as the deepest branching gene clades in plants. Each of these eleven must harbor at least one pathway that is ancient and conserved in plant biology. Some of these clans contain highly conserved solo-function genes like CYP51, the three subfamilies of CYP97, CYP710, CYP711, and probably CYP727 and CYP746, even if we do not yet know what those functions are. The CYP74 clan, like the CYP97 clan, has only one family, but more than one function, with the production of jasmonate in the oxylipin pathway being one major CYP74 contribution and production of hexenal and other C-6 volatiles being another. The CYP86 clan is heavily biased toward fatty acid ω-hydroxylases and is key to cuticle synthesis. The CYP85 clan catalyzes the synthesis of brassinosteroids, while some in the CYP72 clan destroy these important signaling molecules. However, the CYP72 clan is present in *Chlamydomonas* so the original function cannot be breakdown of brassinosteroids, which came later. The earlier role of the ancient CYP72 clan is not yet identified. The CYP71 clan catalyzes many reactions in the phenylpropanoid pathway. Because it is so large it is perhaps very old with five family members seen in moss. With members such as CYP701 in the gibberellin pathway, it is probable that the CYP71 clan has other basic roles to play. The goal of studying P450 evolution in plants will be to reconstruct the "tree of life" for the whole plant clade and assign approximate or at least relative dates of appearance and functions to each P450 family branch and subfamily branch down to the tips of the tree. This has been just the outline of a much larger program with much left to discover.

Fig. 1

A neighbor-joining tree of 119 plant P450 sequences including all 71 from moss. CYP102A1 from Bacillus megaterium is included for the PHYLIP neighbor-joining algorithm to choose as an outgroup. The sequence alignment was created using ClustalW. Moss P450s are labeled in green. Non-moss sequences have initials after the name. At, *Arabidopsis thaliana*; Cs, *Cucumis sativa*; Cr, *Chlamydomonas reinhardtii*; Le, *Lycopersicon esculentum*; Pg, *Panax ginseng*; Os, *Oryza sativa*; *Picea glauca*, Pt, *Pinus taeda*; Vc, *Volvox carteri*; Vs, *Vicia sativa.* Families existing in green algae and land plants are CYP51, CYP97, CYP710 and CYP746 (blue ovals). These families and clans predate land plants. The additional plant CYP clans found in algae are CYP72, CYP85 and CYP711 (yellow ovals). These clans existed before land plants, but no CYP families in them are shared between land plants and algae. The CYP clans CYP71, CYP74, CYP86 and CYP727 (green ovals) are only found in land plants. The 120 sequences and the alignment used are posted at http://drnelson.utmem.edu/mosstreeseqs.html and http://drnelson.utmem.edu/mosstreealn.html.

References

1. Index of completed genomes http://www.nmpdr.org/index.php?id=87

National Microbial Pathogen Data Resource Center [homepage on the Internet]. Chicago: Computation Institute, University of Chicago/Argonne National Laboratory, 2005 [cited 2006 June 23]. Available from:www.nmpdr.org. Accessed June 23, 2006

2. Venter Institute marine microbial genome project.

http://www.venterinstitute.org/press/news/news_2005_02_24.php

Accessed June 23, 2006

3. Gordon and Betty Moore Foundation Microbial Genome sequencing project

http://www.moore.org/microgenome/microb_list.asp Accessed June 23, 2006

4. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements.

Nature 423, 241-254.

5. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M. (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science. 301, 71-76.

6. Fungal Genome Sequencing

http://www.genome.gov/11008243 Accessed June 23, 2006

7. Government list of genome projects and white papers

http://www.genome.gov/10002154 Accessed June 23, 2006

8. Sequence status of fungal genomes in the Fungal Genome Initiative

http://www.broad.mit.edu/annotation/fungi/fgi/ Accessed June 23, 2006

9. GiardiaDB http://gmod.mbl.edu/perl/site/giardia?page=intro

Accessed June 23, 2006

10. PlasmoDB: The Plasmodium Genome Resource

http://www.plasmodb.org/plasmo/home.jsp Accessed June 23, 2006

11. El-Sayed et al. N.M. (2005) Comparative genomics of trypanosomatid parasitic protozoa. Science 309, 404-409.

12. ToxoDB

http://toxodb.org/ToxoDB.shtml

Accessed June 23, 2006

13. Ensmbl Trace Repository Statistics http://trace.ensembl.org/perl/traceview?stats=1

Accessed June 22, 2006

14. UCSC Genome Browser, University of California, Santa Cruz

http://genome.ucsc.edu/cgi-bin/hgBlat Accessed June 22, 2006

15. Baylor College of Medicine sea urchin BLAST server

http://www.hgsc.bcm.tmc.edu/blast/blast.cgi?organism=Spurpuratus

Accessed June 22, 2006

16. Assembly/Alignment/Annotation of twelve related Drosophila species.

http://rana.lbl.gov/drosophila/ Accessed June 23, 2006

17. Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK Jr, Maiti R, Chan AP, Yu C, Farzad M, Wu D, White O, Town CD. Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. BMC Biol. 2005 Mar 3(1):7.

18. International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. Nature 436, 793–800.

19. Bevan M, Walsh S. (2005) The Arabidopsis genome: a foundation for plant research. Genome Res. 15, 1632-1642.

20. NSF 2010 Project: Functional Genomics of Arabidopsis P450s. http://arabidopsis-p450.biotec.uiuc.edu/ Accessed June 23, 2006

21. Ohyanagi H, Tanaka T, Sakai H, Shigemoto Y, Yamaguchi K, Habara T, Fujii Y, Antonio BA, Nagamura Y, Imanishi T, Ikeo K, Itoh T, Gojobori T, Sasaki T. (2006) The Rice Annotation Project Database (RAP-DB): hub for Oryza sativa ssp. japonica genome information. Nucleic Acids Res. 34, D741-744.

22. Joint Genome Institute *Populus trichocarpa* homepage http://genome.jgi-psf.org/Poptr1/Poptr1.home.html Accessed June 23, 2006

23. Why sequence Physcomitrella patens? http://www.jgi.doe.gov/sequencing/why/CSP2005/physcomitrella.html Accessed June 23, 2006

24. Community Sequencing Project plans 2005 http://www.jgi.doe.gov/sequencing/cspseqplans.html Accessed June 23, 2006

25. Joint Genome Institute *Chlamydomonas reinhardtii* homepage http://genome.jgi-psf.org/Chlre3/Chlre3.home.html Accessed June 23, 2006

26. NSF Genomics of Loblolly Pine Embryogenesis Project http://www.tigr.org/tdb/e2k1/pine/pine_status.shtml Accessed June 23, 2006

27. Lorenz WW, Sun F, Liang C, Kolychev D, Wang H, Zhao X, Cordonnier-Pratt MM, Pratt LH, Dean JF. (2006) Water stress-responsive genes in loblolly pine (Pinus taeda) roots identified by analyses of expressed sequence tag libraries. Tree Physiol. 26, 1-16.

28. Pavy N, Paule C, Parsons L, Crow JA, Morency MJ, Cooke J, Johnson JE, Noumen E, Guillet-Claude C, Butterfield Y, Barber S, Yang G, Liu J, Stott J, Kirkpatrick R, Siddiqui A, Holt R, Marra M, Seguin A, Retzel E, Bousquet J, MacKay J. Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. BMC Genomics. 2005 Oct 19;6:144.

29. ESTdb release 012006, June 16, 2006 summary by organism http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html Accessed June 23, 2006

30. Cytochrome P450 homepage http://drnelson.utmem.edu/CytochromeP450.html Accessed June 23, 2006

31. Chlamydomonas reinhardtii cytochrome P450s http://drnelson.utmem.edu/chlamydomonas.htm Accessed June 23, 2006

32. 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 ε-ring hydroxylation activity. Proc Natl Acad Sci USA. 101, 402-407.

33. Kim J, DellaPenna D. (2006) Defining the primary route for lutein synthesis in plants: the role of Arabidopsis carotenoid β-ring hydroxylase CYP97A3. Proc Natl Acad Sci U S A. 103, 3474-3479.

34. 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.

35. Kizawa H, Tomura D, Oda M, Fukamizu A, Hoshino T, Gotoh O, Yasui T, Shoun H. (1991) Nucleotide sequence of the unique nitrate/nitrite-inducible cytochrome P-450 cDNA from Fusarium oxysporum. J Biol Chem. 266, 10632-10637.

36. Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S. (2004) Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol. 135, 756-772.

37. Tissue profiling expression data for Arabidopsis P450 genes http://arabidopsis-p450.biotec.uiuc.edu/tissue_profiling.shtml Accessed June 23, 2006

38. 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.

39. Booker J, Sieberer T, Wright W, Williamson L, Willett B, Stirnberg P, Turnbull C, Srinivasan M, Goddard P, Leyser O. (2005) MAX1 encodes a cytochrome P450 family member that acts downstream of MAX3/4 to produce a carotenoid-derived branch-inhibiting hormone. Dev Cell. 8, 443-449.

40. Durst F, Nelson DR. (1995) Diversity and evolution of plant P450 and P450-reductases. Drug Metabol Drug Interact. 12, 189-206.

41. Batard Y, Schalk M, Pierrel MA, Zimmerlin A, Durst F, Werck-Reichhart D. (1997) Regulation of the cinnamate 4-hydroxylase (CYP73A1) in Jerusalem artichoke tubers in response to wounding and chemical treatments. Plant Physiol. 113, 951-959.

42. Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C. (2002) The *Arabidopsis REF8* gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J. 30, 33-45.

43. Humphreys JM, Chapple C. (2002) Rewriting the lignin roadmap. Curr Opin Plant Biol. 5, 224-229.

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

45. Helliwell CA, Poole A, Peacock WJ, Dennis ES. (1999) *Arabidopsis ent*-kaurene oxidase catalyzes three steps of gibberellin biosynthesis. Plant Physiol. 119, 507-510.

46. Winkler RG, Helentjaris T. (1995) The maize *Dwarf3* gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis.
Plant Cell 7, 1307-1317.

47. Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ. (2001) The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway.
Proc Natl Acad Sci USA. 98, 2065-2070.

48. Warne, T.R. and Hickok, L.G. (2005) Evidence for a gibberellin biosynthetic origin of Ceratopteris antheridiogen 1. Plant Physiol. 139, 1935-1945.

49. Yamane, H Takajhashi, N. Takeno, K. Furuya M. (1979) Identification of gibberellin A9 methyl ester as a natural substance regulating formation of reproductive organs in *Lygodium japonicum*. Planta 1437, 251-256.

50. Malonek S, Bomke 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. Phytochem. 66, 1296-1311.

51. 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.

52. 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.

53. Imaishi H, Matsumoto Y, Ishitobi U, Ohkawa H. (1999) Encoding of a cytochrome P450-dependent lauric acid monooxygenase by CYP703A1 specifically expressed in the floral buds of petunia hybrida. Biosci. Biotechnol. Biochem. 63, 2082-2090.

54. 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.

55. Kim TW, Hwang JY, Kim YS, Joo SH, Chang SC, Lee JS, Takatsuto S, Kim SK. (2005) Arabidopsis CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. Plant Cell 17, 2397-2412.

56. Nomura T, Kushiro T, Yokota T, Kamiya Y, Bishop GJ, Yamaguchi S. (2005) The last reaction producing brassinolide is catalyzed by cytochrome P-450s, CYP85A3 in tomato and CYP85A2 in *Arabidopsis*. J Biol Chem. 280, 17873-17879.

57. Bancos S, Nomura T, Sato T, Molnar G, Bishop GJ, Koncz C, Yokota T, Nagy F, Szekeres M. (2002) Regulation of transcript levels of the *Arabidopsis* cytochrome p450 genes involved in brassinosteroid biosynthesis.
Plant Physiol. 130, 504-513.

58. 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 the internodes elongation by modulating GA responses in rice. Plant Cell Physiol. 47, 181-191

59. Howe GA, Schilmiller AL. (2002) Oxylipin metabolism in response to stress. Increasing evidence indicates that the collective biological importance of oxylipins in plants is comparable to that of the eicosanoid family of lipid mediators in animals. Curr Opin Plant Biol. 5, 230-236.

60. 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 ω-hydroxylation in development. Proc Natl Acad Sci USA. 98, 9694-9699.

61. 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.

62. Duan H, Schuler MA. (2005) Differential expression and evolution of the *Arabidopsis* CYP86A subfamily. Plant Physiol. 137, 1067-1081.

63. The abscisic acid site. http://cbr-rbc.nrc-cnrc.gc.ca/abscisicacid/ Accessed June 23, 2006

64. Nambara E, Marion-Poll A. (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol. 56, 165-185.

65. 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.

66. 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.

67. Martens S, Mithofer A. Flavones and flavone synthases. (2005) Phytochem. 66, 2399-2407.

68. de Vetten N, ter Horst J, van Schaik HP, de Boer A, Mol J, Koes R. (1999) A cytochrome b5 is required for full activity of flavonoid 3', 5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. Proc Natl Acad Sci USA. 96, 778-783.

69. Jung W, Yu O, Lau SM, 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.

70a. AkashiT, AokiT, Kameya N, Nakamura I and Ayabe S (1997) Two new cytochrome P450 cDNAs (Accession Nos. AB001379 and AB001380) from elicitor-induced licorice (*Glycyrrhiza echinata* L.) cells (PGR97-167). Plant Physiol. 115, 1288

70b. Liu, C.J., Huhman, D., Summer, L.W. and Dixon, R.A. (2003) Regiospecific hydroxylation of isoflavones by cytochrome P450 81E enzymes from *Medicago truncatula*. Plant J. 36, 471-481.

71. MetaCyc Pathway: maackiain biosynthesis http://biocyc.org/META/NEW-IMAGE?object=PWY-2464 Accessed June 23, 2006

72. Stafford, HA (1991) Flavonoid evolution: an enzymatic approach. Plant Physiology 96, 680-685. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1080830 Accessed June 23, 2006

73. Nelson DR. (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys. 369, 1-10.

74. Schoendorf A, Rithner CD, Williams RM, Croteau RB. (2001) Molecular cloning of a cytochrome P450 taxane 10 β-hydroxylase cDNA from Taxus and functional expression in yeast. Proc Natl Acad Sci USA. 98, 1501-1506.

75. 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.

76. Dudareva N, Pichersky E, Gershenzon J. (2004) Biochemistry of plant volatiles. Plant Physiol. 135, 1893-1902.

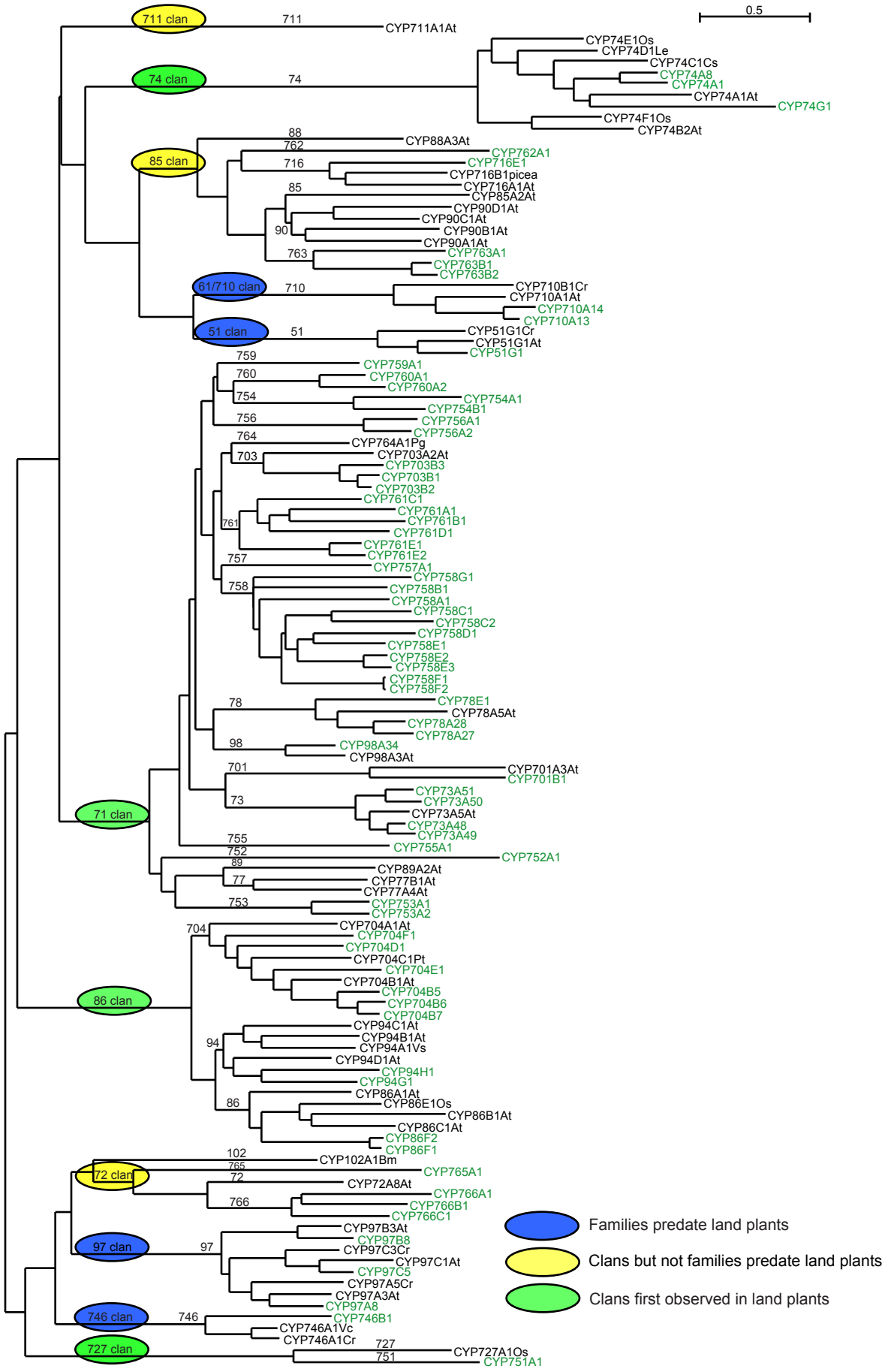
77. Aharoni A, Giri AP, Verstappen FW, Bertea CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ. (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell 16, 3110-3131.

78. Bozak KR, Yu H, Sirevag R and Christoffersen RE (1990) Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit Proc Natl Acad Sci USA 87, 3904-3908.

79. Schroder G, Unterbusch E, Kaltenbach M, Schmidt J, Strack D, De Luca V, Schroder J. (1999) Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. FEBS Lett. 458, 97-102.

80. Irmler S, Schroder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, Strack D, Matern U, Schroder 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.

81. 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.


0.5
711 clan
711
CYP711A1At
74 clan
74
CYP74E1Os
CYP74D1Le
CYP74C1Cs
CYP74A8
CYP74A1
CYP74A1At
CYP74G1
CYP74F1Os
CYP74B2At
85 clan
88
CYP88A3At
762
CYP762A1
716
CYP716E1
CYP716B1picea
CYP716A1At
85
CYP85A2At
CYP90D1At
CYP90C1At
90
CYP90B1At
CYP90A1At
763
CYP763A1
CYP763B1
CYP763B2
61/710 clan
710
CYP710B1Cr
CYP710A1At
CYP710A14
CYP710A13
51 clan
51
CYP51G1Cr
CYP51G1At
CYP51G1
759
CYP759A1
760
CYP760A1
CYP760A2
754
CYP754A1
CYP754B1
756
CYP756A1
CYP756A2
764
CYP764A1Pg
703
CYP703A2At
CYP703B3
CYP703B1
CYP703B2
CYP761C1
CYP761A1
761
CYP761B1
CYP761D1
CYP761E1
CYP761E2
757
CYP757A1
CYP758G1
758
CYP758B1
CYP758A1
CYP758C1
CYP758C2
CYP758D1
CYP758E1
CYP758E2
CYP758E3
CYP758F1
CYP758F2
78
CYP78E1
CYP78A5At
CYP78A28
CYP78A27
98
CYP98A34
CYP98A3At
701
CYP701A3At
CYP701B1
73
CYP73A51
CYP73A50
CYP73A5At
CYP73A48
CYP73A49
71 clan
755
CYP755A1
752
CYP752A1
89
CYP89A2At
77
CYP77B1At
CYP77A4At
753
CYP753A1
CYP753A2
704
CYP704A1At
CYP704F1
CYP704D1
CYP704C1Pt
CYP704E1
CYP704B1At
CYP704B5
CYP704B6
CYP704B7
86 clan
CYP94C1At
CYP94B1At
94
CYP94A1Vs
CYP94D1At
CYP94H1
CYP94G1
CYP86A1At
86
CYP86E1Os
CYP86B1At
CYP86C1At
CYP86F2
CYP86F1
102
CYP102A1Bm
765
CYP765A1
72 clan
72
CYP72A8At
766
CYP766A1
CYP766B1
CYP766C1
CYP97B3At
CYP97B8
97 clan
97
CYP97C3Cr
CYP97C1At
CYP97C5
CYP97A5Cr
CYP97A3At
CYP97A8
746 clan
746
CYP746B1
CYP746A1Vc
CYP746A1Cr
727 clan
727
CYP727A1Os
751
CYP751A1
Families predate land plants
Clans but not families predate land plants
Clans first observed in land plants