1

Cytochrome P450 Nomenclature, 2004

David R. Nelson

Summary

Aspects of cytochrome P450 (CYP) nomenclature are addressed. The rules for naming a P450 are outlined, though individuals should not name their own genes. The nomenclature is presented as a unifying principle to enhance communication across disciplines. Because of the historical nature of gene sequencing, sometimes names have to be changed, but this is kept to a bare minimum to avoid confusion in the literature. CYP names have now reached four digits owing to proliferation of CYP families in the fungi and lower eukaryotes. For example, *CYP5034A1* is from *Ustilago maydis*. P450 sequence motifs are described that are useful in making global alignments. CYP clans are defined as clusters of CYP families. The clan names are useful in describing higher-order evolution of the gene superfamily. The nomenclature of orthologs and pseudogenes is also discussed.

Key Words: Cytochrome P450; CYP; P450 clans; nomenclature; motifs.

1. Introduction

1.1. Moving From Hundreds to Thousands of Sequences

The previous publication of this chapter recognized 753 named P450 sequences in mid-1998. In September 2004, the Cytochrome P450 (CYP) count was 3811 and rapidly moving to 4000. Eukaryotic genomes are being sequenced in months, not years, and annotation has become the rate-limiting step. The nomenclature system for cytochrome P450, first devised in 1987, has become strained, but it is not broken ***(1)***. This system relies on evolutionary relationships as depicted in phylogenetic trees. There is an arbitrary 40% amino acid sequence identity rule for membership in a family and a 55% rule for membership in a subfamily. The actual decision to include a sequence in an existing group largely depends on how it clusters on a tree and not so much on the absolute percentage of identity, which is more or less a rule of thumb. Owing to the great diversity of P450s in insects, fungi, and bacteria, there is a need for additional layers of nomenclature, above the family/subfamily level. This is similar to the multiple levels of the Linnean classification scheme for species. The concept of clans has

From: *Methods in Molecular Biology, vol. 320: Cytochrome P450 Protocols: Second Edition*
Edited by: I. R. Phillips and E. A. Shephard © Humana Press Inc., Totowa, NJ

been introduced *(2,3)* as a level above family rank. It is possible that subclans or superclans may be needed, but the exact details of this are only now being discussed by those concerned with P450 nomenclature. The general Web repository for P450 nomenclature and sequence data is http://drnelson.utmem.edu/cytochromeP450.html.

1.2. Nomenclature Is Philosophy

The average person who needs to use a gene name does not consider how the name was arrived at, or what its implications are. Under the 40% rule, a new sequence submitted for naming by an individual may be 38% identical to an existing family such as CYP1, with three subfamilies. On a tree, this family may be well separated from its neighbors, so the new sequence seems to belong in the CYP1 family. In this case, it would become *CYP1D1*. However, someone under pressure for grant funding and tenure decisions, would strongly like to see the new sequence in a new family. A decision must be made based on the best interests of the nomenclature; this may disappoint an individual.

1.3. Nomenclature Is a Unifying Principal

Biochemists, geneticists, molecular biologists, and others who discover genes, often name those genes in an appropriate manner based on how they work or to what pathway they belong. Often, these are very useful names and well understood by researchers in these fields. Frequently, the gene has been named for a mutant phenotype before the gene sequence was discovered. *ERG5* and *ERG11* are two P450 genes in the ergosterol pathway of fungi, but they are also *CYP61A1* and *CYP51F1*. The Halloween genes disembodied (*dib*), phantom (*phm*), shade (*shd*), and shadow (*sad*) are required for ecdysone synthesis and its further metabolism in arthropods. They are embryonic lethal mutations in *Drosophila*. These genes are also Cyp302a1, Cyp306a1, Cyp314a1, and *Cyp315a1*, respectively *(4–6)*. Spook (*spo*) is a fifth Halloween P450 gene, but the sequence has not been revealed yet *(7)*. The two nomenclatures exist side by side and will be used by different groups for different audiences. There is nothing wrong with this. The Cyp names identify these genes as P450 genes. The numbers in the 300 range identify them as animal P450s. Furthermore, *Cyp302a1*, *Cyp314a1*, and *Cyp315a1* are in the mitochondrial clan, whereas *Cyp306a1* is in the CYP2 clan (*see* **Subheading 1.4.**). This information follows from the P450 nomenclature and shows relationships to other P450 genes in these clans. The CYP nomenclature should be cross-referenced when the phenotype-specific names are used.

1.4. Nomenclature May Change

Because names are assigned in historical order, often without the benefit of knowing whole genome P450 collections and/or related genes from other species, some names will need to be revised. When 455 P450s were named from the rice genome, 3 P450s from *Arabidopsis* were found to cluster in new locations on the phylogenetic tree *(8)*. *CYP709A1* and *CYP709A2* became *CYP735A1* and *CYP735A2*, respectively, because they clearly separated from the other CYP709 sequences. *CYP721B1* became *CYP734A1* for similar reasons. From the 272 *Arabidopsis* genes and pseudogenes to the total collection now of more than 1100 plant P450s, only three names were changed

in *Arabidopsis*. In the future, name changes in existing plant sequences should be even more rare or nonexistent, because the nomenclature stabilizes as the sequence space or diversity is more completely known.

1.5. Nomenclature Growing Pains

The 3811 P450s named to date have exceeded the capacity of the original nomenclature system that was based on fewer than 100 P450 families in eukaryotes and an open-ended number in bacteria. The first 100 families were divided among animals (CYP1–49), lower eukaryotes (CYP51–69), and plants (CYP71–99). Bacteria began with CYP101 and could go upward from there. Plants were the first to break the system. The solution was to give three-digit CYP names, allowing for about 1000 new CYP families. The numbering scheme is similar to that of the first 100, except bacteria already were in the 100 and higher range. The four divisions were assigned as follows: bacteria, CYP101–299; animals, CYP301–499; lower eukaryotes, CYP501–699; and plants, CYP701–999. Bacteria, animals, and plants still fit this three-digit system, but lower eukaryotes have already broken the CYP699 limit (*CYP699A1* is from *Agaricus bisporus*, a mushroom). A four-digit system has been set up with similar ranges: bacteria, CYP1001–2999; animals, CYP3001–4999; lower eukaryotes, CYP5001–6999; and plants, CYP7001–9999. This historical naming system may induce some CYP name prejudice, with the one- and two- digit names occupying a privileged place in the eyes of some. *CYP1A1* from vertebrates may be accorded higher status than *CYP5034A1* from *Ustilago maydis*. This is not the intent of the nomenclature, but it may reflect where the early experimental efforts on P450 first uncovered these genes. For instance, there are no vertebrate P450s above *CYP51*.

2. Naming a New P450 Gene

New P450 genes should be submitted to the Committee on Standardized Cytochrome P450 Nomenclature for naming. That committee consists of David Nelson and any other expert needed in case of a tricky nomenclature problem. Usually, that means David Nelson will name the gene, and the sequence is kept confidential. Names should not be assigned by an author without help from this central committee in order to avoid inaccurate or duplicate names. One author named a rabbit gene *CYP12* because it hydroxylated its sterol substrate on the 12 position ***(9)***. However, CYP12 is a family of insect P450s found in mitochondria, so this gene was renamed *CYP8B1*. Because many confidential sequences are named, the only way to be sure one's sequence has not already been assigned a name is to consult the committee. In addition, as described in **Subheading 4.**, there are other considerations, such as orthology, that can affect a name. The 40% rule is a guideline that can be broken for good cause. It is very helpful to do BLAST searches and identify the best matches in the public database, but one should remember that there may be even better matches in the confidential set of P450s.

There is no limit to how many sequences one may submit for naming. The record as of this writing is all the P450s found in *Fusarium graminearum, Magnaporthe grisea, Aspergillus nidulans,* and *Neurospora crassa*. The *N. crassa* genes had already been

named, but the others included 341 P450 genes. It took more than a month to assign names to these genes.

The process that is used to name a gene is first BLAST searching known public P450s on the P450 BLAST server at http://132.192.64.52/p450.html. Thousands of P450s on this server have been assembled from genomic DNA or mRNA sequences, and sometimes overlapping expressed sequence tags have been used to make a complete gene. This set does not exist anywhere else in this useful form, because it is a curated nonredundant collection. The sequences are sorted by species or, in some cases, by larger taxonomic groups such as bacteria. Just a few searches here often can identify the best BLAST match in the named set of P450s and narrow the process down to a single family or subfamily. If the sequence does not match any known P450 at the 40% identity level or higher, then the sequence may be in a new family. This is often the case for fungal or bacterial sequences. It is never the case for vertebrate sequences, because no new families of P450s have been discovered in vertebrates for at least 5 yr.

If the sequence is in a new family, more extensive searches need to be done in Genbank for related sequences. It may be necessary to name a whole new family of sequences, rather than just the one submitted. Sometimes, a sequence that is from an animal or a plant may not match any families in those groups. It is a good practice to search fungal and bacterial sequences to detect a potential contamination. I have had to notify researchers conducting genome-sequencing projects of these problems.

Once a sequence is determined to belong to a family, the naming process can be completed quickly if the sequence has 55% or greater identity to a known sequence. If it is less, then the subfamily must be determined. New subfamilies are continually being found. They can be hard to name in the large families of CYP4, CYP6, and CYP71. The CYP4 family has 43 subfamilies. The first 24 use single letters of the alphabet, as in CYP4A, CYP4B, and so on. Once CYP4Z was reached (*CYP4Z1* is a human sequence), double letters were used for subfamilies, CYP4AA, CYP4AB, up to *CYP4AX2*, a silkworm sequence. This process will continue using CYP4BA, CYP4BB, and so on. Note that the letters *O* and *I* are not used for subfamily names to avoid confusion with one and zero. Potentially 24^2, or 576, subfamilies are possible with double letters.

To assign a new subfamily in a complex family such as CYP4, a tree must be made to see how the new branch clusters with all the other subfamilies. Usually this is done with a subset of sequences, but it may include up to 160 sequences for detailed results. Typical tree-drawing programs do not make readable trees if more than 160 sequences are included, but this is usually not required. For final assignment, it may be necessary to make both UPGMA and neighbor-joining trees to get a good idea where the new sequence belongs. Sometimes a judgment call is needed to decide whether a new sequence should be given a new subfamily name or be included in an existing subfamily. Additional information such as intron-exon locations can be quite helpful. Branches/sequences with low bootstrap support may shift from place to place in different trees, which can be problematic in naming these sequences.

3. Benchmarks for P450 Sequence Alignment

The construction of a tree is dependent on the sequence alignment. The program ClustalW is often used for this purpose. Alignment of P450 sequences that are in the same family is quite good by this automated method, but more distant sequences will pose problems for even the best algorithm. Alignments should be checked by eye to ensure that the benchmarks for P450 sequences line up properly. If they do not, they will need to be manually adjusted.

These P450 benchmarks or motifs are few in the N-terminal region, and more abundant in the C-terminal from the I-helix to the end of the P450 molecule. Many of these benchmark regions are specific for families or higher taxonomic groups, such as bacterial or mitochondrial P450s that tend to cluster together. When aligning difficult sequences with low similarity to other sequences, pay particular attention to these benchmark regions. Alignment algorithms cannot use this information, and they are often quite inaccurate when it comes to distantly related P450 sequences. This is understandable when a bacterial sequence may be only 8 to 9% identical to a mammalian sequence.

At the N-terminal of many plant and animal P450s is a proline-rich sequence. It is often PPGP, but it may be (hydrophobic; hdr)PGP. This sequence is usually followed four residues later by P(hdr)(hdr)G(polar)(hdr), as in PIIGNL. This region follows the N-terminal membrane-anchor sequence, which is usually not conserved and cannot be aligned with any certainty. About 20 amino acid residues later, there is a common sequence, KYG or RYG. This tripeptide is present in *CYP51*, lanosterol-14α-demethylase, which is one of the more ancient eukaryotic P450s involved in cholesterol biosynthesis. It is also present in *CYP7*, *CYP8*, and *CYP19* sequences, but not in *CYP4* or *CYP52* (fatty-acid and alkane hydroxylases) nor in bacterial-like sequences. The two basic P450 forms are called E-like and B-like P450s for eukaryote and bacterial-like, respectively. A tree computed from representatives of all known bacterial CYP families showed a clear break into E and B branches, with only two sequence families falling outside this pattern. These two were *CYP152* and *CYP198*. *CYP152A1* is a known peroxygenase that uses hydrogen peroxide instead of molecular oxygen in its reaction ***(10)***. *CYP152A2* is the only known P450 from an anaerobe, *Clostridium acetobutylicum* (GenPept AAK81262). This P450 in an anaerobe may serve a protective role against hydrogen peroxide. This major substrate difference may explain why *CYP152* falls outside the E and B clusters. The function of *CYP198* (GenPept AAM42184) is unknown.

The ancient origin of the KYG motif is further confirmed by its presence in *CYP110* from *Anabaena*, a cyanobacterium. In modified form (ELG) it is also seen in *CYP11B* and *CYP102* ($P450_{BM-3}$) from *Bacillus megaterium*, where it occurs just before the b 1-1 strand. Both *CYP110* and *CYP102* are E-like bacterial P450s, suggesting that the fundamental split between E- and B-like P450 structures occurred before eukaryotes developed. For comparison of the E-like P450 crystal structure of *CYP102*, and the B-like P450 crystal structures of *CYP101* ($P450_{cam}$) and *CYP108* ($P450_{terp}$), *see* **ref. *11***.

In bacterial B-like P450s, near amino acid 50 there is a conserved trp residue, often seen in the pattern WXXT(R,K), as in WIATK or WLVTR. This is the first region to look for near the N-terminus of a new bacterial sequence. About 100 amino acid residues from the N-terminus of bacterial sequences, there is another region, around the C-helix, that is critical when aligning bacterial sequences with eukaryotic sequences. The motif is DPPXHXXXR. This motif corresponds to a C-helix sequence, WXXXR, conserved in most eukaryotic P450s. The R is often followed by another basic group, as in WREQRR. Some lower eukaryotic P450s (CYP53, 57, and 58) have an H instead of the W seen in most eukaryotes.

After the C-helix, there is a long nonconserved stretch with few alignment cues. One worth mentioning is the (N,D,S)(V,I,T)(V,I) sequence around positions 175–180 in eukaryotic sequences. This E-helix region can be diagnostic for some families, as in the 3A subfamily, where the sequence reads GAYSMD(V,I)I, or the 4A subfamily SLMTLDT(I,V). After this region, the next well-conserved area is the I-helix. From here (near position 300) the P450s are much more strongly conserved and alignments are much easier. The I-helix has a characteristic sequence A(A,G)X(E,D)T, in which T is part of the molecular oxygen-binding site. This region is not conserved in enzymes that use peroxides as substrates, such as allene oxide synthase (CYP74), because they do not have to bind molecular oxygen. Consequently, the CYP74 family is one of the most distant branches on the P450 tree. This may not reflect its true evolutionary history, because one of the key regions used in computing trees is not conserved.

Exactly 16 amino acid residues from the T previously mentioned is a conserved P. This P is present even in many bacterial sequences, so it is a good marker for the junction between the I and J helices. The sequence conservation continues for another 15 residues in eukaryotes to a highly conserved G or P residue. At this point, there seems to be some variation in sequence length in the region between the J-helix and the K-helix. The K-helix is the best conserved feature in P450s, with its invariant EXXR charge pair. About 22 amino acid residues beyond the K-helix R, there is an (aromatic)X(I,V,L)P(K,A)G sequence that spans the connection between the β-2-2 strand and the β-1-3 strand (nomenclature of Peterson and Graham-Lorence; *see* **ref. *11***). The P(K,A) pair lies in this gap and causes a sharp change in direction. The region beyond this area is not structurally well defined, but 17 residues from the G there is a sequence, (aromatic)XX(P,D), that is very helpful in making alignments. It is followed four to five residues later by the PERF sequence, with some variations, as in PSRF, PDNF, PQRW, and so on. The PERF motif is not present in bacterial B-like P450 sequences, although it is in CYP102, CYP110, and CYP118 bacterial E-like P450s. It is part of the meander before the heme thiolate ligand (nomenclature of Peterson and Graham-Lorence; *see* **ref. *11***). After this point, there are length variations before reaching the signature sequence of all P450s, FXXGXXXCXG, with some variations allowed except at Cys. Even this one invariant residue has been replaced by His in two sequences from *Ciona intestinalis* (Genbank AK173774) and *Ciona savignyi*. Beyond the signature, 18 residues from the Cys, there is a conserved tetrapeptide, LQNF, or variants of it. The C-terminal region is quite variable, with an unexplained occurrence of a PR approx 33–34 residues from the F in LQNF. In many

Table 1
Percentage of Identity Among Some Orthologs in *Drosophila, Anopheles,* and *Apis*[a]

	Drosophila melanogaster	*Drosophila pseudoobscura*	*Anopheles gambiae*	*Apis mellifera*
18A1	100	96	Absent	60
49A1	100	91	63	46
301A1	100	89	75	69
302A1 dib	100	82	56	48[b]
303A1	100	89	49	43[b]
306A1 phm	100	66	48	47
314A1 shd	100	90	49	45
315A1 sad	100	79	36	38

[a]Values are from a comparison to *Drosophila melanogaster*. *dib, phm, shd,* and *sad* are Halloween genes important in development.

[b]Incomplete *Apis* sequences.

sequences, these are the last two amino acid residues, whereas in others there is an extension beyond the PR sequence.

These motifs form the main alignment benchmarks that are used in making P450 sequence alignments. Often the distance between benchmarks is absolutely the same, and a cursory check for reasonable homologies between the segments can confirm that the alignment is probably correct without introducing gaps. Of course, this is most difficult to do in the N-terminal region and when aligning bacterial or lower eukaryotic sequences that are very divergent from the other sequences.

4. Orthologs and Nomenclature

The world of P450 genes has expanded well beyond the limited mammalian realm. Because mammals dominated the early P450 studies, the rules for naming P450s were based largely on P450 statistics from mammals. Orthologs in mammals tend to be very conserved, ranging from 66% for *CYP17A1* to 96% for *CYP26B1* for mouse and human. The average is about 81%. These values fit nicely with the 40% rule for family membership and the 55% rule for subfamily membership. However, when more diverse phyla are considered, such as Arthropoda, the orthologs are not as strongly conserved. **Table 1** shows that orthologs in the single genus *Drosophila* are about as divergent as the orthologs between mouse and human. Orthologs among *Drosophila, Anopheles,* and *Apis* (honeybee) fall outside the 55% rule frequently and they even fall outside the 40% rule for *CYP315A1*. The *Anopheles* and *Apis 315A1* sequences have very low identities to *Drosophila 315A1*; however, the best BLAST scores for these sequences are clearly to *CYP315A1*. Niwa et al. ***(5)*** biochemically identified silkworm *315A1* as the ortholog of the *Drosophila* shadow (sad) gene, yet it only has a 36% identity to the *Drosophila* sequence, so the low percentage identity is not owing to mislabeling the orthologs. The value of the nomenclature is diminished if orthologs are placed in different subfamilies and even different families. When possible,

orthologs are given the same name, even if it violates the 55% rule or even the 40% rule.

5. Pseudogenes and Nomenclature

Pseudogenes are abundant in humans and mice and less common in fungi or bacteria. These faulty genes pose a nomenclature problem *(12)*. They were originally tagged with a P on the end of the CYP name to indicate a pseudogene, as in *CYP2T2P*. As whole genomes were sequenced, it became apparent that there were at least four classes of pseudogene. The nearly intact pseudogene is still tagged with a P. The solo exon pseudogene is a small piece of a gene, just one or a few exons, far away from any other intact P450 genes. These have been named with an extension on the CYP name, as in *CYP4F-se1[6:7]*, in which -se1 stands for solo exon one of the CYP4F subfamily. The [6:7] indicates which exons of the *CYP4F* gene are present in the pseudogene. Detritus exons are indicted by the extension -de. These pseudogenes are close to a known gene and represent recent duplications of one or more exons from that gene. These exons may be alternative splice exons. The fourth type of pseudogene is an internal exon duplication. These pseudogenes may be intact or partial exons that occur inside a known gene. These exons form alternative transcripts. They are named with the extension -ie, for internal exon. The exact details of pseudogene naming are complicated. These are described in an article on all human and mouse P450 genes *(13)*. Also in that article is a nomenclature for naming P450 alternative transcripts.

6. P450 Clans

The nomenclature for P450s used to be small, with only a few dozen families. Taxa such as vertebrates had fewer than 20 families, and one could easily memorize the family names. More diverse groups such as plants now have 62 families, and this is slowly growing. Fungi and bacteria are so diverse that every second or third sequence is in a new family. One cannot memorize the family names and recognize relationships between sequences in that way with these highly diverse groups. To restore some order to this exploding nomenclature system, higher-order groups were proposed. These are called clans, and they are essentially like clades, although clades technically refer to species with a common ancestor and not to sequences.

Clans have been defined as groups of P450 families that consistently cluster together on phylogenetic trees. No percentage of identity cutoff is given, because that would not work in this case. The 62 families of plant P450s sort into just 10 clans, once again a number that is reasonable and can be grasped and remembered. For details on plant P450 clans, *see* the article by Nelson et al. *(8)* on comparative genomics of rice and *Arabidopsis* P450s. The animal clans are still being defined, but it is clear that the insect CYP6 and CYP9 families belong in a clan with vertebrate CYP3 and CYP5. This has been called the CYP3 clan for the lowest family number in the group. Insects appear to have four clans, CYP2, CYP3, CYP4, and mitochondrial. These clans have families that cluster with vertebrate CYP2, CYP3, CYP4, and mitochondrial (CYP11, CYP24, CYP27) families. Vertebrates have about 10 clans, although some, such as CYP19, have only one family. The clans that are in common between insects and

vertebrates must have had a common ancestral sequence in the bilaterian ancestor of animals. For a tree with vertebrate P450 clans shown, see my article on the comparison of human and Fugu P450s ***(14)***. There is no comprehensive assignment of vertebrate and invertebrate clans yet.

7. Of CYPs and CYPs

The CYP root of cytochrome P450 gene names is not exclusively used for P450 genes. Without realizing that this root term was used for cytochrome P450, it has also been used for cyclophilins. This has happened in *Caenorhabditis elegans*, mammals, *Arabidopsis*, and *Chlamydomonas*. The Committee on Standardized Cytochrome P450 Nomenclature has been trying to correct this ***(15)***. The HUGO Human and Mouse Gene Nomenclature Committees recently approved the use of *CYN* in human and *Cyn* in mouse for the official gene names for cyclophilins. The *C. elegans* community also agreed to change 17 official names from *CYP* to *CYN*. Unfortunately, in the issue of *Plant Physiology* preceding the one that carried the official nomenclature for 455 rice *CYP* genes ***(8)***, an article was published naming 29 *Arabidopsis* cyclophilins as *CYP* genes ***(16)***. This has not yet been addressed. Only two cyclophilin genes in *Chlamydomonas* had been named *CYP*, and it has been agreed to change them to *CYN*. These types of nomenclature problems are long lasting because they are published in the literature and are perpetuated in databases. It may take many years to clear up this confusion.

References

1. Nebert, D. W., Adesnik, M., Coon, M. J., et al. (1987) The P450 gene superfamily: recommended nomenclature. *DNA* **6,** 1–11.
2. Nelson, D. R. (1998) Metazoan cytochrome P450 evolution. *Comp. Biochem. Physiol. Pt. C* **121,** 15–22.
3. Nelson D. R. (1999) Cytochrome P450 and the individuality of species. *Arch. Biochem. Biophys.* **369,** 1–10.
4. Gilbert, L. I. (2004) Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in Drosophila melanogaster. *Mol. Cell. Endocrinol.* **215,** 1–10.
5. Niwa, R., Matsuda, T., Yoshiyama, T., et al. (2004) CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of Bombyx and Drosophila. *J. Biol. Chem.* **279,** 35,942–35,949.
6. Warren, J. T., Petryk, A., Marques, G., et al. (2004) Phantom encodes the 25-hydroxylase of Drosophila melanogaster and Bombyx mori: a P450 enzyme critical in ecdysone biosynthesis. *Insect Biochem. Mol. Biol.* **34,** 991–1010.
7. Petryk, A., Warren, J. T., Marques, G., et al. (2003) Shade is the Drosophila P450 enzyme that mediates the hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone. *Proc. Natl. Acad. Sci. USA* **100,** 13,773–13,778.
8. Nelson, D. R., Schuler, M. A., Paquette, S. M., Werck-Reichhart, D., and Bak, S. (2004) 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.
9. Eggertsen, G., Olin, M., Andersson, U., et al. (1996) Molecular cloning and expression of rabbit sterol 12alpha-hydroxylase. *J. Biol. Chem.* **271,** 32,269–32,275.

10. Matsunaga, I., Yamada, A., Lee, D. S., et al. (2002) Enzymatic reaction of hydrogen peroxide–dependent peroxygenase cytochrome P450s: kinetic deuterium isotope effects and analyses by resonance Raman spectroscopy. *Biochemistry* **41,** 1886–1892.
11. Peterson, J. A. and Graham-Lorence, S. E. (1995) Bacterial P450s: structural similarities and functional differences, in *Cytochrome P450: Structure, Mechanism and Biochemistry,* 2nd ed. (Ortiz de Montellano, P. R., ed.), Plenum, New York, pp. 151–180.
12. Nelson, D. R. (2004) Frankenstein genes, or the Mad Magazine version of the human pseudogenome. *Hum. Genomics* **1,** 310–316.
13. Nelson, D. R., Zeldin, D. C., Hoffman, S. M. G., Maltais, L., Wain, H., and Nebert, D. W. (2004) 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.
14. Nelson, D. R. (2003) Comparison of P450s from human and fugu: 420 million years of yertebrate P450 evolution. *Arch. Biochem. Biophys.* **409,** 18–24.
15. Nebert, D. W., Anastassios, N. Vasiliou, V., and Nelson, D. R. (2004) Cyclophilin nomenclature problems, or, "A Visit from the Sequence Police." *Hum. Genomics* **1,** 381–388.
16. Romano, P. G., Horton, P., and Gray, J. E. (2004) The Arabidopsis cyclophilin gene family. *Plant Physiol.* **134,** 1268–1282.