DNA AND CELL BIOLOGY
Volume 12, Number 1, 1993
Mary Ann Liebert, Inc., Publishers
Pp. 1-51

The P450 Superfamily: Update on New Sequences, Gene Mapping, Accession Numbers, Early Trivial Names of Enzymes, and Nomenclature

DAVID R. NELSON,[1] TETSUYA KAMATAKI,[2] DAVID J. WAXMAN,[3] F. PETER GUENGERICH,[4] RONALD W. ESTABROOK,[5] RENE FEYEREISEN,[6] FRANK J. GONZALEZ,[7] MINOR J. COON,[8] IRWIN C. GUNSALUS,[9] OSAMU GOTOH,[10] KYUICHIRO OKUDA,[11] and DANIEL W. NEBERT[12]

ABSTRACT

We provide here a list of 221 P450 genes and 12 putative pseudogenes that have been characterized as of December 14, 1992. These genes have been described in 31 eukaryotes (including 11 mammalian and 3 plant species) and 11 prokaryotes. Of 36 gene families so far described, 12 families exist in all mammals examined to date. These 12 families comprise 22 mammalian subfamilies, of which 17 and 15 have been mapped in the human and mouse genome, respectively. To date, each subfamily appears to represent a cluster of tightly linked genes. This revision supersedes the previous updates [Nebert *et al.*, DNA 6, 1-11, 1987; Nebert *et al.*, DNA 8, 1-13, 1989; Nebert *et al.*, DNA Cell Biol. 10, 1-14 (1991)] in which a nomenclature system, based on divergent evolution of the superfamily, has been described. For the gene and cDNA, we recommend that the italicized root symbol "*CYP*" for human ("*Cyp*" for mouse), representing "*cy*tochrome *P*450," be followed by an Arabic number denoting the family, a letter designating the subfamily (when two or more exist), and an Arabic numeral representing the individual gene within the subfamily. A hyphen should precede the final number in mouse genes. "*P*" ("*p*" in mouse) after the gene number denotes a pseudogene. If a gene is the sole member of a family, the subfamily letter and gene number need not be included. We suggest that the human nomenclature system be used for all species other than mouse. The mRNA and enzyme in all species (including mouse) should include all capital letters, without italics or hyphens. This nomenclature system is identical to that proposed in our 1991 update.

Also included in this update is a listing of available data base accession numbers for P450 DNA and protein sequences. We also discuss the likelihood that this ancient gene superfamily has existed for more than 3.5 billion years, and that the rate of P450 gene evolution appears to be quite nonlinear. Finally, we describe P450 genes that have been detected by expressed sequence tags (ESTs), as well as the relationship between the P450 and the nitric oxide synthase gene superfamilies, as a likely example of convergent evolution.

[1]Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599.
[2]Faculty of Pharmaceutical Sciences, Hokkaido University, N12 W6, Sapporo, Hokkaido, 060 Japan.
[3]Department of Biological Chemistry and Molecular Pharmacology and Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115.
[4]Department of Biochemistry and the Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, TN 37232.
[5]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75235.
[6]Department of Entomology, University of Arizona, Tucson, AZ 85721.
[7]Laboratory of Molecular Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892.
[8]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109.
[9]Department of Biochemistry, University of Illinois, Urbana, IL 61801.
[10]Saitama Cancer Center, 818 Komuro, Ina, Saitama 362, Japan.
[11]Department of Biochemistry, Hiroshima University School of Dentistry, Hiroshima 734, Japan.
[12]Department of Environmental Health and the Center for Environmental Genetics, University of Cincinnati Medical Center, Cincinnati, OH 45267-0056.

INTRODUCTION

P450 ENZYMES are important in the oxidative, peroxidative, and reductive metabolism of numerous endogenous compounds such as steroids, bile acids, fatty acids, prostaglandins, leukotrienes, and biogenic amines. Many of these enzymes also metabolize a wide range of foreign chemicals including drugs, environmental pollutants, natural plant products, and alcohols. The metabolism of foreign chemicals can frequently produce toxic metabolites, of which some have been implicated as agents that may be responsible for tumor initiation, promotion, and tumor progression.

It should be emphasized that the Nomenclature Committee of the International Union of Biochemistry (NC-IUB) prefers the term "heme-thiolate protein" instead of "cytochrome" for P450 (Palmer and Reedijk, 1991). The original term "cytochrome P-450" represents a holdover from when the protein was originally given its provisional name (Sato and Omura, 1961). These proteins are, in fact, not "cytochromes" in the true meaning of this terminology.

As will be discussed below, it is clear that this superfamily is very ancient, the ancestral gene having existed more than 3.5 billion years ago, at a time that obviously predates drugs, animal-plant interactions, and combusion of organic matter. On the basis of such an ancient phylogenetic tree, Nebert (1991) has proposed that the P450 enzymes, as well as other so-called "drug-metabolizing" enzymes, play an important role in maintaining the steady-state levels of endogenous ligands involved in ligand-modulated transcription of genes effecting homeostasis, growth, differentiation, and neuroendocrine functions.

With less than a dozen cDNA and deduced protein sequences available in 1985, it became apparent that the amino acid sequences could be aligned and classified on the basis of proposed divergent evolutionary relationships of the corresponding genes. A committee was formed to carry out this task. The foundations for the nomenclature system were laid with the first report (Nebert *et al.*, 1987) and extended in subsequent updates (Nebert *et al.*, 1989a, 1991). The latter two reports have also included recommendations of the Committee for Human Gene Mapping and the Committee on Standardized Nomenclature of Mouse Genes.

Each P450 gene almost always produces a single protein. To date, there appear to be only a few exceptions to this rule where "functional" alternative splicing might occur, *i.e.*, differential processing of the P450 transcript such that entire (translated) exons or portions of exons are exchanged to produce an enzyme with a new catalytic activity (Lacroix *et al.*, 1990; Lephart *et al.*, 1990; Miles *et al.*, 1990b) or tissue-specific expression (Means *et al.*, 1991).

NAMING A P450 GENE OR ENZYME

Naming a newly discovered P450 gene will require agreement by the P450 Nomenclature Committee, to prevent incorrect assignments or duplications of gene names. It would therefore be best for this Committee to examine each new P450 (protein) sequence and compare that sequence with all others in the data base. In many cases, the naming of genes will be carried out chronologically, meaning that delays in submitting a new sequence to this Committee might result in assignment of a higher number (or higher letter for subfamily designation).

Recommendations for naming a P450 gene or cDNA include the italicized root symbol "*CYP*" ("*Cyp*" for the mouse), denoting *cy*tochrome *P*450, an Arabic number designating the P450 family, a letter indicating the subfamily when two or more subfamilies are known to exist within that family, and an Arabic numeral representing the individual gene. With mouse genes or cDNAs, the final number is generally preceded by a hyphen. "*P*" ("*p*" in mouse) after the gene number is used to denote a pseudogene. If no second subfamily or second gene exists in a family, the subfamily and gene number need not be included, *e.g.*, *CYP5* or *CYP19*. The same nomenclature for the corresponding gene product (enzyme) is recommended, *e.g.*, italicized "*CYP1A1*" or "*CYP17*" ("*Cyp1a-1*" and "*Cyp17*," respectively, in mouse) for the gene and cDNA; nonitalicized "CYP1A1" or "CYP17" for the mRNA and protein in all species including mouse.

If a trivial name for the enzyme has already been well established, these trivial names might continue to be used in a publication—as long as the officially assigned name and the species under study are provided in the title, summary, or footnote, *e.g.*, "olf1" and "IIC16" are encoded by the rat *CYP2G1* and rabbit *CYP2C16* genes, respectively. Another possibility for designating the protein might be "P450 1A1" or "P450 2G1," or simply "1A1" or "2G1."

The consistent usage of the gene nomenclature is preferred. Authors may continue to use trivial names—as long as the names are compatible with GenBank and other nucleic acid and protein databases. For the protein, this means no hyphen in P450, no Greek letters, and no subscripts or superscripts. For example, $P450_{7\alpha}$, $P450_{scc}$, $P450_{11\beta}$, $P450_{17\alpha}$, $P450_{arom}$, $P450_{c21}$, $P450_{c24}$, and $P450_{c27}$ should be referred to as P450c7, P450scc, P450c11b, P450c17, P450arom, P450c21, P450c24, and P450c27, respectively, or simply c7, scc, c11 (or 11b), c17 (or 17a), arom, c21, c24, and c27.

A P450 protein sequence from one gene family is defined as usually having $\leq 40\%$ amino acid identity to a P450 protein from any other family. This definition of a P450 gene family was an arbitrary decision, but has turned out to be unexpectedly very useful. In those instances that have been examined thus far, genes within a defined subfamily have been found to be nonsegregating, *i.e.*, to lie within the same "gene cluster." For example, the rat *CYP2D1, CYP2D2, CYP2D3, CYP2D4,* and *CYP2D5* genes are located adjacent to one another on the same chromosome and form the "*CYP2D* cluster" (Table 1). These clusters have most likely arisen *via* gene duplication events, mostly during the last 400 million years (Nebert and Gonzalez, 1987; Nelson and Strobel, 1987; Gonzalez and Nebert, 1990; Nebert and Nelson, 1991); therefore, it is expected that each subfamily might be a genetically segregating distinct entity.

TABLE 1. CHROMOSOMAL AND SUBCHROMOSOMAL LOCALIZATION OF *CYP* GENES

P450 gene cluster	*Chromosomal location*	*References*
HUMAN		
CYP1	15q22-qter (near *MPI*)	Hildebrand *et al.* (1985b)
		Jaiswal *et al.* (1987a)
CYP2A	19q13.1–13.2	Phillips *et al.* (1985b)
		Davis *et al.* (1986)
		Davis *et al.* (1987)
		Miles *et al.* (1990c)
CYP2B	19q12–q13.2	Santisteban *et al.* (1988)
		Miles *et al.* (1988)
		Yamano *et al.* (1989b)
CYP2C	10q24.1–24.3	Okino *et al.* (1987)
		Riddell *et al.* (1987)
		Meehan *et al.* (1988a)
		Griffith *et al.* (1989)
		Shephard *et al.* (1989)
CYP2D	22q13.1	Gonzalez *et al.* (1988c)
		Gough *et al.* (1993)
CYP2E	10	McBride *et al.* (1987)
		Umeno *et al.* (1988a)
CYP2F	19	Nhamburo *et al.* (1990)
CYP3	7q22.1	Riddell *et al.* (1987)
		Brooks *et al.* (1988)
		Spurr *et al.* (1989)
		Inoue *et al.* (1992)
CYP4A	1	O.W. McBride and J.P. Hardwick (in preparation)
CYP4B	1p12–p34	Nhamburo *et al.* (1989)
CYP7	8q11–q12	Cohen, J.C., *et al.* (1992)
CYP11A	15q23–q24	Chung *et al.* (1986b)
		Sparkes *et al.* (1991)
CYP11B	8q21–q22	Chua *et al.* (1987)
CYP17	10q24.3	Matteson *et al.* (1986b)
		Sparkes *et al.* (1991)
		Fan *et al.* (1992)
CYP19	15q21	Chen, S., *et al.* (1988)
CYP21	6p (within *HLA*)	White *et al.* (1984c)
		White *et al.* (1985)
CYP27	2q33-qter	Cali and Russell (1991)
MOUSE		
Cyp1	Mid-9 (near *Mpi-1*)	Tukey *et al.* (1984)
		Hildebrand *et al.* (1985a)
Cyp2a	7 (near *Gpi-1*)	Kimura, S., *et al.* (1989c)
		Burkhart *et al.* (1990)
		Matsunaga, T., *et al.* (1990)
		Miles *et al.* (1990c)
Cyp2b	Proximal 7 (*Coh*)	Simmons and Kasper (1983)
		Simmons *et al.* (1985)
		Miles *et al.* (1990c)
Cyp2c	19	Meehan *et al.* (1988b)
Cyp2d	15	Gonzalez *et al.* (1987)
		Wong *et al.* (1989)
Cyp2e	7	Umeno *et al.* (1988c)
Cyp2f	7	Nhamburo *et al.* (1990)
Cyp3	6	Simmons *et al.* (1985)
Cyp4a	4 (near *Mtv-13*)	Kimura, S., *et al.* (1989b)
		Miles *et al.* (1991)

TABLE 1. (CONTINUED)

P450 gene cluster	*Chromosomal location*	*References*
Cyp11a	9	Youngblood *et al.* (1989)
Cyp11b	15	Mouw *et al.* (1989)
		Yoshioka *et al.* (1990)
		Domalik *et al.* (1991)
Cyp-17	19 (distal to *Got-1*)	Youngblood *et al.* (1991)
Cyp-19	9	Youngblood *et al.* (1989)
Cyp-21	17 (within *H-2*)	White *et al.* (1984a)
Cyp-27	1	J.J. Cali and D.W. Russell (personal communication)

It should be emphasized that, if two or more genes exist in a subfamily, only the gene cluster is listed in this Table. For example, *CYP1A1* and *CYP1A2* are located near one another on human chromosome 15 and represent the "*CYP1* cluster." *Cyp1a-1* and *Cyp1a-2* are near one another on mouse chromosome 9 and represent the "*Cyp1* cluster." The murine *Cyp2e-1* gene is listed as the "*Cyp2e* cluster," although it may be the only gene in this mouse subfamily. There are five known *CYP2D* functional genes in the rat *CYP2D* cluster, and the number of genes in this human or mouse subfamily is not yet firmly established.

Within a single family, the P450 protein sequences are >40% identical with several exceptions. (i) The *CYP2D, CYP2J,* and *CYP2K* subfamilies include the most distant members of the *CYP2* family. Exclusion of these subfamily sequences results in all remaining *CYP2* genes encoding enzymes that are ≥40% similar to each other. (ii) Although the CYP4C1, CYP4D1, and CYP4E1 sequences from insects are clearly related to the *CYP4* family, they are only 26–42% identical to the mammalian members of this family. It is clear, however, that these insect genes are related to the mammalian *CYP4* family. (iii) The CYP6A1 and CYP6B2 proteins are <40% similar, but sequences surrounding the conserved cysteine residue make it clear that these two genes are evolutionarily related. (iv) The nuclear genes encoding two mitochondrial P450 proteins, scc and 11b, are included in the same family (*CYP11*), even though the enzyme sequences are only 34–39% identical. It is worth noting that CYP27 is a mitochondrial protein and that the *CYP27* gene branches from the *CYP11* family (Nebert and Nelson, 1991). (v) The CYP52B and CYP52C sequences are 30–44% identical to the CYP52A sequences, but they form a tight cluster that workers in this research area wish to designate as a single gene family. (vi) The bacterial CYP105C1 protein is 39.6% identical to that of the CYP105A1 protein. Placement of these two genes in separate families would be splitting hairs.

Mammalian sequences within the same subfamily are always >55% identical. Inclusion of more distant species (*e.g.*, chicken c17 or trout IA1) within the same subfamily drops this value to >46%.

All genes within a given family that have been examined thus far contain the same number of exons and similar intron–exon boundaries. Interestingly, P450 genes in evolutionarily related families share some intron–exon boundaries; for example, *CYP17* and *CYP21* share all introns except two (Picado-Leonard and Miller, 1987), *CYP2* and *CYP4* families share four introns, and two of those four shared introns are also found in *CYP11*. Conclusions drawn from intron–exon boundaries, in combination with the unweighted pair-group method of analysis, are remarkably similar (Nebert and Nelson, 1991). Thus, comparative data about intron/exon organization might also be considered, in addition to standard amino acid alignments, when one draws conclusions about P450 evolution (Nebert and Gonzalez, 1987).

There are currently 221 P450 genes and 12 putative pseudogenes that have been described in 31 eukaryotes (including 11 mammalian and 3 plant species) and in 11 prokaryotes (Table 2). Of the 36 gene families so far described, 12 exist in all mammals examined to date. These 12 families comprise 22 mammalian subfamilies, or clusters of genes, of which 17 and 15 have been mapped in the human and murine genome, respectively (Table 1). Previous nomenclature updates did not use exhaustive literature search techniques, which permitted many relevant references to be overlooked. By searching the three major protein and nucleic acid sequence databases, we have now found many of these references which are included in Table 2.

This P450 nomenclature system affords a convenient medium for colleagues across distantly related fields to converse with one another. What follows is a discussion of several potential problems that may arise from the application of this system of nomenclature, as is true with any phylogenetic analysis based on evolution.

UNCERTAINTY ABOUT ORTHOLOGOUS GENES

An "orthologous gene" in two species refers to a gene that we know with certainty corresponds to the ancestral gene which existed before evolutionary divergence of the two species. In several subfamilies (*CYP2A, 2B, 2C, 2D, 3A, 4A,* and *11B*), numerous species-specific gene duplications, nonequal crossing-over, and gene conversion events have made orthologue assignments between species impossible. "Gene conversion" refers to an event during replication in which some portion of a gene is replaced by the corresponding part of a nearby gene (or pseudogene). Where the classification of proteins encoded by orthologous genes cannot be certain—particularly between widely diverged species and especially in subfamilies containing three or

TABLE 2. UPDATE OF ALL *CYP* GENES AND THEIR PRODUCTS[a]

Gene symbol	*Trivial name*	*Species*	*References*
CYP1A1	c, βNF-B	Rat	Botelho *et al.* (1979) Haniu *et al.* (1984a) Haniu *et al.* (1984c) Sogawa *et al.* (1984) Yabusaki *et al.* (1984a) Yabusaki *et al.* (1984b) Hines *et al.* (1985) Oeda *et al.* (1985) Cheng *et al.* (1986) Fagan *et al.* (1986) Sogawa *et al.* (1986)
	P_1, c, form 6	Human	Jaiswal *et al.* (1985a) Jaiswal *et al.* (1985b) Quattrochi *et al.* (1985) Kawajiri *et al.* (1986) Wrighton *et al.* (1986a) Hayashi *et al.* (1991b) Kubota *et al.* (1991) Shimada *et al.* (1992)
	Form 6	Rabbit	Okino *et al.* (1985) Kagawa *et al.* (1987)
	IA1	Trout	Heilmann *et al.* (1988)
	Dah1	Dog	Uchida *et al.* (1990)
	IA1	Hamster	Sagami *et al.* (1991)
	HSc 1		Ohgiya (1992b) [GenEMBL D12977]
	MKah1	Monkey	Komori *et al.* (1992a)
	GP 53K	Guinea pig	Cheng *et al.* (1986)
	GPc1		Ohgiya *et al.* (1992b) [GenEMBL D11043]
Cyp1a-1	$P_1$450	Mouse	Kimura, S., *et al.* (1984a) Kimura and Nebert (1985) Gonzalez *et al.* (1985a) Cheng *et al.* (1986) Kimura, S., *et al.* (1987b)
CYP1A2	P-448, d, HCB	Rat	Botelho *et al.* (1982) Haniu *et al.* (1984a) Haniu *et al.* (1984b) Kawajiri *et al.* (1984) Yabusaki *et al.* (1984a) Sogawa *et al.* (1985) Cheng *et al.* (1986) Fagan *et al.* (1986) Haniu *et al.* (1986)
	P_3, d, form 4	Human	Quattrochi *et al.* (1985) Jaiswal *et al.* (1986) Quattrochi *et al.* (1986) Wrighton *et al.* (1986a) Jaiswal *et al.* (1987a) Ikeya *et al.* (1989) Quattrochi and Tukey (1989)
	LM_4	Rabbit	Fujita *et al.* (1984) Okino *et al.* (1985) Ozols (1986) Kagawa *et al.* (1987) Pompon (1988)
	MC4	Hamster	Koga *et al.* (1990) Lai and Chiang (1990) Sagami *et al.* (1991)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
	MKah2	Monkey	Komori *et al.* (1992b)
	P-450-D3	Dog	Ohta *et al.* (1989b)
	P-450-D2		Ohta *et al.* (1990)
	Dah2		Uchida *et al.* (1990)
	pP-450IA-61	Chicken	Murti *et al.* (1991)
Cyp1a-2	P_3	Mouse	Kimura, S., *et al.* (1984a) Kimura, S., *et al.* (1984b) Gonzalez *et al.* (1985a) Cheng *et al.* (1986) Ikeya *et al.* (1989)
	P_2		Kimura and Nebert (1986)
CYP2A1	a1, a, 3, UT-F, RLM2b, IF-3	Rat	Botelho *et al.* (1979) Haniu *et al.* (1984a) Imaoka *et al.* (1987) Nagata *et al.* (1987) Arlotto *et al.* (1989) Matsunaga, T., *et al.* (1990)
CYP2A2	a2, RLM2, UT-4	Rat	Jansson *et al.* (1985) Funae and Imaoka (1987) Matsunaga, T., *et al.* (1988) Arlotto *et al.* (1989) Matsunaga, T., *et al.* (1990)
CYP2A3	a3	Rat	Kimura, S., *et al.* (1989c) Ueno and Gonzalez (1990)
Cyp2a-4	15αoh-1	Mouse	Squires and Negishi (1988) Lindberg *et al.* (1989) Lindberg and Negishi (1989) Burkhart *et al.* (1990)
Cyp2a-5	15αoh-2	Mouse	Squires and Negishi (1988) Lindberg *et al.* (1989) Burkhart *et al.* (1990)
Cyp2a-4 or *Cyp2a-5*			Lang *et al.* (1989) Lange *et al.* (1990)
CYP2A6	IIA3, P450(1), IIA4	Human	Phillips *et al.* (1985a) Landsman (1989) [GenEMBL X13930) Miles *et al.* (1989a) Yamano *et al.* (1989a) Miles *et al.* (1990a) Yamano *et al.* (1990) Yun *et al.* (1991)
CYP2A7	IIA4	Human	Yamano *et al.* (1990)
CYP2A8	AFB, MC1	Hamster	Fukuhara *et al.* (1989) Koga *et al.* (1990) Lai and Chiang (1990)
CYP2A9	MC1-81	Hamster	Lai and Chiang (1990)
CYP2A10	NMa	Rabbit	H.M. Peng, X. Ding, M.J. Coon (personal communication)
CYP2A11	NMc	Rabbit	H.M. Peng, X. Ding, M.J. Coon (personal communication)
CYP2A	(Fragment)	Cow	Lazard *et al.* (1990)
CYP2B1	b, PB-4, PB-B, PBRLM5	Rat	Botelho *et al.* (1979) Fujii-Kuriyama *et al.* (1982a) Fujii-Kuriyama *et al.* (1982b) Fujii-Kuriyama *et al.* (1982c)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
			Waxman and Walsh (1982)
			Gotoh *et al.* (1983)
			Yuan *et al.* (1983b)
			Backes *et al.*, 1985)
			Scholte *et al.* (1985)
			Suwa *et al.* (1985)
			Hashimoto *et al.* (1988)
			Jean *et al.* (1988)
			Oesch *et al.* (1989)
	IIB1-WM (variant)		Kedzie *et al.* (1991)
CYP2B2	3, PB-5, PB-D, PBRLM6	Rat	Fujii-Kuriyama *et al.* (1982a)
			Waxman and Walsh (1982)
			Kumar *et al.* (1983)
			Mizukami *et al.* (1983)
			Phillips *et al.* (1983)
			Yuan *et al.* (1983b)
			Affolter and Anderson (1984)
			Kumar *et al.* (1984)
			Backes *et al.* (1985)
			Frey *et al.* (1985)
			Ravishankar and Padmanaban (1985)
			Scholte *et al.* (1985)
			Atchison and Adesnik (1986)
			Jaiswal *et al.* (1987b)
			Rangarajan *et al.* (1987)
			Jean *et al.* (1988)
			Oesch *et al.* (1989)
			Rangarajan and Padmanaban (1989)
			Lacroix *et al.* (1990)
CYP2B3	IIB3	Rat	Affolter *et al.* (1986)
			Labbé *et al.* (1988)
			Jean *et al.* (submitted)
CYP2B4	LM2	Rabbit	Ozols *et al.* (1981)
			Heinemann and Ozols (1982)
			Heinemann and Ozols (1983)
			Tarr *et al.* (1983)
	B0, B1		Gasser *et al.* (1988)
	b14, b46, b54		Komori *et al.* (1988b)
CYP2B4P	(Pseudogene)	Rabbit	Zaphiropoulos *et al.* (1986)
CYP2B5	b52, HP1	Rabbit	Komori *et al.* (1988b)
	B2		Gasser *et al.* (1988)
CYP2B6	LM2, IIB1, hIIB	Human	Miles *et al.* (1988)
			Miles *et al.* (1989b)
			Yamano *et al.* (1989b)
			Miles *et al.* (1990b)
CYP2B7P	IIB2 (pseudogene)	Human	Yamano *et al.* (1989b)
CYP2B8	gene IV	Rat	Giachelli *et al.* (1989)
Cyp2b-9	pf26	Mouse	Noshiro *et al.* (1988)
Cyp2b-10	pf3/46	Mouse	Noshiro *et al.* (1988)
CYP2B11	IIB	Dog	Graves *et al.* (1990)
CYP2B12	IIB-gene 4	Rat, preputial gland	Friedberg *et al.* (1992)
Cyp2b-13	16αoh-b	Mouse	Stupans *et al.* (1984)
			Lakso *et al.* (1991)
CYP2B14	2By	Rat	Jean *et al.* (submitted)
CYP2B14P	2Bx (pseudogene)	Rat	Jean *et al.* (submitted)
CYP2B	(Fragment)	Rat	Shayiq and Avadhani (1990)
CYP2B	(Fragment)	Guinea pig	Narimatsu *et al.* (1990)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
			Oguri *et al.* (1991)
CYP2B	(Fragment)	Sheep	Kaddouri *et al.* (1992)
CYP2C1	PBc1	Rabbit	Leighton *et al.* (1984) Zhao *et al.* (1990) Kim and Kemper (1991)
CYP2C2	PBc2, K, pHP2	Rabbit	Leighton *et al.* (1984) Govind *et al.* (1986) Imai *et al.* (1988) Kim and Kemper (1991)
CYP2C3	PBc3	Rabbit	Ozols *et al.* (1981) Leighton *et al.* (1984)
	3b		Ozols *et al.* (1985) Chan and Kemper (1990)
	2C3v		E.F. Johnson (personal communication)
CYP2C4	1-88 PBc4	Rabbit	Johnson *et al.* (1987) Zhao *et al.* (1987) Zhao *et al.* (1990)
CYP2C5	Form 1	Rabbit	Tukey *et al.* (1985) Johnson *et al.* (1987) Pendurthi *et al.* (1990) Zhao *et al.* (1990)
CYP2C6	PB1, k, PB-C, *pTF2*, RLM5a	Rat	Waxman and Walsh (1983) Gonzalez *et al.* (1986a) Friedberg *et al.* (1986)
	PB2		Funae and Imaoka (1987) Schenkman *et al.* (1987)
	2C6 (Product of alternative splicing)		Kimura, H., *et al.* (1988) Umeno *et al.* (1988b) Kimura, H., *et al.* (1989)
CYP2C7	f, RLM5b	Rat	Haniu *et al.* (1984a) Gonzalez *et al.* (1986a)
	pTF1		Friedberg *et al.* (1986) Favreau *et al.* (1987) Kimura, H., *et al.* (1988) M.C. Lechner (personal communication)
CYP2C8	IIC2 Form 1 mp-12, mp-20 hP2-1	Human	Kimura, S., *et al.* (1987a) Okino *et al.* (1987) Ged *et al.* (1988) Komori *et al.* (1989a) Shephard *et al.* (1989)
	HPH pB8		Kolyada (1990) Ged and Beaune (1991)
CYP2C9	MP-1, MP-2 IIC1 human-2 mp-4 HM2 pHLS.5 hPA6	Human	Shimada *et al.* (1986) Kimura, S., *et al.* (1987a) Yasumori *et al.* (1987) Ged *et al.* (1988) Komori *et al.* (1988a) Meehan *et al.* (1988a) Ohgiya *et al.* (1989) Romkes *et al.* (1991) Veronese *et al.* (1991)
	hPA22		Ohgiya *et al.* (1992a)
CYP2C10	mp, mp-8, [?]cloning artifact of *CYP2C9*		Umbenhauer *et al.* (1987) Ged *et al.* (1988) Srivastava *et al.* (1991)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
CYP2C11	h, M-1, 16α, 2c, UT-A, RLM5	Rat	Cheng and Schenkman (1982) Haniu *et al.* (1984a) Waxman (1984) Yoshioka *et al.* (1987)
	male, UT-2		Morishima *et al.* (1987) Ström *et al.* (1988) Zaphiropoulos *et al.* (1988)
CYP2C12	i, 15β, 2d, UT-I	Rat	Haniu *et al.* (1984a) Zaphiropoulos *et al.* (1988)
	female, F-2		Zaphiropoulos *et al.* (1990b)
CYP2C13	+g, −g, RLM3	Rat	Cheng and Schenkman (1982) Haniu *et al.* (1984a) Matsumoto *et al.* (1986)
	UT-5		Funae and Imaoka (1987) McClellan-Green *et al.* (1989) Yeowell *et al.* (1990) Zaphiropoulos *et al.* (1990a) Eguchi *et al.* (1991)
CYP2C14	pHP3	Rabbit	Imai (1987)
CYP2C15	b32-3	Rabbit	Imai *et al.* (1988)
CYP2C16	IIC16	Rabbit	Hassett and Omiecinski (1987) Hassett and Omiecinski (1990)
CYP2C17	254c [?]splice variant of *CYP2C18/CYP2C19*		Romkes *et al.* (1991) Romkes *et al.* (1992)
CYP2C18	29c, 6b	Human	Romkes *et al.* (1991) Furuya *et al.* (1991) Ged and Beaune (1992) Nunoya (1992) [GenEMBL X63904]
CYP2C19	11a	Human	Romkes *et al.* (1991)
CYP2C20	MKmp13	Monkey	Ohta *et al.* (1989a) Komori *et al.* (1992b)
CYP2C21	DM 1-1	Dog	Uchida *et al.* (1990)
CYP2C22	Md	Rat	Emi *et al.* (1990) Nagata *et al.* (1990b)
CYP2C23	c1 17	Rat	Cook *et al.* (1990)
CYP2C24	IIC24	Rat prostate	Zaphiropoulos (1991)
CYP2C25	hsm1	Hamster	Sakuma *et al.* (1992) [GenEMBL X63022]
CYP2C26	hsm2	Hamster	Sakuma *et al.* (1992) [GenEMBL D11435]
CYP2C27	hsm3	Hamster	Sakuma *et al.* (1992) [GenEMBL D11436]
CYP2C28	hsm4	Hamster	Sakuma *et al.* (1992) [GenEMBL D11437]
CYP2C	HM2 (Fragment)	Human	Komori *et al.* (1988a)
CYP2C	(Fragment)	Rat	Imaoka *et al.* (1990a)
Cyp2c	(Fragment)	Mouse	Watanabe *et al.* (1991)
CYP2D1	UT-7 db1 CMF1a	Rat	Funae and Imaoka (1987) Gonzalez *et al.* (1987) Ishida *et al.* (1988b) Matsunaga, E., *et al.* (1989)
CYP2D2	db2 CMF2	Rat	Gonzalez *et al.* (1987) Ishida *et al.* (1988b) Matsunaga, E., *et al.* (1989) Matsunaga, E., *et al.* (1990)
CYP2D3	db3	Rat	Matsunaga, E., *et al.* (1989) Matsunaga, E., *et al.* (1990)
CYP2D4	CMF3 db4	Rat	Ishida *et al.* (1988b) Matsunaga, E., *et al.* (1990)
CYP2D5	CMF1b	Rat	Ishida *et al.* (1988b) Ishida *et al.* (1989)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
	db5		Matsunaga, E., *et al.* (1989) Matsunaga, E., *et al.* (1990)
CYP2D6	db1	Human	Gonzalez *et al.* (1988b) Gonzalez *et al.* (1988c) Kimura, S., *et al.* (1989d) Manns *et al.* (1989) Gough *et al.* (1990)
CYP2D7P	IID7 (pseudogene)	Human	Kimura, S., *et al.* (1989d)
CYP2D8P	IID8 (pseudogene)	Human	Kimura, S., *et al.* (1989d)
Cyp2d-9	16α, ca	Mouse	Wong *et al.* (1987) Ichikawa *et al.* (1989) Wong *et al.* (1989) Yoshioka *et al.* (1990)
Cyp2d-10	cb	Mouse	Ichikawa *et al.* (1989) Wong *et al.* (1989) Yoshioka *et al.* (1990)
Cyp2d-11	cc	Mouse	Wong *et al.* (1989)
Cyp2d-12	cd	Mouse	M. Negishi (personal communication)
Cyp2d-13	ce	Mouse	M. Negishi (personal communication)
CYP2D14	2D	Cow	Tsuneoka *et al.* (1992)
CYP2D	(Fragment)	Rat	Sugita *et al.* (1988)
CYP2D	(Fragment)	Rat	Suzuki *et al.* (1992)
CYP2E1	j	Human	Song *et al.* (1986) Wrighton *et al.* (1986a) Wrighton *et al.* (1986b) Lasker *et al.* (1987) Wrighton *et al.* (1987a) Umeno *et al.* (1988a) Hayashi *et al.* (1991a) Uematsu *et al.* (1991) Adams *et al.* (1992)
	j, RLM6	Rat	Ryan *et al.* (1985) Song *et al.* (1986) Favreau *et al.* (1987) Wrighton *et al.* (1987b) Umeno *et al.* (1988c)
	DM		Imaoka *et al.* (1990b)
	3a	Rabbit	Khani *et al.* (1987) Imai *et al.* (1988) Khani *et al.* (1988a) Khani *et al.* (1988b)
	MKj1	Monkey	Komori *et al.* (1992b)
Cyp2e-1	j	Mouse	Voliva and Paigen (1991) Freeman *et al.* (1992)
CYP2E2	3d,IIE2	Rabbit	Khani *et al.* (1988a) Khani *et al.* (1988b)
CYP2F1	IIF1	Human	Nhamburo *et al.* (1990)
Cyp2f-2	Nah-2	Mouse	Ritter *et al.* (1991)
CYP2G1	olf1	Rat	Nef *et al.* (1989) Nef *et al.* (1990)
	NMb	Rabbit	Ding *et al.* (1991b)
CYP2H1	pCHP3, PB15	Chicken	Hobbs *et al.* (1986) Hansen and May (1989) Sinclair *et al.* (1990)
CYP2H2	pCHP7	Chicken	Hansen and May (1989) Sinclair *et al.* (1990)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
CYP2J1	ib	Rabbit intestine	Kikuta *et al.* (1991)
CYP2K1	LMC2	Trout	D.R. Buhler (personal communication)
?CYP2 EST	14F12	*C. elegans*	Waterston *et al.* (1992)
CYP3A1	pcn1, PCNa	Rat	Gonzalez *et al.* (1985) Graves *et al.* (1987) Halpert (1988)
	6β-4		Nagata *et al.* (1990a) Burger *et al.* (1992)
	pIGC2		Ribeiro and Lechner (1992)
CYP3A2	pcn2 PCNb/c	Rat	Gonzalez *et al.* (1986b) Graves *et al.* (1987) Halpert (1988)
	PB-1 6β-1/3		Imaoka *et al.* (1988b) Nagata *et al.* (1990a) Miyata *et al.* (1991) M.C. Lechner (personal communication)
CYP3A3	HLp	Human	Watkins *et al.* (1985) Molowa *et al.* (1986)
CYP3A4	nf-25 hPCN1 nf-10	Human	Beaune *et al.* (1986) Gonzalez *et al.* (1988a) Bork *et al.* (1989) Hashimoto (1992) [GenEMBL D11131]
CYP3A3 or *CYP3A4*		Human	Komori *et al.* (1988a)
CYP3A5	hPCN3 HLp2	Human	Aoyama *et al.* (1989) Schuetz *et al.* (1989)
CYP3A5P	3A5P (pseudogene)	Human	J.D. Schuetz and P.S. Guzelian (personal communication)
CYP3A6	3c	Rabbit	Dalet *et al.* (1988) Potenza *et al.* (1989)
CYP3A7	HFLa HFL33 HLp2	Human	Kitada *et al.* (1987) Komori *et al.* (1988a) Wrighton and Vandenbranden (1989) Komori *et al.* (1989a) Komori *et al.* (1989b) Itoh *et al.* (1992)
CYP3A8	MKnf2	Monkey	Ohta *et al.* (1989a) Komori *et al.* (1992b)
CYP3A9	olf3	Rat	P. Nef (personal communication)
CYP3A10		Hamster	Teixeira and Gil (1991)
Cyp3a-11	IIIAm1	Mouse	Yanagimoto *et al.* (1992)
CYP3A12	PBD-1	Dog	Ciaccio *et al.* (1991)
Cyp3a-13	IIIAm2	Mouse	Yanagimoto (1991) [GenEMBL X63023]
CYP3A	(Fragment)	Rat	Nagata *et al.* (1990a)
CYP3A	(Fragment)	Sheep	Kaddouri *et al.* (1992)
CYP4A1	LA_{ω}	Rat	Hardwick *et al.* (1987) Earnshaw *et al.* (1988) Kimura, S., *et al.* (1989a)
CYP4A2	IVA2, k-5	Rat	Yoshimoto *et al.* (1986) Kimura, S., *et al.* (1989a)
	k-2		Kawashima *et al.* (1992)
CYP4A3	IVA3, DM-2	Rat	Imaoka and Funae (1986) Imaoka *et al.* (1988a) Kimura, S., *et al.* (1989b) Imaoka *et al.* (1990b)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
CYP4A4	p-2	Rabbit	Matsubara *et al.* (1987)
	LPGω		Kikuta *et al.* (1989)
CYP4A5	KDB3	Rabbit	Johnson *et al.* (1990)
	kd		Yoshimura *et al.* (1990)
			Yokotani *et al.* (1991)
CYP4A6	ka-1	Rabbit kidney	Kusunose *et al.* (1989)
			Yokotani *et al.* (1989)
	KDA6		Johnson *et al.* (1990)
	LPGAω-1	Rabbit liver	Kikuta *et al.* (1990)
			Muerhoff *et al.* (1992)
CYP4A7	ka-2	Rabbit kidney	Kusunose *et al.* (1989)
	R4		Yokotani *et al.* (1989)
			Johnson *et al.* (1990)
	LPGAω-2	Rabbit liver	Kikuta *et al.* (1990)
CYP4A8	PP1	Rat	Imaoka *et al.* (1990a)
			Strömstedt *et al.* (1990)
CYP4A9	HL14Acon	Human	J.P. Hardwick (personal communication)
Cyp4a-10	A14	Mouse	C.J. Henderson (personal communication)
	clone 1		D.R. Bell (personal communication)
CYP4A11	HK_ω	Human	Kawashima *et al.* (1992)
			F.J. Gonzalez (personal communication)
			E.F. Johnson (personal communication)
Cyp4a-12	clone 2	Mouse	D.R. Bell (personal communication)
CYP4A13		Guinea pig	D.R. Bell (personal communication)
CYP4B1	Lung P450	Human	Nhamburo *et al.* (1989)
	p-2-like		Yokotani *et al.* (1990)
	HLCF1		J.P. Hardwick (personal communication)
	Form 5	Rabbit	Gasser and Philpot (1989)
	Form 5	Rat	Gasser and Philpot (1989)
	L-2		Imaoka and Funae (1990)
CYP4C1	P-450	*Blaberus discoidalis* (cockroach)	Bradfield *et al.* (1991)
CYP4D1	4D1	*Drosophila melanogaster* (fruit fly)	Gandhi *et al.* (1992)
CYP4E1	In cuticle gene cluster	*D. melanogaster*	Snyder *et al.* (1981)
			Snyder and Davidson (1983)
CYP4F1	A3	Rat	Chen and Hardwick (1992)
CYP4F2	4F2	Human	L. Chen and J.P. Hardwick (personal communication)
CYP4F3	LTBω	Human	M. Kusunose (personal communication)
?CYP4 EST	8B12	*Caenorhabditis elegans*	Waterston *et al.* (1992)
?CYP4 EST	wEST00713	*C. elegans*	McCombie *et al.* (1992)
CYP5	TXA synthase	Human	Nüsing *et al.* (1990)
			Yokoyama *et al.* (1991)
			Ohashi *et al.* (1992)
CYP6A1	VIA1	*Musca domestica* (house fly)	Feyereisen *et al.* (1989)
CYP6A2	DM P450-B1	*D. melanogaster*	Waters *et al.* (1992)
CYP6B1	CYP6B1v1, CYP6B1v2	*Papilio polyxenes* (black swallowtail butterfly)	Cohen, M.B., *et al.* (1992)
CYP7	7α	Rat	Noshiro *et al.* (1989)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
			Jelinek and Russell (1990)
			Jelinek *et al.* (1990)
			Li *et al.* (1990)
			Noshiro *et al.* (1990)
			Nishimoto *et al.* (1991)
			Chiang *et al.* (1992)
		Human	Noshiro and Okuda (1990)
			Molowa *et al.* (1992)
			Cohen, J.C., *et al.* (1992)
			Karam and Chiang (1992)
		Rabbit	D.F. Jelinek and D.W. Russell (personal communication)
		Cow	D.F. Jelinek and D.W. Russell (personal communication)
CYP10	P-450	*Lymnea stagnalis* (pond snail)	Teunissen *et al.* (1992)
CYP11A1	scc	Human	Chung *et al.* (1986b)
			Matteson *et al.* (1986a)
			Morohashi *et al.* (1987a)
			Chung (1989) [GenEMBL X14257]
			Moore *et al.* (1990)
			Inoue *et al.* (1991)
			Lin *et al.* (1991a)
			Moore *et al.* (1992)
			Hum *et al.* (1993)
		Cow	Ogishima *et al.* (1983)
			Morohashi *et al.* (1984)
			Chung *et al.* (1985b)
			Chashchin *et al.* (1986)
			Adamovich *et al.* (1989)
			Ahlgren *et al.* (1990)
		Pig	Mulheron *et al.* (1989)
			Iwahashi *et al.* (1991)
			Kuwada *et al.* (1991)
		Rat	McMasters *et al.* (1987)
			Oonk *et al.* (1989)
			Oonk *et al.* (1990)
		Chicken	Mathew *et al.* (1990)
		Trout	Takahashi *et al.* (1993)
CYP11B1	11β	Human	Chua *et al.* (1987)
			Mornet *et al.* (1989)
			Kawamoto *et al.* (1990b)
			White *et al.* (1991)
			Kawamoto *et al.* (1992)
	CB11β-7	Cow	Ogishima *et al.* (1983)
			Chua *et al.* (1987)
			Morohashi *et al.* (1987b)
			Kirita *et al.* (1988)
			Hashimoto *et al.* (1989)
	pcP-450(11β)-3		Kirita *et al.* (1990)
	11β-4		Mathew *et al.* (1991)
	11β	Rat	Nonaka *et al.* (1989)
			Ogishima *et al.* (1989)
			Malee and Mellon (1991)
			Mukai *et al.* (1991)
			Nomura *et al.* (submitted)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
Cyp11b-1	11β	Mouse	Mouw *et al.* (1989) Domalik *et al.* (1991)
CYP11B2	11β-2	Human	Mornet *et al.* (1989) Kawamoto *et al.* (1990a) Kawamoto *et al.* (1992)
	Aldo-2	Rat	Matsukawa *et al.* (1990) Malee and Mellon (1991)
	11β		Imai *et al.* (1990)
	Aldo		Mukai *et al.* (1991) Nomura *et al.* (submitted)
		Cow	Kirita *et al.* (1990)
Cyp11b-2	Aldo synthase	Mouse	Domalik *et al.* (1991)
CYP11B3	B3	Rat	K. Mukai (personal communication)
CYP11B4	CB11β-20	Cow	Kirita *et al.* (1990)
CYP11B5P	CB11β-1 (pseudogene)		Kirita *et al.* (1990)
CYP11B6P	CB11β-3 (pseudogene)		Kirita *et al.* (1990)
CYP11B7P	λB11β(15-1), (15-2) (pseudogene)		Hashimoto *et al.* (1989)
	CB11β-21		Kirita *et al.* (1990)
CYP11B8P	B4 (pseudogene)	Rat	K. Mukai (personal communication)
CYP17	17α	Human	Chung *et al.* (1987) Picado-Leonard and Miller (1987) Bradshaw *et al.* (1987) Kagimoto *et al.* (1988) Brentano *et al.* (1990) Lin *et al.* (1991b)
		Cow	Zuber *et al.* (1986) Bhasker *et al.* (1989)
		Pig	Nakajin *et al.* 1984) Chung *et al.* (1987) Conley *et al.* (1992) Zhang *et al.* (1992)
		Chicken	Ono *et al.* (1988)
		Rat	Namiki *et al.* (1988) Nishihara *et al.* (1988) Fevold *et al.* (1989) Mellon and Vaisse (1989) Zhang *et al.* (1992)
		Trout	Sakai *et al.* (1992)
		Guinea pig	Ogishima *et al.* (1983)
Cyp-17	17α	Mouse	Youngblood *et al.* (1991) Youngblood and Payne (1992)
CYP19	Arom	Human	Chen, S., *et al.* (1986) Evans *et al.* (1986) Hall *et al.* (1987) Simpson *et al.* (1987) Chen, S., *et al.* (1988) Corbin *et al.* (1988) Harada (1988) Means *et al.* (1989) Pompon *et al.* (1989) Toda *et al.* (1989) Harada *et al.* (1990) Toda *et al.* (1990) Mahendroo *et al.* (1991) Means *et al.* (1991)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
			Harada *et al.* (1992)
		Chicken	McPhaul *et al.* (1988)
			Matsumine *et al.* (1991)
		Rat	Hickey *et al.* (1990)
			Lephart *et al.* (1990)
		Trout	Tanaka *et al.* (1992)
		Goldfish	G. Callard (personal communication)
Cyp19	gES-M10	Mouse	Terashima *et al.* (1991)
CYP21A1	c21	Cow	White *et al.* (1984b)
			Chung *et al.* (1985a,b)
			Chung *et al.* (1986a)
			John *et al.* (1986)
			Yoshioka *et al.* (1986)
		Pig	Yuan *et al.* (1983a)
			Bienkowski *et al.* (1984)
			Chung *et al.* (1987)
			Haniu *et al.* (1987)
			Burghelle-Mayeur *et al.* (1992)
		Sheep	Crawford *et al.* (1992)
Cyp21a-1	c21A	Mouse	Amor *et al.* (1985)
			Chaplin *et al.* (1986)
			Ogata and Zepf (1991)
CYP21A1P	(pseudogene c21A)	Human	Higashi *et al.* (1986)
			White *et al.* (1986)
			Morel *et al.* (1989)
			Gitelman *et al.* (1992)
			Bristow *et al.* (1993)
CYP21A2	c21B	Human	Carroll *et al.* (1985)
			Higashi *et al.* (1986)
			White *et al.* (1986)
			Matteson *et al.* (1987)
			Rodrigues *et al.* (1987)
			Amor *et al.* (1988)
			Morel *et al.* (1989)
			Chiou *et al.* (1990)
			Gitelman *et al.* (1992)
			Helmberg *et al.* (1992)
			Owerbach *et al.* (1992)
Cyp21a-2p	(pseudogene c21B)	Mouse	Chaplin *et al.* (1986)
CYP24	cc24	Rat	Ohyama *et al.* (1991)
CYP27	26-ohp	Rabbit	Dahlbäck (1988)
			Andersson *et al.* (1989)
			Dahlbäck (1989)
	25-hydroxylase	Rat	Shayiq and Avadhani (1989)
			Usui *et al.* (1990b)
	P-450 26/25		Su *et al.* (1990)
	27-hydroxylase	Human	Cali and Russell (1991)
			Cali *et al.* (1991)
			Meiner *et al.* (1992)
CYP51	14DM	*Saccharomyces cerevisiae*	Kalb *et al.* (1986)
			Kalb *et al.* (1987)
			Ishida *et al.* (1988a)
	ERG11		Turi *et al.* (1991)
		Candida tropicalis	Chen, C., *et al.* (1987)
			Chen, C., *et al.* (1988)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
		C. albicans	Lai and Kirsch (1989)
CYP52A1	alk1	*C. tropicalis*	Sanglard *et al.* (1987) Sanglard and Loper (1989) Sanglard and Fiechter (1989) Seghezzi *et al.* (1991)
CYP52A2	alk2	*C. tropicalis*	Sanglard and Fiechter (1989) Seghezzi *et al.* (1991)
CYP52A3	Cm1 ALK1-A ALK1-B	*C. maltosa*	Schunck *et al.* (1989) Takagi *et al.* (1989) Ohkuma *et al.* (1991a) Ohkuma *et al.* (1991b) Schunck *et al.* (1991)
CYP52A4	Cm2 ALK3-A, ALK3-B	*C. maltosa*	Schunck *et al.* (1991) Ohkuma *et al.* (1991b)
CYP52A5	ALK2-A, ALK2-B	*C. maltosa*	Ohkuma *et al.* (1991b)
CYP52A6	alk3	*C. tropicalis*	Seghezzi *et al.* (1992) [GenEMBL Z13010]
CYP52A7	alk4	*C. tropicalis*	Seghezzi *et al.* (1992) [GenEMBL Z13011]
CYP52A8	alk5	*C. tropicalis*	Seghezzi *et al.* (1992) [GenEMBL Z13012]
CYP52A9	ALK5-A	*C. maltosa*	Ohkuma *et al.* (1993)
CYP52A10	ALK7-A	*C. maltosa*	M. Takagi (personal communication) [GenEMBL D12719]
CYP52A11	ALK8-A	*C. maltosa*	M. Takagi (personal communication) [GenEMBL D12719]
CYP52B1	alk6	*C. tropicalis*	D. Sanglard (personal communication) [GenEMBL Z13013]
CYP52C1	alk7	*C. tropicalis*	D. Sanglard (personal communication) [GenEMBL Z13014]
CYP52C2	ALK6-A	*C. maltosa*	M. Takagi (personal communication) [GenEMBL D12718]
CYP52D1	ALK4-A	*C. maltosa*	M. Takagi (personal communication) [GenEMBL D12716]
CYP53	bphA	*Aspergillus niger*	van Gorcom *et al.* (1990) Boschloo *et al.* (1991)
CYP54	CI-1	*Neurospora crassa*	Attar *et al.* (1989)
CYP55	dNIR	*Fusarium oxysporum*	Kizawa *et al.* (1991)
CYP56	DIT2	*Saccharomyces cerevisiae*	Briza *et al.* (1990)
CYP57	*pda*T9	*Nectria haematococca*	A. Maloney and H.D. VanEtten (personal communication)
CYP71		*Persea americana* (avocado)	Bozak *et al.* (1990)
CYP72	P450	*Catharanthus roseus* (Madagascar periwinkle)	J. Schröder (personal communication)
CYP73	Ca4h	*Helianthus tuberosus* (Jerusalem artichoke)	Gabriac *et al.* (1991)
CYP101	cam	*Pseudomonas putida*	Haniu *et al.* (1982a) Haniu *et al.* (1982b)

TABLE 2. (CONTINUED)

Gene symbol	*Trivial name*	*Species*	*References*
			Poulos *et al.* (1985) Poulos *et al.* (1987) Unger *et al.* (1986)
CYP102	BM-3	*Bacillus megaterium*	Narhi and Fulco (1986) Narhi and Fulco (1987) Ruettinger *et al.* (1989)
CYP103	pinF1	*Agrobacterium tumefaciens*	Kanemoto *et al.* (1989)
CYP104	pinF2	*A. tumefaciens*	Kanemoto *et al.* (1989)
CYP105A1	SU1	*Streptomyces griseolus*	Omer *et al.* (1990)
CYP105B1	SU2	*S. griseolus*	Omer *et al.* (1990)
CYP105C1	choP	*Streptomyces* spp.	Ishizaki *et al.* (1989) Horii *et al.* (1990)
CYP105D1	soy	*S. griseus*	Trower *et al.* (1992) Sariaslani and Omer (1992)
CYP106	BM-1	*B. megaterium*	He *et al.* (1989)
CYP107A1	*eryF*	*Saccharopolyspora erythraea*	Haydock *et al.* (1991) Weber *et al.* (1991)
CYP107B1	orf405	*S. erythraea*	Andersen and Hutchinson (1992)
CYP108	terp	*Pseudomonas* spp.	Peterson *et al.* (1992)
CYP109	ORF405	*Bacillus subtilis*	Lewis and Wake (1989) Ahn and Wake (1991)
CYP110	ORF3	*Anabaena* spp.	Lammers *et al.* (1990)
CYP111	lin	*Pseudomonas* spp.	I.C. Gunsalus (personal communication)
CYP112	BJ-1	*Bradyrhizobium japonicum*	R.E. Tully and D.L. Keister (1992) [GenEMBL L02323]
Unidentified fragment		Chicken	Gupta *et al.* (1990) Sinclair *et al.* (1990)
Unidentified fragment		Rat	Imaoka *et al.* (1990b)
Unidentified fragment		Chicken	Sinclair *et al.* (1990)
Unidentified fragment		Chicken	Sinclair *et al.* (1990)
Unidentified fragment		Pig	Bergman and Postlind (1990)
Unidentified fragment		Pig	Bergman and Postlind (1991)
Unidentified fragment		Human	Hu *et al.* (1991)

[a]References followed by Accession Numbers are not otherwise published.

more genes—sequential numbering is recommended, usually on a chronological basis, as the protein sequences become available in the literature (*e.g.*, *CYP2C1, CYP2C2, CYP2C3, CYP2C4,...CYP2C28*). The same rules might become necessary for genes in other subfamilies, once it becomes apparent that three or more closely related genes exist in that subfamily.

On the basis of amino acid similarities and in some cases cDNA-expressed catalytic activities, we can be certain about the orthologous *CYP1A1* and the orthologous *CYP1A2* genes in all mammals examined to date (Nebert and Nelson, 1991). On the other hand, trout appears to have only the *CYP1A1* gene in its *CYP1* family, suggesting that *CYP1A2* arose *via* a gene duplication event in land

animals following the divergence of land animals and sea animals approximately 370 million years ago (Heilmann *et al.*, 1988).

DROSOPHILA P450 SEQUENCE REPORTED IN 1983

One of the earliest P450 sequences published, other than amino-terminal sequences, has only recently been uncovered by Dr. Debra Nero, who was searching GenBank for alignments with the cockroach *CYP4C1* sequence. The most significant match was to an unknown *Drosophila* genomic sequence encoding 153 amino acids (Snyder and Davidson, 1983). This sequence is definitely a P450 carboxy-terminal region, including the heme-binding motif; obviously, at the time this sequence was published, it could not be recognized as a P450 protein. Figure 2c of the Snyder and Davidson (1983) paper shows a nucleic acid sequence encoding the carboxy-terminal half of a P450 protein that is homologous after translation to amino acids 306–531 of the *Drosophila* CYP4D1 sequence.

Upon closer examination of the published sequence (Snyder and Davidson, 1983), one of us (D.R.N.) found that the sequence similarity extends 5′-ward of the identified sequence. Dr. Nero's computer failed to extend the match because there appear to be three introns that interrupt the sequence. The three putative introns begin with GT, end with AG, and have similarity at their ends to *Drosophila* consensus splice sites defined by Snyder and Davidson (1983). If we count the first nucleotide in Figure 2c of Snyder and Davidson (1983) as nucleotide number 1, the first intron occurs between nucleotides 170 and 235, the second intron runs between 410 and 463, and the third putative intron occurs between nucleotides 721 and 870. The carboxy-terminal fragment of 22 amino acids (nucleotides 871–936) resembles the carboxy-terminal sequences of the CYP4A subfamily and is followed by three in-frame stop codons. Isolation and sequence of the cDNA for this gene would allow confirmation of these predicted introns.

The Snyder and Davidson (1983) gene clearly represents the first published member of the *CYP4* family. However, we have named this gene *CYP4E1* (Table 2), because the first four subfamilies in the *CYP4* family of our database had already been assigned.

UNCERTAINTY ABOUT ENZYME ACTIVITIES

It had been postulated that catalytic activity might aid us in assignment of orthologous genes across species (Nebert *et al.*, 1989b). This hypothesis is clearly no longer true. It has been shown that a single amino acid difference—in a peptide of about 500 residues—can be critical in changing the catalytic activity from "comarin 7-hydroxylase" to "testosterone 15α-hydroxylase" (Lindberg and Negishi, 1989). These data indicate that a particular ancestral gene existing 20 million years ago, for example, could have undergone a single nucleotide substitution in rat but not in mouse 15 million years ago, leading to a single amino acid change, such that the orthologous gene in the rat and mouse would exhibit entirely different catalytic activities. [The rat and mouse species are believed to have separated from one another about 17 million years ago (reviewed in Nebert and Gonzalez, 1987).] Another striking difference between rat and mouse occurs with debrisoquine metabolism in the *CYP2D* subfamily: high levels of debrisoquine-4-hydroxylation occur in the Sprague-Dawley rat but not mouse liver. In studies too numerous to describe here, genetically engineered cDNAs encoding chimeric P450 proteins have also been shown to cause changes in catalytic activities.

GENE REGULATION

It had originally been considered that families or subfamilies of genes might be regulated in a similar fashion, *e.g.*, 3-methylcholanthrene-inducible genes, phenobarbital-inducible cluster, rifampicin-inducible gene cluster, *etc.* This is definitely not the case. P450 "induction" often represents a combination of transcriptional activation, post-transcriptional, and, in some cases, post-translational regulation (Nebert and Gonzalez, 1987; Gonzalez, 1988). Induction of subsets of drug-metabolizing enzymes most likely represents a cell type-specific response to overwhelming concentrations of a foreign compound whose chemical structure resembles that of an endogenous ligand (Nebert, 1991).

Up-regulation (induction) by a particular class of inducers need not be confined to a single family or subfamily. For example, although the mechanism for *CYP2A* gene activation is not understood, it appears that particular CYP2A enzymes—as well as CYP1A1 and CYP1A2—are inducible by polycyclic hydrocarbons (Matsunaga, T., *et al.*, 1988). Also, it appears that certain mammalian CYP2B, CYP2C, and CYP3A enzymes but not others are inducible by phenobarbital. In addition, the house fly CYP6A1 and bacterial CYP102 enzymes are also inducible by phenobarbital (Waxman and Azaroff, 1992). These findings suggest that certain of the regulatory mechanisms involved in enzyme induction evolutionarily preceded the division of P450 genes into families and subfamilies.

ACCESSION NUMBERS FOR THE P450 GENES AND GENE PRODUCTS

We have collected all available P450 accession numbers from SwissProt, NBRF-PIR, and GenBank/EMBL (Table 3). By examining more than 800 pages of printed database output, we have linked these numbers to the correct P450 gene or gene product. To obtain the output, the databases were searched using the GCG package STRINGSEARCH for "P450" and "P-450" in the definitions of the database entries. To find sequences that were P450 but not called P450 in these definitions, we searched the two protein databases with FINDPATTERNS to match "FXXGXXX-CXG." This search uncovered thromboxane A_2 synthase and several other sequences not previously found. The output was extracted by the command FETCH@filename, where "filename" was the filename of files created by STRINGSEARCH and FINDPATTERNS. Some acces-

TABLE 3. ACCESSION NUMBERS FOR CYTOCHROME P450 SEQUENCES

Gene	Species	PIR Database		SwissProt		GenEMBL			
CYP1A1	hum	B23585	A24797	P04798		M12079	M31664->M31667		X04300
		S15803	S16714			X02612	K03191		
		S19336							
	mon	S21761							
	rat	A00185	A24406*	P00185		K02246	M12170	M14633	M26126
		S09617*	S15584			M26129	X00469		
	ham	JX0189				D10251	D12977		
	gpi	F24406*				D11043			
	rab	A27821	B27821	P05176		D00212	M11727	M19917	X05685
	dog	C37222							
	tro	A28789		P10609		M21310			
Cyp1a-1	mou	A00184	S15588	P00184		Y00071	M10021		
		A23923							
CYP1A2	hum	A23585	S07373	P05177		L00383->L00389		M12078	M14337
		S16718	S22433			M31664->M31667		Z00036	M55053
	mon								
	rat	A22562	B24406*	P04799		K02422	K03241	M26127	X01031
		A20963	S16875						
	ham	S13885	JX0190	P24453		M34446	M63787	D10252	
		PX0036*							
	rab	A00187	S02038	P00187		D00213	M11728	M36538	X05686
		B25143	A00188			X13853			
	dog	B37222							
	chi	JT0575				M64537			
Cyp1a-2	mou	A00186	B23923	P00186		X00479	X04283		
		A26373							
CYP2A1	rat	A29560	A34272	P11711		J02669	M33312		
		S03981*							
CYP2A2	rat	A31887	B34272	P15149		J04187	M33313	M33325	M34392
		S03982*							
CYP2A3	rat	S15056	A32030	P20812		M33190	J02852		
		S12708							
Cyp2a-4	mou	A33531	S16067	P15392		M25146	M25147	M26202	M26203
		S16068				M26205->M26208		J03549	M19319
Cyp2a-5	mou	B33531		P20852		M25204->M25211		M26204	
CYP2A6	hum	A00190	S05946	P11509	P00190	K03192	M33316	M33318	X06401
		S04581	S09329	P10890		X13897	X13929	X13930	
		A34271	B34271						
		S04698							
CYP2A7	hum	C34271		P20853		M33317			
CYP2A8	ham	S13884	A33293	P24454		M27906	M34446	M34447	M63788
		PX0037*							
CYP2A9	ham			P24455		M34446	M34448	M63789	
CYP2A10	rab								
CYP2A11	rab								
CYP2A	bov	A35704*		P22779					
CYP2B1	rat	A00176	S03854	P00176		D00250	K01721	L00313->L00320	
		A22363	S19172			M11251	M26854*	M26855*	M37134
		B92255				J00719	M19972	M21412	
CYP2B2	rat	A00177	A21872	P04167		J00720->J00728		K00996	K01626
		S15589	A29298*			K01721	K02427	K02428	M13234
		S03855*	A32736			M15947	M19972	M26853*	M27076
		A21162	B00176			M34452	Y00410	M21403	M13650
CYP2B1 or 2B2		A34259*							
CYP2B3	rat	A25459	A29818	P13107		M20406	M19973		
CYP2B4	rab	A00178	A00179	P00178	P00179	M20856	M20857		
		B27717	C27717						
		E27717							
CYP2B5	rab	D27717	A27717	P12789		M18820	M20855		
CYP2B6	hum	S04578	A32969	P20813		J02864	M29874	X06399	X06400
		S07598	S04579			X13494	X16864		
CYP2B7P	hum					J02864	M29873		
CYP2B8	rat					J04808			
Cyp2b-9	mou	A31047		P12790		M21855			
Cyp2b-10	mou	B31047		P12791		M21856			
CYP2B11	dog	S11305		P24460		M33575	M92447		
CYP2B12	rat	S18907				X63545			
Cyp2b-13	mou					K02409*			
CYP2B14	rat								
CYP2B	gpi	A36154*	S15135*						

CYP2C1	rab	A00181	A34257*	P00180		K01522	M74199	M76597	
CYP2C2	rab	A00182	A27718	P00181		K01521	M14955	M19137	M76596
		S15587*							
CYP2C3	rab	A00183	A22606	P00182		J02901	M31245->M31254		K01523
		A34534							
CYP2C4	rab	A26731	B34257*	P11371		J02716	M74200->M74203		
CYP2C5	rab	A00180	A37828	P00179		J05575	M11299	M74204->M74206	
		S20227				M17026			
CYP2C6	rat	A25954	B25585	P05178		J03509	J04466	K03501	M13711
		A28516	S00955*			M18335	M18336	M24237->M24239	
						M36848	X06712*	M14776	
CYP2C7	rat	A25585	B25954	P05179		M14775	M18774	M31031	N00038
		B28516				M18335			
CYP2C8	hum	A29782	S06306	P10632		J02832	J03472	J05326	M17397
		E28951	C28951			M17398	M21941	X51535	X54807
		S12688	A38462			X54808	M21942	X65962	
		S21423							
CYP2C9	hum	A28530	S06863	P11712		D00173	J02832	J05326	M21940
		C38462	B38462			M61855	M61857		
		A28951*	B28951*						
		PX0013*	A41506						
CYP2C10	hum	D28951	A27541	P11713		J02832	M15331	M21939	
CYP2C11	rat	A29421	A26685	P08683		J02657	M18356->M18363		
CYP2C12	rat	A32140	A34258	P11510		J03786	M33544->M33550		M33656
CYP2C13	rat	A32470	A36122	P20814	P22693	J02861	J05352	M32277	M33994
		S15586				M81311	M82846	M82848->M82855	
CYP2C14	rab	A26921		P17666		D00190	D00191		
CYP2C15	rab	B27718		P11372		M19234			
CYP2C16	rab	S12765	A27479*	P15123		M18376	M29968	X13853*	M29661
CYP2C17	hum	G38462				(numbers removed)			
CYP2C18	hum	D38462	E38462			J05326	M61853	M61856	X63904
CYP2C19	hum	F38462				J05326	M61854		
CYP2C20	mon								
CYP2C21	dog	A37222							
CYP2C22	rat	A39257	S11160	P19225		M58041	X53477		
CYP2C23	rat	S13101		P24470		X55446			
CYP2C24	rat	JH0451				M86677	M86678		
CYP2C25	ham					X60322			
CYP2C26	ham					D11435			
CYP2C27	ham					D11436			
CYP2C28	ham					D11437			
CYP2C	rat								
Cyp2c	mou	A23739							
CYP2D1	rat	A31579	A26822	P10633		J02867	M16654	M22328	
		C32970	B32970						
CYP2D2	rat	C31579	B26822	P10634		M16655	M22330	X52027	X52455
		D32970	S16871						
CYP2D3	rat	E32970	S16872	P12938		J02868	X52028	X52456	
CYP2D4	rat	D31579	S16873	P13108		M22331	X52029	X52457	
CYP2D5	rat	B31579	S09611	P12939		J02869	M22329	M25143	X52030
		A32970	S16874			X52458			
CYP2D6	hum	A28883	S01199	P10635		M19697	M20403	M33388	
		A30335*	A33629			X07618->X07620		X08006	
						X16865->X16867		Y00300	
CYP2D7P	hum					M33387			
CYP2D8P	hum					M33387			
Cyp2d-9	mou	A27384	B30247	P11714		M23998	M23997	M24262	M24267
		S15806	B27384						
Cyp2d-10	mou	A30247	S15807	P24456		M27167	M27168	M24263	
		S19168							
Cyp2d-11	mou	S15808	S19169	P24457		M24264			
Cyp2d-12	mou								
Cyp2d-13	mou								
CYP2D14	bov					X68013	X68481		
CYP2D	rat	A28702*							
CYP2E1	hum	A31949	B25341	P05181		D10014	J02625	J02843	M77918
		A29660*							
	mon								
	rat	A28145	A25341	P05182		J02627	M20131		
		S09072*							
	rab	A26579	C27718	P08682		M15061	M18770	M19235	M21364
		A27750	A27680*			M21365	M21366		
Cyp2e-1	mou	S18037	S19657			X62595			
CYP2E2	rab	B27750	B27680*			M18771	M21349->M21351		
						M21358->M21372			

CYP2F1	hum	A36036	B36036	P24903	J02906			
Cyp2f-2	mou	A39302			J05349	M77497		
CYP2G1	rat	A33875	A35551	P10610	J04715	M31923->M31931		M33296
					M34444			
	rab	S13907		P24461				
CYP2H1	chi	A24814		P05180	M13454			
CYP2H2	chi	A31418	S10683*	P20678	M25469			
CYP2J1	rab	A40938			D90405			
CYP2K1	tro							
CYP2(EST)	Cel				M89049			
CYP3A1	rat	A22631	PX0035*	P04800	M10161	X64401	X62086	
CYP3A2	rat	A25222	PX0032*	P05183	M13646	X62087		
		PX0034*						
CYP3A1 or 3A2		B26997*	A26997*		M86850			
		C26997*						
CYP3A3	hum	A25170*	A29410	P05184	D00003	M13785	N00003	X12387
CYP3A4	hum	A25517	A29815	P08684	M14096	M18907	X12387	D11131
		A32199*	S16900					
		S03851						
CYP3A3 or 3A4		PX0012*						
CYP3A5	hum	S06491	A34101	P20815	J04813			
CYP3A6	rab	A29487	A34236	P11707	J05034	M19139		
CYP3A7	hum	S04983	S02152*	P24462	D00408			
		JX0062	PX0014*					
CYP3A8	mon	SO4509*						
CYP3A9	rat							
CYP3A10	ham	A40843						
Cyp3a-11	mou	S22334			X60452			
CYP3A12	dog	S14275		P24463	X54915			
Cyp3a-13	mou	S18155			X63023			
CYP3A	rat	PX0033*						
CYP4A1	rat	A26137	B32965	P08516	M14972	M33937	M57718	X07259
		S01336						
CYP4A2	rat	A32965		P20816	M33938	M57719		
CYP4A3	rat	A27700*	A32966	P20817	M33936			
CYP4A2 or 4A3		A26380*	S09074*					
CYP4A4	rab	A29368		P10611	J02818	L04758		
CYP4A5	rab	A34260	S14761	P14579	M28655	X57209		
CYP4A6	rab	B34260	A34160	P14580	M28656	M29531	L04755	
CYP4A7	rab	C34260	B34160	P14581	J05150	M29530	M28657	
CYP4A8	rat	SO8300*	A36304	P24464	M37828			
CYP4A9	hum							
Cyp4a-10	mou							
CYP4A11	hum				L04751			
CYP4B1	hum	A33414 *	S07765	P13584	J02871	X16699		
		S17971						
	rat	PQ0092*	B40164	P15129	M29853			
	rab	A40164		P15128	M29852			
CYP4C1	Bld	A39381			M63798			
CYP4D1	Dme							
CYP4E1	Dme				K00045			
CYP4F1	rat				M94548			
CYP4F2	hum							
CYP4F3	hum							
CYP4(EST)	Cel				M89401			
CYP4(EST)	Cel				M80176			
CYP5	hum	JQ1143	S10750	P24557				
CYP6A1	Mdo	A32157		P13527	M25367			
CYP6A2	Dme				M88009			
CYP6B1	Ppx				M80828	M83117		

CYP7	hum	S11051	JH0659	P22680	X56088	M89803	M93133	
	rat	S06632	A38736	P18125	J05460	J05509	M59184->M59189	
		A36450	A35376		J02926	J05430	X17595	Z14108
		A37071						
	rab							
	bov							
CYP10	Lst							
CYP11A1	hum	A25922	S08081*	P05108	D00161->D00169	M14565	M28253	
		S16069*	S16716		M60421	X05367->X05374		
					X14257*	X58981		
	rat	A27321	A34164	P14137	J05156	M22615	M63125->M63133	
		A23688						
	bov	A00189	A24067	P00189	J05245	K02130	M25920	M25921
		A28860*	S04947					
		S15865*						
	pig	A30825	S03188	P10612	X13768			
		PN0018						
	chi	S11144						
CYP11B1	hum	A34181	S11338	P15538	J02985	M24667	X55764	X55765
	rat	B34281*	S05666	P15393	X15431			
	bov	A28415	B28860*	P15150	D00185	D00186	D00361	
		JX0050	JX0071		D00449->D00457		J02985	M17843
		S15805	A38819		M17844	M36535		
Cyp11b-1	mou	B41552		P15539	J04451			
CYP11B2	hum	JX0151	B34181	P19099	X54741			
		A37088						
	rat	A35342	B35342		D00567	D00568	X52766	
		S09736						
	bov							
Cyp11b-2	mou	A41552						
CYP11B3	rat							
CYP11B4	bov	cf.CYP11B1, different flanking regions						
CYP11B5P	bov	JU0316						
CYP11B6P	bov	JU0317						
CYP11B7P	bov	JU0318			D00458	D00459		
CYP11B8P	rat							
CYP17	hum	A29587	A26366	P05093	M14564	M19489	M63871	
		A40908	S16717		M31146->M31153			
		A40921						
	rat	A27659	A30828	P11715	M21208	M22204	M27282	M31681
		A33980	S16719		X14086	Z11902	S40343	
	gpi	D28860*						
	bov	S04346	A26289	P05185	M12547			
	tro	S21125						
	pig	B26366	S22339	P19100	M63507	Z11854->Z11856		
					S87722	S40340->S40342		
	chi	JT0318 *		P12394	M21406			
Cyp17	mou	A39072		P27786	M64863	S41708		
CYP19	hum	S03962	A31255	P11511	J04127	JO5105	M28420	
		A31580	A29480		M30795->M30804		M74714	X55983
		A23546*	A34451		Y07508	M18856	X13589	
		A24344*	D24344*					
		B24344*	E24344*					
		C24344*	S22908					
		A40142						
	rat	A36121	S16901	P22443	M33986			
	chi	A31916		P19098	J04047			
	tro							
	gol							
Cyp19	mou	S13912			D00659			
CYP21A1	bov	A00192	A24101	P00191	K01333	M11267	M12918	M13461
		A27555	A21181		M13545			
		C28860						
	pig	A32525		P15540	M83939			
	she				M92836	M92837		
Cyp21a-1	mou	A00193	A26660	P03940	K03234	M64933	M73820	M15009
CYP21A1P	hum	A21889*	A25446	P08686	M12793	M12792		
		A27865	A29406		M25813	M13936		
		A32715	A33725					
CYP21A2	hum	A00191	A33725	P04033	K02771	M12792	M25813	M17252
Cyp21a-2p	mou				K03234	D15008	M19711	

CYP24	rat	S13918							
CYP27	hum	A39740				M62401			
	rat	A36239	A34558	P17178		M38566	Y07534		
		A33406*	S09198						
	rab	A30293*	A33813	P17177		J04717			
		A32279*							
CYP51	Sce	A25563*	A27491	P10614		M15663	M21483	M21484	M18109
		B31569							
	Ctr	A26828*	A31854	P14263		M17595	M23673		
	Cal	S02713		P10613		X13296			
CYP52A1	Ctr	A29297	JS0203	P10615		M15945	M24894	M63258	
CYP52A2	Ctr	S06148	JT0980			M63258	X17560		
CYP52A3	Cma	JQ1040	S08667	P16496	P20017	D00481	M27081	X51931	
		JU0095	A33254	P24458		D01168			
		JQ1039							
CYP52A4	Cma	S08668	B40576	P16141		X51932	X55881	D12715	
CYP52A5	Cma	A40576				X55881	D12714		
CYP52A6	Ctr	S22972				Z13010			
CYP52A7	Ctr	S22973				Z13011			
CYP52A8	Ctr	S22974				Z13012			
CYP52A9	Cma	JS0723				D12717			
CYP52A10	Cma	JS0725				D12719			
CYP52A11	Cma	JS0726				D12719			
CYP52B1	Ctr	S22975				Z13013			
CYP52C1	Ctr	S22976				Z13014			
CYP52C2	Cma	JS0724				D12718			
CYP52D1	Cma	JS0722				D12716			
CYP53	Ani	S10453	S12015	P17549		X52521			
CYP54	Ncr	S09643*				X15033			
CYP55	Fox			P23295		M63340			
CYP56	Sce	B36395	S14228	P21595		X55713			
CYP57	Nha								
CYP71	Pam	A35867		P24465		M32885			
CYP72	Cro								
CYP73	Htu								
CYP101	Psp	A00194	A25660	P00183		M12546			
CYP102	Bme	A34286		P14779		J04832			
CYP103	Atu	A32306		P24466		M19352			
CYP104	Atu	B32306		P24467		M19352			
CYP105A1	Sgl	A35401		P18326		M32238			
CYP105B1	Sgl	B35401		P18327		M32239			
CYP105C1	Str	S15809		P23296		J03356	M31939		
CYP105D1	Stg	S18924		P26911		X63601			
CYP106	Bme	S07764	S17973	P14762		X16610			
CYP107A1	Ser	S16745	S18531			M54983	X60379		
CYP107B1	Ser					M83110			
CYP108	Pse					M91440			
CYP109	Bsu			P27632		M36988			
CYP110	Ana	C37842				M38044			
CYP111	Pse								

CYP112	Bja		L02323
Unidentified sequences			
	hum	S14367*	
	hum	S08282* (2C like or 2A like)	
	rat	S09073*	
	rab		M12280->M12288 (pseudogene)
	pig	S11471*	
	pig	S15850*	
	chi	S10680*	
	chi	S10681* S13263*	
	chi	S10682*	

species abbreviations: rat = rat; hum = human; rab = rabbit; mon = monkey; dog = dog; mou = mouse; ham = hamster; gpi = guinea pig; chi = chicken; tro = trout; bov = bovine; Bld = Blaberus discoidalis (cockroach); Cel = Caenorhabditis elegans; Dme = Drosophila melanogaster; Mdo = Musca domestica (house fly); Ppx = Papilio polyxenes (black swallowtail butterfly); Lst = Lymnea stagnalis (pond snail); pig = pig; she = sheep; gol = goldfish; Sce = Saccharomyces cerevisiae; Ctr = Candida tropicalis; Cal = Candida albicans; Cma = Candida maltosa; Ani = Aspergillus nigra; Ncr = Neurospora crassa; Fox = Falciparum oxysporum; Nha = Nectria haematococca; Pam = Persea americana (avocado); Cro = Catharanthus roseus (syn. Vinca rosea, Madagascar periwinkle); Htu = Helianthus tuberosus, Jerusalem artichoke; Psp = Pseudomonas putida; Bme = Bacillus megaterium; Atu = Agrobacterium tumefaciens; Sgl = Streptomyces griseolus; Stg = Streptomyces griseus; Str = Streptomyces species; Ser = Saccharopolyspora erythraea; Pse = Pseudomonas species; Bsu = Bacillus subtilis; Ana = Anabaena species; Bja = Bradyrhizobium japonicum.

GenEMBL (GenBank/EMBL/DDBJ) is a combined listing of GenBank, EMBL, and the DNA Data Bank of Japan accession numbers, because they use common numbering.

Some sequences can be identified from amino-terminal sequences as belonging to a subfamily. These have been included in the listing of the subfamily, even though it is premature to assign a specific number to these sequences.

* Indicates peptides of less than 100 amino acids in length (only complete for PIR and SwissProt).

sion numbers were also found as cross references to other databases, and some numbers were obtained from published papers. If you know of additional accession numbers not listed in Table 3, please let us know.

Some accession numbers appear more than once in Table 3. GenBank and EMBL give two or more genes the same accession number if they exist on one contiguous segment of DNA, as seen with *CYP2D7P* and *CYP2D8P*. On the other hand, exons that are separated by unsequenced introns are given different, usually consecutive, accession numbers. The SwissProt database gives one accession number per sequence (or sometimes two), and all references to that sequence appear under that one number. This is why there are fewer SwissProt accession numbers than NBRF-PIR accession numbers. Although there are almost 1,100 numbers in Table 3, there are some sequences that have not yet been submitted to any of the databases, or of which we are unaware of their accession numbers at the time of submission of this manuscript.

ALLELIC VARIANTS

We are all aware of mistakes in sequence reading and translation, typographical errors, inclusion of vector DNA, and other cloning artifacts (Lamperti *et al.*, 1992; Lenstra, 1992), all of which can contribute to a lack of homogeneity of a particular sequence in the DNA and protein databases. In addition, it must be emphasized that outbred species—not only rats and other laboratory animals, but also agricultural livestock and especially humans—can contain multiple alleles of the same gene. A large number of allelic variants ("microheterogeneity") have already been identified in the human, cow, and rodent. The Nomenclature Committee has arbitrarily assigned proteins having $\leq 3\%$ divergence as derived from two alleles of the same gene, unless (i) functional differences (catalytic activities) have been demonstrated or (ii) nontranslated regions are clearly divergent, indicating distinct genes.

The 3% cut-off connotes fewer than 15 amino acid differences among a 500-residue P450 protein. There are several cases, however, in which this arbitrary cut-off does not hold. Even though the rat CYP2B1 and CYP2B2 proteins are 97.4% identical, and the mouse CYP2A4 and CYP2A5 proteins differ at only 11 residues, they are known in both cases to represent products from two distinct genes and thus are not allelic variants from the same locus. In other cases as well, it has been difficult to distinguish allelic variants from products derived from two distinct genes. More rigorous gene mapping and cloning studies should help clarify some of this confusion during the next several years. Studies of inheritance patterns will provide the clearest evidence for establishing whether two alleles or two genes exist.

SCOPE OF THE NOMENCLATURE SYSTEM

The difficulties noted between minor variants of P450 DNA or protein have caused some criticism that the nomenclature is not meaningful at the microlevel. For example, the human *CYP2C9* and *CYP2C10* sequences are nearly identical in their coding sequences and vary only in their noncoding sequences. There is concern that the variation is artificial and due to a spurious ligation during preparation of the cDNA library, a problem not uncommon in certain cDNA library construction protocols. Thus, *CYP2C10* may be the result of a ligation of the coding region and a small portion of the 3′ noncoding region of the *CYP2C9* cDNA with a fragment of an unrelated cDNA. Others believe there is evidence that both sequences are genuine and that there are two genes. Human *CYP2C17* appears to be a splice variant of *CYP2C18* and *CYP2C19*. It is impossible for the Nomenclature Committee to make a ruling on these and other similar problems; we can only note these concerns with question marks in Table 2. In this sense the nomenclature is meaningless at this level, but the workers involved do have the framework of the nomenclature for defining the problem.

Because the human *CYP2C10* and *CYP2C17* genes are reported in the literature and have accession numbers assigned to them (Table 3), however, we shall continue to list them (Table 2). Furthermore, it is obvious that we cannot use these gene designations for any other gene.

The suggestion has also been made that chimeric forms of P450 be given a standarized nomenclature. At this time, we believe that hybrids are unique to each laboratory, that they should be given sensible names, and that this not be a duty of the Nomenclature Committee. As long as the parent sequences are correctly named and the fusion joints are clearly given, this should be a matter for each individual laboratory.

TRIVIAL P450 NAMES THAT HAVE BEEN ASSOCIATED WITH A SPECIFIC CATALYTIC ACTIVITY

Part of the confusion in the P450 literature has been the extensive use of trivial names, which have often been assigned to the same gene or gene product. In a number of cases, it is now possible to list many of these names, in which a particular gene or gene product can clearly be identified with a purified protein, reconstituted catalytic activity, or an unambiguously defined genetic difference (Table 4). The Table 4 data should aid those outside the field, and especially colleagues who have searched the older literature and become confused about correlations between the enzyme studied then and the current gene/gene product nomenclature system.

EXPRESSED SEQUENCE TAGS (ESTs)

The sequencing of cDNA libraries, encouraged by the Human Genome Project, has recently identified P450 genes in the nematode *Caenorhabditis elegans*. Three expressed sequence tags were reported in a recent summary of this project (McCombie *et al.*, 1992; Waterston *et al.*, 1992). Although these represent partial P450 cDNA sequences, there are sufficient data to assign one of these tentatively to the *CYP2* family and the other two quite definitely to the *CYP4* family (Tables 2 and 3). Whether these sequences will represent members of a new family or subfamily must await further sequencing of full-length cDNA clones.

The putative *CYP2* EST is coded on a fragment of 479 bp (GenEMBL: M89049). The sequence starting at nucleotide 1 encodes a fragment homologous to amino acids 71–208 of CYP2J1. There are several frameshifts in this expressed sequence tag, making it difficult in some regions to identify which frame is correct. Nucleotides 1–288, 353–391, 393–401, and 403–429 were used as the open reading frame; nucleotides 289–352 were excluded because they could not be assigned to any reading frame. Some of these choices may be incorrect, and a more careful sequencing of this clone might resolve the problem of shifting reading frames. This fragment is in the amino-terminal region and appears most similar to *CYP2* family sequences. Once the more highly conserved regions of this gene are sequenced, it is possible that the gene might be assigned to a particular *CYP2* subfamily or to a new gene family.

One of the *CYP4* EST sequences is 528 bp (GenEMBL: M89401). It was not clear to us which reading frame to use over the first 62 nucleotides, so these bases were not included in the translation and alignment. The translated segment from 63 to 528, however, was homologous to amino acids 277 to 434 of the cockroach CYP4C1 sequence; this segment includes the P450 "I helix" (Poulos *et al.*, 1987), but stops short of the conserved cysteine.

The other *CYP4* EST sequence is 376 bp (GenEMBL: M80176). Nucleotides 206–171 on the complementary strand code for a sequence AGPRNCIGQKFA, which is identical to that of CYP4C1, 4D1, and 4E1 surrounding the conserved cysteine. The strong sequence similarity predicts that a PFS sequence should occur upstream of the AGPRN... sequence, but this was absent in all three possible reading frames between 376 and 207. Therefore, since the first 170 nucleotides on the complementary strand do not appear to match the *CYP4* family sequences in any reading frame, these nucleotides are presumed to be non-P450 in origin. The sequence 170-27 on the complementary strand continues the sequence to a stop codon at 26–24, which appears to be the true end of this putative *CYP4* EST coding sequence.

NONLINEARITY OF THE RATE OF EVOLUTION

As has been noted for genes in other superfamilies, the rate of divergence of the P450 genes has been found to be markedly nonlinear (Nebert and Gonzalez, 1987; Nebert *et al.*, 1989b; Nebert and Nelson, 1991). This nonlinearity might be explained by paleontologic evidence, suggesting that evolutionary rates in different lineages have indeed not been linear. The human and chicken *CYP17* genes, for