DNA AND CELL BIOLOGY
Volume 14, Number 1, 1995
Mary Ann Liebert, Inc., Publishers
Pp. 73–82

A Cluster of Cytochrome P450 Genes of the CYP6 Family in the House Fly

MICHAEL B. COHEN[1] and RENÉ FEYEREISEN

ABSTRACT

A cluster of genes of the *CYP6* family was found in a series of overlapping λDASH clones from a genomic library of the house fly, *Musca domestica*. Four complete genes, *CYP6A3, CYP6A4, CYP6A5*, and *CYP6C1*, and fragments of two other genes, *CYP6A6* and *CYP6C2*, were closely linked on a 24-kb segment of DNA. Restriction fragment length polymorphism (RFLP) analysis of PCR-amplified segments of two of the genes showed that the cluster is localized on chromosome V of the house fly. Each gene contained a short intron of 57 to 125 bp interrupting a conserved Glu codon, as in the previously described *CYP6A1* gene. The gene fragment *CYP6A6* consisted only of the coding region downstream from this intron, *i.e.*, about one-third of the complete P450. The gene fragment *CYP6C2* was missing a short amino-terminal part of the coding region, and may represent the two last exons of a larger gene. Gene duplication and chromosomal inversion events may explain the origin of this cluster. The P450 proteins deduced from the nucleotide sequences shared 39–71% amino acid identity with each other. This low identity and the lack of evidence of recent gene conversion events suggested that this cluster may be evolutionarily ancient and that homologous clusters may be found in other holometabolous insects. Evidence for transcription of the genes and for correct splicing of the introns was obtained by northern blotting and reverse transcription polymerase chain reaction (RT-PCR) experiments. No overexpression was observed in any of three insecticide-resistant house fly strains. RT-PCR and sequencing also revealed the existence of other genes or alleles closely related to the members of this cluster.

INTRODUCTION

THE DIVERSITY OF GENES OF THE CYTOCHROME P450 superfamily across taxa and within a single species (Nelson *et al.*, 1993) reflects the ancient origin of P450 genes but also bursts of gene duplications within a phylogenetic lineage: The presence of P450s that are undoubtedly homologous in bacteria, fungi, plants, and animals suggests that an ancestral P450 gene evolved more than 2 billion years ago (Nebert *et al.*, 1989), and it has been estimated that the initial P450 gene duplication took place perhaps as early as 1.4 billion years ago (Nelson and Strobel, 1987). Fifty P450 genes have been reported in the rat (Nelson *et al.*, 1993) and the total number of documented P450 sequences has grown from five in 1985 to more than 300 to date. The diversity of P450s of the CYP2 family in mammalian species has been hypothesized to result from animal–plant "warfare" (Gonzalez and Nebert, 1990), and it appears now that the diversity of plant P450s responsible for the synthesis of so-called secondary plant chemicals is indeed matched by a diversity of herbivore P450s responsible for the detoxification of such chemicals. An example is provided by CYP6B1, a P450 isolated from the caterpillar of the black swallowtail *Papilio polyxenes* (Cohen *et al.*, 1992). CYP6B1 is induced by toxic linear furanocoumarins found in the diet of this specialist herbivore and was recently shown to metabolize these compounds (Ma *et al.*, 1994). It is plausible that the inexhaustible biosynthetic repertoire of plants, fungi, and bacteria has constituted a strong selective force driving the diversification of animal P450s.

Insects of importance to public health and agriculture have seen another strong selective force in the last 50 years, that of synthetic insecticides. Increased detoxification by P450 enzymes plays a major role in many cases of insecticide

Department of Entomology and Center for Insect Science, University of Arizona, Tucson, AZ 85721.
[1]Present address: International Rice Research Institute, P.O. Box 933, 1099 Manila, Philippines.

resistance but until now, relatively little has been known of the molecular basis of P450-mediated insecticide resistance. The *CYP6A1* gene of the house fly *Musca domestica* (Feyereisen *et al.*, 1989) and the *CYP6A2* gene of *Drosophila melanogaster* (Waters *et al.*, 1992) were both shown to be overexpressed in insecticide-resistant strains. Overexpression of *CYP6A1* was not the result of gene amplification, but of a mutation in the resistant strain that altered regulation of the gene by a *trans*-acting factor (Cariño *et al.*, 1994). CYP6A1, which metabolizes cyclodiene insecticides (Andersen *et al.*, 1994), is, however, not the only P450 involved in insecticide resistance in the house fly, and a multiplicity of P450s in this species is inferred from biochemical (*e.g.*, Ronis *et al.*, 1988) and molecular (Cariño *et al.*, 1992) data. In a sense, the selection of P450 diversity by natural products over hundreds of millions of years may have preadapted insects to selection by insecticides over the last few decades. Therefore, it was expected that multiple P450 genes would be found in insects. Here we describe the characterization of a cluster of P450 genes in the house fly. This cluster contains at least four functional P450 genes of the CYP6 family. Gene organization, chromosomal localization, and expression of these genes is documented. Sequence divergence of the P450s in this cluster is much greater than the divergence of P450s in the mammalian CYP2D cluster.

MATERIALS AND METHODS

Library screening

A genomic library of the *sbo* strain of the house fly was constructed in the λ DASH vector (Stratagene) as described for the Rutgers strain by Koener *et al.* (1993). Approximately 350,000 plaques of the library were probed with a 500-bp coding-region fragment of *CYP6A1* from the Rutgers strain of the house fly. Because our initial intent was to isolate *CYP6A1* from the *sbo* strain, we screened at high stringency (50% formamide, 5× SSPE, 3× Denhardt's solution, 0.5% $NaDodSO_4$, and 10 μg/ml salmon sperm DNA, at 42°C) and conducted the final wash at high stringency as well (0.1% SSPE, 0.1% $NaDodSO_4$ at 65°C). After secondary screens of positive plaques, single plaques were picked for phage purification and characterization.

An additional 350,000 plaques of the library were later screened under the same conditions with four additional probes: a 1.4-kb cDNA clone of *CYP6A1* (Feyereisen *et al.*, 1989) and three probes derived from the two *sbo* genomic clones (*sbo*λ A and *sbo*λ B) isolated from the first library screen. These were a 1.3-kb *Eco* RI restriction fragment of *CYP6A3*, a 4.0-kb *Eco* RI fragment containing portions of *CYP6C1* and *CYP6C2*, and a 2.8-kb *Eco* RI fragment containing a portion of *CYP6A6* (see Fig. 1).

Sequencing of genomic clones

As a first step, we subcloned into the the pBS II SK vector (Stratagene) all *Eco* RI restriction fragments of the four genomic clones *sbo*λ A, B, D, and E. Genes were located by comparing DNA sequences from the subclones, translated into all six reading frames by the Blastx procedure (Gish and States, 1993), to protein sequences in the GenBank and EMBL databases. The correct order of all subcloned fragments was confirmed by cloning and sequencing restriction fragments or PCR products bridging adjacent subclones. PCR products were generated with vector and/or internal primers and were cloned into the pCR II vector with the TA cloning kit (Invitrogen). The genomic subclones were sequenced using a combination of vector and internal primers and exonuclease III deletions (Erase-a-base kit, Promega). All genes in the cluster were sequenced on both strands, from at least 370 bp upstream to 100 bp downstream of the coding region. Over 11 kb of sequence covering the cluster were deposited in GenBank, as listed in the legend to Fig. 1.

Northern blotting

Poly$(A)^+$RNA was isolated from house flies (different strains, stages, or chemical treatments) exactly as described in Cariño *et al.* (1992). Four-microgram samples were separated on 1% agarose-formaldehyde gels (Davis *et al.*, 1986) and transferred to Zetaprobe membranes (Biorad). The

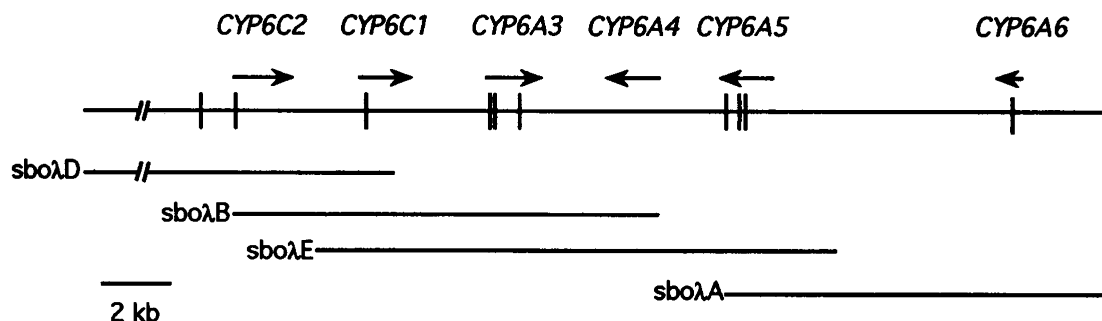

FIG. 1. Map of the *CYP6* gene cluster. Arrows indicate the direction of transcription; vertical marks indicate *Eco* RI restriction sites. The extent of the four clones (*sbo*λ A, B, D, and E) isolated from the *sbo* genomic library is shown by the lines below. The GenBank accession numbers of the genes and the length of sequence deposited are as follows: *CYP6C2*, U09345 (1,801 bp); *CYP6C1*, U09233 (2,301 bp); *CYP6A3*, U09231 (2,142 bp); *CYP6A4*, U09232 (2,089 bp); *CYP6A5*, U09343 (2,110 bp); *CYP6A6*, U09344 (718 bp).

blots were serially probed at high stringency (50% formamide, 0.12 *M* $NaHPO_4$, 0.25 *M* NaCl, 5 m*M* EDTA, 7% $NaDodSO_4$) with 300–1,500-bp coding-region fragments of each of the six genes in the cluster and with a fragment of the actin 2 gene of *Drosophila melanogaster*. Final membrane washes were also at high stringency (0.2% SSC, 0.1% $NaDodSO_4$ 65°C).

Reverse-transcription PCR

Two methods were used to amplify fragments of the genes in the cluster by RT-PCR. For *CYP6A4* and *CYP6A5*, poly(A)$^+$RNA from *sbo* house fly larvae and adults, respectively, was reverse-transcribed with the Superscript enzyme (BRL) and primers designed to regions downstream from the presumed intron in each of the genes. The RT products were extracted with phenol-chloroform, precipitated with ethanol, and resuspended in water. For PCR amplification, we used the RT primer and a second primer corresponding to a region upstream of the presumed intron.

For *CYP6A3*, *CYP6C1*, and *CYP6C2*, we used poly (A)$^+$RNA from *sbo* adult house flies and the *rTth* enzyme kit (Perkin-Elmer), following the manufacturer's instructions. This enzyme allowed us to reverse-transcribe at higher temperatures for greater specificity. As above, we used specific primer pairs in which one was downstream and one upstream of the presumed intron in each gene.

The sequences (5'→3') of the upstream and downstream primers for the various genes were: *CYP6A3*, TGGCTATCAGGAACAGG and AATTTCGAGGGAGAGTC; *CYP6A4*, CTATTTTAAACCGGTGGG and CTCAAGGATCCCAATTC; *CYP6A5*, CCTACAATCTCACAATGGTAT and ATGGCGTCGCTTGTGCAAACGATT; *CYP6C1*, CCTCCTCGAATACTGAAGAC and TCAGCTGGAAATGGATG; *CYP6C2*, CTCCTTGGTGTTGAAGACTC and CCCGTTTCACCACCGATG.

RT-PCR products of the expected size were gel purified with the Sephaglas Band Prep kit (Pharmacia) and cloned with the TA cloning kit (Invitrogen). We sequenced from 450 to 1,000 bp of each product to confirm that it corresponded to the expected gene and that the intron had been excised as predicted.

Chromosomal localization

To determine the chromosomal localization of the gene cluster, we compared restriction patterns of PCR-amplified fragments of two of the clustered genes in two strains of the house fly, Rutgers and *aabys*, and in backcross progeny of F_1 hybrids to *aaybs*. The *aabys* strain bears recessive mutations on each of the house fly's five autosomes (*ali-curve*, *aristapedia*, *brown body*, *yellow eye*, and *snip wing* for chromosomes I–V, respectively), whereas the Rutgers strain is wild-type for these characters. Five phenotypes of the backcross progeny were analyzed, each homozygous for one *aabys* chromosome, *e.g.*, ++++*s*.

Genomic DNA was isolated from one male and one female fly of each parental strain and the F_1 and backcross progeny. A 1.6-kb fragment of *CYP6A5* was amplified with primers CCTACAACAATCTCACAATGGTAT and ATGGCGTCGCTTGTGCAAACGATT. The PCR product was digested directly with *Taq* I; 1 μl of REact 1 buffer (BRL) was added per 9 μl of final reaction volume. The digestion products were separated on a 2% agarose gel and stained with ethidium bromide.

A fragment of CYP6A3 was amplified with primers AATTTCGAGGGAGAGTC and ACCGAGCTCCCACGCCTTTAAAGGAA. It was necessary to gel-purify the expected 1.2-kb product from nonspecific products, using the Sephaglas Band Prep kit (Pharmacia). The purified product was digested with *Mbo* I and restriction products were separated on a 2% agarose gel and stained with ethidium bromide.

RESULTS

Isolation of genomic clones

Our first screening of the genomic library from the *sbo* strain of the house fly with a fragment of *CYP6A1* from the Rutgers strain yielded two clones: *sbo*λ A, 12.3 kb in length and *sbo*λ B, 13.4 kb (Fig. 1). These clones contained two complete P450 genes, *CYP6A4* and *CYP6C1*, and fragments of four other genes. To isolate the remaining portions of these incomplete genes, we plated out another aliquot of the library and probed with restriction fragments from the ends of *sbo*λ A and B. We also screened with a second *CYP6A1* probe, the cDNA clone HFP61 (Feyereisen *et al.*, 1989). The second screen yielded two clones containing additional portions of the gene cluster: *sbo*λ D of 17 kb and *sbo*λ E of 16.6 kb (Fig. 1), but no clones containing *CYP6A1*.

The overlapping clones *sbo*λ A, B, D, and E covered 39 kb, with six P450 genes arranged along a region of 24 kb (Fig. 1). The transcriptional direction of the adjacent genes *CYP6C2*, *CYP6C1*, and *CYP6A3* was opposite of that of the three remaining genes. To determine if additional P450 genes were present in the cluster, we obtained DNA sequence at a minimum of 1 kb intervals from all the gaps between the six genes and from the 11 kb of clone *sbo*λ D upstream of *CYP6C2*. The 2 kb upstream of the *CYP6A6* fragment on *sbo*λ A was completely sequenced. The 5'-terminal portion of *CYP6A6* was not found nor were any additional P450s found.

Gene structure

By amino acid sequence alignment with known P450 proteins of the CYP6 family (Fig. 2), it was evident that four of the genes in the cluster, *CYP6C1*, *CYP6A3*, *CYP6A4*, and *CYP6A5*, were complete, while *CYP6C2* and *CYP6A6* were not. Only about one-third of *CYP6A6* was present, corresponding to approximate amino acid positions 380–500. The *CYP6C2* genomic sequence appeared to be interrupted by an intron towards the 5' end (Fig. 2) because the open reading frame was interrupted by a stop codon 45 bp upstream of the first amino acid shown (Glu-52 of the alignment, Fig. 2) and because the conserved proline-rich region situated after the hydrophobic transmembrane region, seen in the other *CYP6* genes, was missing from the deduced sequence. We selected

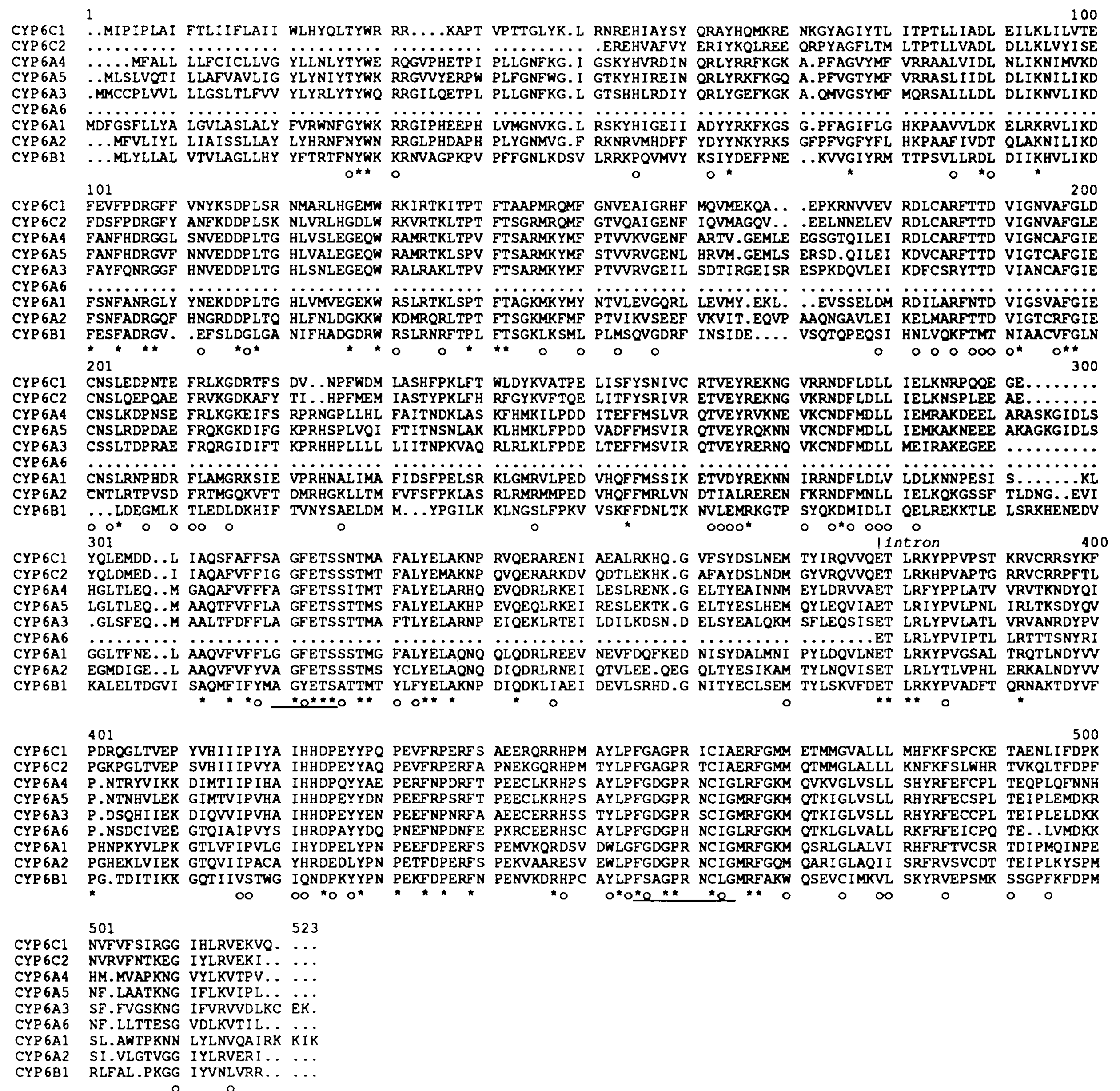

FIG. 2. Alignment of the deduced amino acid sequences of the six genes in the cluster and three other members of the *CYP6* family. The position of the intron in *CYP6A1* (Cohen *et al.*, 1994), *CYP6A2* (Dunkov and Feyereisen, unpublished results), and the clustered genes is marked on top of the alignment (I*intron*). The other members of the *CYP6* family are house fly *CYP6A1* (Feyereisen *et al.*, 1989); *CYP6A2* from *Drosophila melanogaster* (Waters *et al.*, 1992); and *CYP6B1* from *Papilio polyxenes* (Cohen *et al.*, 1992). The regions surrounding the conserved Thr in distal helix I and the conserved Cys in the heme binding domain are underlined. Residues conserved in all CYP6 proteins (*) or in all members of the cluster (o) are marked.

this Glu as the first amino acid because the amino acid homology with the other genes extended to this site, which contained the best of three neighboring consensus 3′ intron splice sites. We were unsuccessful in characterizing the 5′ end of the *CYP6C2* transcript by use of the RACE procedure.

All of the genes in the cluster had a short intron in the same position as that of *CYP6A1* (Cohen *et al.*, 1994), with the exception of *CYP6A6*, which was disrupted at this position (Table 1). The splice sites of each intron were confirmed by sequencing RT-PCR products generated by primers designed to regions upstream and downstream of its predicted location.

Potential promoter sequences upstream of the genes were sought. In *CYP6A3*, a TATAAAA sequence at 437 bp upstream of the start codon was followed by an arthropod initiator sequence (TCAGT; Cherbas and Cherbas, 1993) at 33 bp after the TATA box, and by sequences (AT-rich, consensus splice sites) compatible with the presence of an

TABLE 1. SHARED INTRON POSITION AND STRUCTURE OF INSECT *CYP6* GENES

Gene	*Intron length*	*Amino acid position*	*5′ Splice site*	*3′ Splice site*
CYP6A1[a]	60 bp	364	GTATTGAAT G /**gtaagt**	tgc**ag**/ AAACACTCCGC
			ValLeuGlnG	luThrLeuArg
CYP6A2[b]	69 bp	369	GTCATCTCA G/**gtaggt**	ttc**ag**/ AAACCCTGAGG
			ValIleSerG	luThrLeuArg
CYP6A3	62 bp	362	AGCATATCG G/**gtgaga**	cac**ag**/AAACCCTTCGC
			SerIleSerG	luThrLeuArg
CYP6A4	66 bp	365	GTGGTGGCG G /**gtaagt**	tac**ag**/ AGACCCTACGT
			ValValAlaG	luThrLeuArg
CYP6A5	57 bp	368	GTGATTGCT G /**gtaagt**	tgc**ag**/ AAACCCTACGC
			ValIleAlaG	luThrLeuArg
CYP6A6	(Gene disrupted upstream of presumed intron?)			ttc**ag**/ AAACCCTACGC
				ThrLeuArg
CYP6C1	64 bp	356	GTGGTGCAA G/**gtaagg**	ctt**ag**/ AGACCCTGCGC
			ValValGlnG	luThrLeuArg
CYP6C2	125 bp	n.a.	GTGGTGCAG G/**gtaaga**	tcc**ag**/ AAACCCTTCGC
			ValValGlnG	luThrLeuArg

[a]Data from Cohen *et al.* (1994).
[b]*(Drosophila melanogaster)* Data from Dunkov and Feyereisen (unpublished results).

intron in the 5′ untranslated region. In *CYP6A4*, an imperfect TATA box preceded an arthropod initiator sequence, TCAGT, which was at 99 bp upstream of the start codon. In *CYP6A5*, a TATAAAA sequence was located 185 bp upstream of the start codon, and a TATAA sequence was found 110 bp upstream of the start codon of *CYP6C1*. Primer extension analysis would be needed to map the 5′ transcription start site of the genes correctly. No barbie box (Shaw and Fulco, 1993) was found upstream of the start codon of the genes.

Sequence identity among genes in the cluster

All the genes in the cluster coded for proteins with the signature sequence of cytochrome P450 proteins, a conserved heme binding domain near the carboxyl terminus (Nelson *et al.*, 1993). However, CYP6C1 and CYP6C2 had an unusual substitution of alanine for glycine at position +2 from the heme liganding cysteine and their heme binding motif is thus F--G--C-A, instead of F--G---C-G. All the deduced protein sequences also had a stretch of hydrophobic residues preceding a conserved GFETS around the threonine residue of helix I involved in catalysis (homologous to Thr-252 of P450cam).

The six genes in the cluster encoded proteins that shared 39–71% amino acid identity with each other and 37–47% identity with CYP6A1, the only other *CYP6* gene sequenced from the house fly (Table 2, Fig. 3). The clustered house fly P450s were less closely related (31–34% identical) to CYP6B1 from the black swallowtail butterfly (Cohen *et al.*, 1992). CYP6C1 and CYP6C2, which shared the substitution of alanine for glycine in the heme binding domain, were 66% identical to each other but only 37–46% identical to the CYP6A proteins in the cluster. P450s are usually placed in the same subfamily if they are more that 55% identical (Nelson *et al.*, 1993).

CYP6A4 and CYP6A5, coded from adjacent genes in the same orientation, were more closely related to each other

TABLE 2. IDENTITY MATRIX OF GENES IN THE *CYP6* FAMILY

	CYP6C1	*CYP6C2*	*CYP6A4*	*CYP6A5*	*CYP6A3*	*CYP6A6*	*CYP6A1*	*CYP6A2*	*CYP6B1*
CYP6C1	*	66.1	41.5	41.3	39.3	—	36.9	36.7	31.2
CYP6C2	68.6	*	45.9	46.8	42.8	—	41.2	39.5	32.6
CYP6A4	44.9	43.5	*	70.9	60.4	—	45.5	44.5	32.1
CYP6A5	45.6	47.1	68.8	*	65.0	—	47.5	45.8	34.1
CYP6A3	45.4	45.7	61.6	71.7	*	—	42.7	43.5	32.7
CYP6A6	42.6	44.1	52.2	60.3	56.6	*	—	—	—
CYP6A1	36.9	40.0	47.1	50.7	46.5	43.4	*	49.1	32.5
CYP6A2	42.4	41.7	42.7	46.4	48.2	41.9	56.8	*	35.1
CYP6B1	37.7	35.5	42.7	42.7	39.9	33.8	42.0	40.6	*

Distances were calculated with the DISTANCES program of the GCG package, with a threshold of comparison score of 1.5 (*i.e.*, only perfect matches scored) and a denominator of length of shorter sequence without gaps. Values on top of the diagonal: identities over the full length of the protein. Below the diagonal, identities over the length of the second exon only, calculated by the same method.

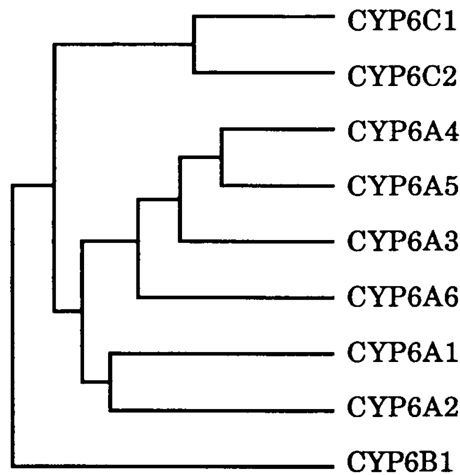

FIG. 3. Dendrogram showing the relatedness of the members of the *CYP6* family, obtained by UPGMA with the PILEUP program of GCG. The alignment used is shown in Fig. 2 and the distance matrix is in Table 2.

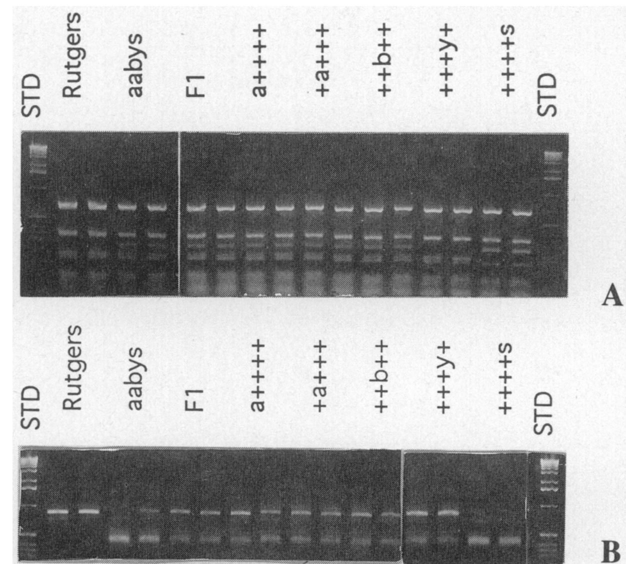

FIG. 4. Chromosomal localization of the *CYP6* gene cluster. Restriction digest patterns for fragments of *CYP6A3* (Fig. 4A) and *CYP6A5* (Fig. 4B) are shown for the two parental strains (Rutgers and *aabys*), F_1 hybrids, and backcross progeny to *aabys*. For each strain, the PCR product from a single male (left) or female (right) was digested with the appropriate restriction enzyme (*Taq* I, A; *Mbo* I, B). In A, a ~250-bp fragment in the Rutgers strain is cut into ~200- and 50-bp fragments in *aabys* as a result of an additional *Taq* I site. In B, a 900-bp fragment in the Rutgers strain is cleaved into two ~450-bp fragments in *aabys* due to an additional *Mbo* I site. Molecular weight markers are the 1-kb ladder (BRL).

(71% identical) than either was to the product of the *CYP6A3* gene, adjacent to *CYP6A4* but lying in the opposite orientation. Thus, it is possible that the duplication and rearrangement that gave rise to *CYP6A3* occurred before the duplication that gave rise to *CYP6A4* and *CYP6A3*.

CYP6A6 was less closely related to the other three CYP6A proteins in the cluster, sharing only 43–60% identity. CYP6A3, CYP6A4, and CYP6A5 were 62–72% identical to each other over the region included in the CYP6A6 fragment, *i.e.*, downstream of the intron.

Chromosomal localization

Both *CYP6A3* and *CYP6A6* were mapped by a PCR-RFLP procedure. Several differences in the sequence of these genes between the Rutgers and *aabys* strains were found to result in the loss or gain of recognition sites for restriction endonucleases. PCR products from backcross progeny of male F_1 Rutgers/*aabys* hybrids to female *aabys* were analyzed for RFLP patterns characteristic of the two parent strains (Fig. 4). Meiotic recombination is rare or absent in male house flies, and each parental chromosome is therefore inherited as an intact unit. Male and female backcross progeny that were homozygous for chromsomes I, II, III, or IV of the *aabys* strain showed the heterozygous restriction pattern for each of the two genes, while + + + + *s* flies, homozygous for chromosome V of the *aabys* strain, were homozygous for the *aabys* RFLP pattern (Fig. 4). These results demonstrated that the *CYP6* gene cluster is located on chromosome V of the house fly.

Gene expression

Northern blot analysis of poly(A)$^+$ RNA from house fly adults and larvae revealed that *CYP6A5*, with a transcript size of ~1.8 kb, was more highly expressed in larvae than adults (Fig. 5A). After standardization against actin mRNA, the relative abundance of CYP6A5 mRNA was found to be 30–40 times higher in larvae than in adults. The other five genes showed only faint bands on Northern blots of poly(A)$^+$RNA from either larvae or adults (data not shown). We did not have a complete sequence of *CYP6A6* and *CYP6C2*, and we could not exclude *a priori* the possibility that these sequences represent gene fragments or pseudogenes. Thus, the faint bands on Northern blots probed with these two sequences might be attributable to cross-reaction with related transcripts.

There were no differences in the expression of *CYP6A5* (Fig. 5A), or any of the other genes (data not shown), between the insecticide-susceptible *sbo* strain and three resistant strains of the house fly, Rutgers, R-Fc, and LPR (Fig. 5A), despite the fact that these three strains are known to have elevated P450 levels (Cariño *et al.*, 1992). In addition, we analyzed poly(A)$^+$ RNA from adult flies of the Rutgers strain reared for 24 hr in the presence of phenobarbital, dieldrin, DDT, piperonyl butoxide, ethanol, naphthalene, or β-naphthoflavone, at a concentration known to induce one or more P450 activities in the house fly (Cariño *et al.*, 1992). Neither *CYP6A5* (Fig. 5B) nor any of the other genes (data not shown) was inducible by any of these seven xenobiotic compounds.

Results of the Northern blot analysis prompted us to confirm by a more sensitive method that five of the genes were transcribed. Poly(A)$^+$RNA was reverse-transcribed

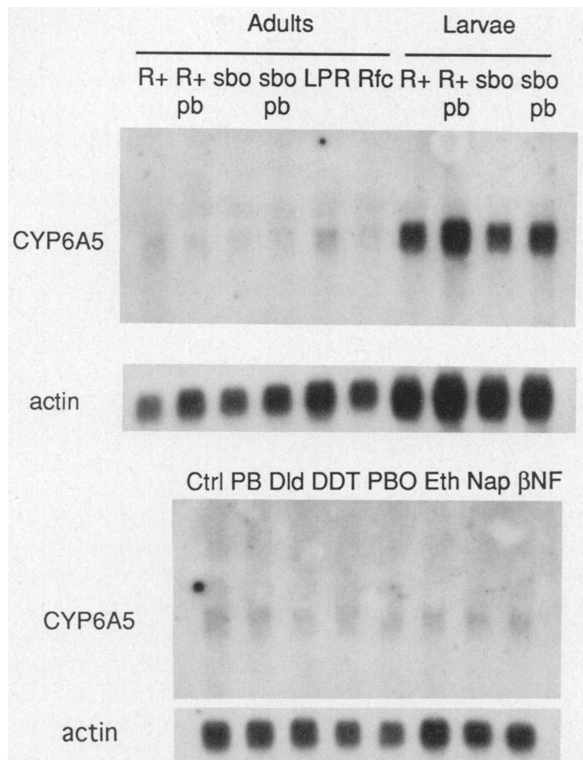

FIG. 5. Northern blots of house fly RNA probed with *CYP6A5*. A. Poly $(A)^+$RNA from larvae and adults of Rutgers (R+) and *sbo* strains reared on control or phenobarbital (pb) diets, and from adults of Rfc and LPR strains reared on control diet. The size of the transcript was approx. 1.8 kb. B. Poly$(A)^+$RNA from Rutgers adult flies reared on control (Ctrl), phenobarbital (PB), dieldrin (Dld), DDT, piperonyl butoxide (PBO), ethanol (Eth), naphthalene (Nap), or β-naphthoflavone (bNF) diets. The lower part of each panel shows hybridization of the same blot to the actin probe DmA2 as described (Cariño *et al.*, 1992, 1994).

with specific primers for each gene downstream from the common intron. The cDNA was then amplified by PCR with primers situated on each side of the intron thus allowing us to discriminate between the amplification of authentic cDNA and that of a possible genomic DNA contamination. This RT-PCR analysis (Table 3) demonstrated that *CYP6C1*, *CYP6C2*, *CYP6A3*, *CYP6A4*, and *CYP6A5* were transcribed and correctly spliced in the *sbo* strain. The selected primer combinations for *CYP6C1*, *CYP6C2*, *CYP6A3*, *CYP6A4*, and *CYP6A5* yielded RT-PCR products that matched the size of the expected cDNA sequences (Table 3). We did not have any sequence of the first exon for *CYP6A6*, and thus could not attempt to amplify the cDNA of spliced mRNA to establish whether this gene was transcribed.

Allelic variants

It was felt that the correct size of the PCR fragments did not constitute sufficient evidence for the identity of the cDNAs that had been amplified. This evidence was obtained when we cloned the RT-PCR products and sequenced several of the cloned products. The results showed that the RT-PCR products of *CYP6C1*, *CYP6C2*, *CYP6A3*, *CYP6A4*, and *CYP6A5* were authentic. However, we also found that additional products were formed for four of the genes (Table 3). These sequences were closely related to the target gene and may be allelic variants. We have cloned in this way fragments of allelic variants of the *CYP6A4* and *CYP6A5* genes that differed by 5–7% in their nucleotide sequence. For *CYP6C1*, two such allelic variants were found, one of which had an unspliced intron with five changes in the 64-bp intron. We do not know whether the latter variant was obtained by PCR of contaminating genomic DNA, or represented an unspliced transcript. A fragment of an unrelated P450 gene was amplified with the primers for *CYP6C2*. The deduced amino acid sequence of this PCR product was only 53% identical to that of CYP6C2. This new gene is currently being characterized.

DISCUSSION

We have identified and characterized a cluster of P450 genes from the house fly by screening a genomic library with a probe for a known insect P450, *CYP6A1*. A multiplicity of

TABLE 3. PCR AMPLIFICATION[a] AND CLONING OF CYP6 cDNAS AND ALLELIC VARIANTS

			Allelic variants[c]		
				Identity to cognate gene	
Gene	Expected length (bp) of RT-PCR product (genomic PCR product)	Expected sequence found[b]	Clones found	Nucleotides	Amino acids
CYP6A3	667 (729)	+	0		
CYP6A4	989 (1,055)	+	2	95%	100%
CYP6A5	1,536 (1,593)	+	1	93.2%	96.6%
CYP6C1	618 (682)	+	2	99.2%	96.7%
			1[d]	96.2%	97.7%
CYP6C2	938 (1,063)	+	0		

[a]Primer sequences are listed in the Materials and Methods section.
[b]The expected haplotype sequence with correct splicing of the intron was found in at least two independent clones.
[c]Clones for which nucleotide (and amino acid) changes from the cognate haplotype sequence were found.
[d]The sequence of this variant showed an unspliced intron (see text) with 5 nucleotide changes. The percent identity is given for the coding sequence only.

P450 genes was not unexpected, but P450 gene clusters have been described only in a few cases, and the *CYP6* gene cluster therefore offered an interesting insight into the structure, expression, and evolution of insect P450 genes.

A first observation was that *CYP6A1* was not part of the cluster, even though it was the probe used for screening. It is possible that the cluster extends beyond the region covered by the genomic clones that we have analyzed and that *CYP6A1* is located on a part of the cluster not yet isolated. However, we have surveyed 11 kb of DNA upstream of *CYP6C2* and 2 kb upstream of *CYP6A6* for the presence of additional P450 genes (and for the missing portion of those genes flanking the cluster) and none were found. We have also sequenced 1.5 kb of DNA upstream of *CYP6A1* and 0.5 kb downstream (Cohen *et al.*, 1994) without finding sequences that matched the cluster. On the other hand, *CYP6A1* and the *CYP6* cluster are both localized on chromosome V of the house fly (Fig. 4 and Cohen *et al.*, 1994) and they are clearly derived from a common ancestor (Fig. 3). Future work, for instance using pulsed-field gel electrophoresis, may reveal the genomic distance between *CYP6A1* and the cluster described here. Alternatively, the genomic organization of P450 genes for the *CYP6* family may be elucidated by *in situ* hybridization to polytene chromosomes of the fruit fly, *Drosophila melanogaster*. The latter approach may be of particular interest because a cluster of *CYP6* genes is likely to be found in other Diptera, as will be discussed below.

The high degree of sequence divergence between the members of the *CYP6* cluster suggests that this cluster is evolutionarily ancient. The rate of evolution of P450 genes is not known in insects, and there are no obvious orthologous P450 genes that can be used as markers for divergence. If a UEP (unit evolutionary period = millions of years needed for 1% divergence of amino acid sequence) value of 2.8 is used (Nelson and Strobel, 1987), then the genes in the cluster may have diverged between 80 and 170 Mybp (million years before present). This low UEP is thought to represent the rapid P450 evolution in mammalian species. A comparison of trout and mouse CYP1A1 leads to a UEP value of 9.0, and by using this larger UEP, a three-fold greater divergence time would be estimated. If *CYP6A1* (Feyereisen *et al.*, 1989) and *CYP6A2* (Waters *et al.*, 1992) are considered orthologous genes, their 51% divergence would lead to a 1.96 UEP. Tentative arguments for an orthology of *CYP6A1* and *CYP6A2* rely on structural and regulatory features and have been presented elsewhere (Feyereisen *et al.*, 1994). The lineages to *Musca domestica* and *Drosophila melanogaster* are believed to have diverged approximately 100 Mybp (Beverley and Wilson, 1984). Thus, even in the case of a very rapid evolution of P450s in insects, it is likely that a homologous P450 cluster will be found on 2R in *Drosophila*, in all Diptera, and possibly in other homometabolous insects. Chromosome V of the house fly is homologous to the right arm of chromosome 2 of *Drosophila* (Foster *et al.*, 1981).

The rate of P450 evolution has been nonlinear, and has apparently been influenced by gene conversion events (Gotoh, 1993; Nelson *et al.*, 1993). Stretches of high sequence identity with abrupt changes to regions of lower identity indicate that several recent gene conversions have occurred among the genes of the *CYP2D* cluster (Matsunaga *et al.*, 1990), but such patterns are not found in the *CYP6* cluster. This suggests that the *CYP6* cluster is much older than the rat *CYP2D* cluster, and/or that gene duplication in the *CYP6* cluster was followed by rapid divergence. Divergence among duplicated genes can be accelerated by the removal of selective constraints or by positive selection for new function (Walsh, 1987). Gotoh (1993) has studied the pattern of excess of nonsynonymous over synonymous nucleotide changes in the CYP2 family and showed a remarkable excess of nonsynonymous changes in the "substrate recognition sites" (SRS). This strongly supports the postulate of positive selection for new function. If microsomal P450s evolve mainly in response to new chemicals insults from the environment, then P450 clusters may be the relic of stepwise selection for new functions by duplication and divergence. The substrate specificity of different P450s of the *CYP6* cluster is not known. Reconstitution of CYP6A1 as a cyclodiene insecticide epoxidase has recently been achieved (Andersen *et al.*, 1994), but preliminary results (Andersen, Stevens, and Feyereisen, unpublished results) indicate that CYP6A5 has a very different substrate specificity.

Other P450 gene clusters have been reported in yeast and in mammals. In the yeast, *Candida tropicalis*, two tandem arrays of P450 genes have been described, each covering about 5 kb of DNA: *CYP52A1* and *CYP52A2*, which code for proteins that are 68% identical, and *CYP52A8* and *CYP52B1*, which are 41% identical (Seghezzi *et al.*, 1992). The *CYP2D* gene cluster in the rat contains four genes, *CYP2D2, 2D3, 2D4, and 2D5* (Matsunaga *et al.*, 1990). All four genes are oriented in the same way, and the proteins are 79–84% identical. The human *CYP2D* cluster contains *CYP2D6* and two pseudogenes *2D8P* and *2D7P* the latter being present in one or two nonallelic variants (Heim and Meyer, 1992). In this cluster, all genes and pseudogenes are also in the same orientation, and the deduced amino acid sequences are 87–99% identical. The human *CYP2D* cluster is believed to have started with a gene duplication about 18 million years ago that led to *CYP2D8P* and the *CYP2D6* and *2D7* genes (Heim and Meyer, 1992). Most recently, Johansson *et al.* (1994) have documented an inherited 12-fold amplification of the *CYP2D6L* variant in individuals with ultrarapid metabolism of debrisoquine. It appears that *CYP2D* genes may have diverged independently as nonorthologous clusters in different mammalian species. Other mammalian P450 clusters include the *CYP2A/2B/2F* cluster, which comprises many genes, all involved with the metabolism of xenobiotics. The pattern of positive selection, duplication, and divergence may be invoked for this cluster as well. In contrast, the *CYP1/CYP11A/CYP19* cluster in the mouse and in humans (Nelson *et al.*, 1993) may have a different evolutionary origin. In humans, for instance, these genes are clustered between 15q21 and 15q24, but they are very distant on every published P450 phylogeny (*e.g.*, Nebert *et al.*, 1989; Gotoh, 1993). The reason for this nonrandom genomic organization is not easily discerned. Very recently, Frolov and Alatortsev (1994) reported the presence of a P450 gene, *CYP4D2*, at about 20 kb from the *CYP4D1* gene (Gandhi *et al.*, 1992) on the X chromosome of *Drosophila*. The two genes are in opposite orientation, and their products are 58% identical.

The relative orientation of the genes in the *CYP6* cluster

may be the result of tandem duplication events but also of chromosomal inversion events, which are common in Diptera. Chromosomal inversion events may be responsible for the disruption of the two genes, *CYP6C2* and *CYP6A6*, situated at each end of the cluster. Both these genes appear to be interrupted at the site of an intron. In the case of *CYP6C2*, evidence that the gene is functional was obtained by RT-PCR and sequencing of a portion of the cDNA. The precise location of the first intron was only tentatively assigned and its length is unknown. In the case of *CYP6A6*, only the second exon was found, although it is possible that the 2.0 kb upstream of the *CYP6A6* fragment is part of a long intron and that the gene is functional. The intron, common to *CYP6A* and *CYP6C* genes, that interrupts the conserved ETLR motif is 57–125 bp long, *i.e.*, a length typical of introns found in *Drosophila* (Mount *et al.*, 1992). Longer introns are considerably less frequent in *Drosophila*. However, the genome size of *Musca domestica* is 10 times larger than that of *Drosophila*, and its genomic organization is clearly different (Cockburn and Mitchell, 1989). Furthermore, there is no correlation between intron size and genomic size. The precise structure of *CYP6A6*, *i.e.*, unusually long intron or gene disruption, thus remains uncertain.

We have obtained evidence that each of the *CYP6* genes described here, except for *CYP6A6*, are functional genes. The deduced amino acid sequences have all the features of normal microsomal P450s, although CYP6C1 and CYP6C2 share an Ala instead of the common Gly found as the second residue past the invariant Cys. Their heme binding motif is therefore F--G---C-A instead of F--G---C-G. This Gly is absolutely conserved among P450s except in CYP55 from the fungus *Fusarium oxysporum* (Kizawa *et al.*, 1991), in CYP113 from *Saccharopolyspora erythraea* (Stassi *et al.*, 1993), and in two P450-like sequences from *Arabidopsis thaliana* (GenBank Z27299 and Z17988) where it is replaced by Ala, and in CYP71A6 from the same plant where it is replaced by Ile (GenBank Z17988). Site-directed mutagenesis of this Gly to glu in rat liver CYP1A2 abolished heme binding (Shimizu *et al.*, 1988), but this Ala for Gly substitution is not expected to disrupt heme binding severely. It may slightly affect the orientation of the heme in the catalytic site.

CYP6A1 is overexpressed in several insecticide-resistant house fly strains, including Rutgers and LPR (Cariño *et al.*, 1992, 1994) and *CYP6A2* is overexpressed in some resistant *Drosophila* strains (Waters *et al.*, 1992). However, the genes in the *CYP6* cluster did not appear to have a role in insecticide resistance in the three strains, Rutgers, R-Fc and LPR, that we tested. Overexpression of P450 and resistance in the Fc strain, therefore, is caused by a P450 gene (or genes) that has not been identified yet. Moreover, none of the genes appear to be inducible by phenobarbital in the adult stage, and we failed to identify a barbie box (Shaw and Fulco, 1993) upstream of the putative TATA box of these genes. Two barbie boxes were observed in the 5′ region of the *CYP6A1* and *CYP6A2* genes, and these two genes are inducible by phenobarbital (Cohen *et al.*, 1994). The lack of induction of the clustered *CYP6* genes by phenobarbital and by other inducers indicates that the patterns of P450 induction in the house fly by these chemicals (dieldrin, DDT, β-naphthoflavone, naphthalene, ethanol) is the result of the induction of P450 genes that have not been identified yet. Thus, insecticide resistance and P450 induction are two paradigms that should continue to stimulate the search for new insect P450 genes.

ACKNOWLEDGMENTS

This work was supported by National Institute of Health grant GM39014. We thank J.F. Koener for the *sbo* λ DASH library, F.A. Cariño for fly mRNA samples, J. Longbottom and C. Summers for technical assistance, J. Scott (Cornell University) for the *aabys* strain and LPR flies, and D. Nelson (University of Tennessee, Memphis) for advice in naming the genes.

REFERENCES

ANDERSEN, J.F., UTERMOHLEN, J.G., and FEYEREISEN, R. (1994). Expression of house fly CYP6A1 and NADPH-cytochrome P450 reductase in *Escherichia coli* and reconstitution of an insecticide-metabolizing P450 system. Biochemistry **33,** 2171–2177.

BEVERLEY, S.M., and WILSON, A.C. (1984). Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. J. Mol. Evol. **21,** 1–13.

CARIÑO, F., KOENER, J.F., PLAPP, F.W., JR., and FEYEREISEN, R. (1992). Expression of the cytochrome P450 gene CYP6A1 in the house fly, *Musca domestica*. ACS Symp. Ser. **505,** 31–40.

CARIÑO, F.A., KOENER, J.F., PLAPP, JR., F.W., and FEYEREISEN, R. (1994). Constitutive overexpression of the cytochrome P450 gene CYP6A1 in a house fly strain with metabolic resistance to insecticides. Insect Biochem. Mol. Biol. **24,** 411–418.

CHERBAS, L., and CHERBAS, P. (1993). The arthropod initiator: the capsite consensus plays an important role in transcription. Insect Biochem. Mol. Biol. **23,** 81–90.

COCKBURN, A.F., and MITCHELL, S.E. (1989). Repetitive DNA interspersion patterns in Diptera. Arch. Insect Biochem. Physiol. **10,** 105–113.

COHEN, M.B., SCHULER, M.A., anD BERENBAUM, M.R. (1992). A host-inducible cytochrome P-450 from a host-specific caterpillar; Molecular cloning and evolution. Proc. Natl. Acad. Sci. USA **89,** 10920–10924.

COHEN, M.B., KOENER, J.F., and FEYEREISEN, R. (1994). Structure and chromosomal localization of *CYP6A1*, a cytochrome P450-encoding gene from the house fly. Gene **146,** 267–272.

DAVIS, L.G., DIBNER, M.D., and BATTEY, J.F. (1986). *Basic Methods in Molecular Biology*. (Elsevier, New York).

FEYEREISEN, R., KOENER, J.F., FARNSWORTH, D.E., and NEBERT, D.W. (1989). Isolation and sequence of cDNA encoding a cytochrome P-450 from an insecticide-resistant strain of the house fly, *Musca domestica*. Proc. Natl. Acad. Sci. USA **86,** 1465–1469.

FEYEREISEN, R., ANDERSEN, J.F., CARIÑO, F.A., COHEN, M.B., and KOENER, J.F. (1994). Cytochrome P450 in the house fly: structure, catalytic activity and regulation of expression of *CYP6A1* in an insecticide resistant strain. Pesticide Sci. (in press).

FOSTER, G.G., WHITTEN, M.J., KONOVALOV, C., ARNOLD, J.T.A., and MAFFI, G. (1981). Autosomal genetic maps of the Australian sheep blowfly, *Lucilia cuprina dorsalis* R.-D.

(Diptera: Calliphoridae), and possible correlations with the linkage maps of *Musca domestica* L. and *Drosophila melanogaster* (Mg.). Genet. Res. Camb. **37,** 55–69.

FROLOV, M.V., and ALATORTSEV, V.E. (1994). Cluster of cytochrome P450 genes on the X chromosome of *Drosophila melanogaster*. DNA Cell Biol. **13,** 663–668.

GANDHI, R., VARAK, E., and GOLDBERG, M.L. (1992). Molecular analysis of a cytochrome P450 gene of family 4 on the *Drosophila* X chromosome. DNA Cell Biol. **11,** 397–404.

GISH, W., and STATES, D.J. (1993). Identification of protein coding regions by database similarity search. Nature Genetics **3,** 266–272.

GONZALEZ, F.J., and NEBERT, D.W. (1990). Evolution of the P450 gene superfamily: animal-plant "warfare", molecular drive, and human genetic differences in drug oxidation. Trends Genet. **6,** 182–186.

GOTOH, O. (1993). Evolution and differentiation of P-450 genes. In *Cytochrome P-450*. T. Omura, Y. Ishimura, and Y. Fujii-Kuriyama, eds. (VCH Publishers, Weinheim) pp. 255–272.

HEIM, M.H., and MEYER, U.A. (1992). Evolution of a highly polymorphic human cytochrome P450 gene cluster: CYP2D6. Genomics **14,** 49–58.

JOHANNSON, I., LUNDQVIST, E., BERTILSSON, L., DAHL, M.L., SJÖQVIST, F., and INGELMAN-SUNDBERG, M. (1993). Inherited amplification of an active gene in the cytochrome P450 *CYP2D* locus as a cause of ultrarapid metabolism of debrisoquine. Proc. Natl. Acad. Sci. USA **90,** 11825–11829.

KIZAWA, H., TOMURA, D., ODA, M., FUKAMIZU, A., HOSHINO, T., GOTOH, O., YASUI, T., and SHOUN, H. (1991). Nucleotide sequence of the unique nitrate/nitrite-inducible cytochrome P-450 cDNA from *Fusarium oxysporum*. J. Biol. Chem. **266,** 10632–10637.

KOENER, J.F., CARIÑO, F.A., and FEYEREISEN, R. (1993). The cDNA and deduced protein sequence of house fly NADPH-cytochrome P450 reductase. Insect Biochem. Mol. Biol. **23,** 439–447.

MA, R., COHEN, M.B., BERENBAUM, M.R., and SCHULER, M.A. (1994). Black swallowtail (*Papilio polyxenes*) alleles encode cytochrome P450s that selectively metabolize linear furanocoumarins. Arch. Biochem. Biophys. **310,** 332–340.

MATSUNAGA, E., UMENO, M., and GONZALEZ, F.J. (1990). The rat P450 IID subfamily: Complete sequences of four closely linked genes and evidence that gene conversions maintained sequence homogeneity at the heme-binding region of the cytochrome P450 active site. J. Mol. Evol. **30,** 155–169.

MOUNT, S.M., BURKS, C., HERTZ, G., STORMO, G.D., WHITE, O., and FIELDS, C. (1992). Splicing signals in *Drosophila:* Intron size, information content, and consensus sequences. Nucleic Acids Res. **20,** 4255–4262.

NEBERT, D.W., NELSON, D.R., and FEYEREISEN, R. (1989). Evolution of the cytochrome P450 genes. Xenobiotica **19,** 1149–1160.

NELSON, D.R., and STROBEL, H.W. (1987). Evolution of cytochrome P-450 proteins. Mol. Biol. Evol. **4,** 572–593.

NELSON, D.R., KAMATAKI, T., WAXMAN, D.J., GUENGERICH, F.P., ESTABROOK, R.W., FEYEREISEN, R., GONZALEX, F.J., COON, M.J., GUNSALUS, I.C., GOTOH, O., OKUDA, K., and NEBERT, D.W. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol. **12,** 1–51.

RONIS, M.J.J., HODGSON, E., and DAUTERMAN, W.C. (1988). Characterization of multiple forms of cytochrome P-450 from an insecticide resistant strain of house fly (*Musca domestica*). Pest. Biochem. Physiol. **32,** 74–90.

SEGHEZZI, W., MEILI, C., RUFFINER, R., KUENZI, R., SANGLARD, D., and FIECHTER, A. (1992). Identification and characterization of additional members of the cytochrome P450 multigene family CYP52 of *Candida tropicalis*. DNA Cell Biol. **11,** 767–780.

SHAW, G.-C., and FULCO, A.J. (1993). Inhibition by barbiturates of the binding of BM3R1 repressor to its operator site on the barbiturate-inducible cytochrome P450BM-3 gene of *Bacillus megaterium*. J. Biol. Chem. **268,** 2997–3004.

SHIMIZU, T., HIRANO, K., TAKAHASHI, M., HATANO, M., and FUJII-KURIYAMA, Y. (1988). Site-directed mutageneses of rat liver cytochrome P-450_d: Axial ligand and heme incorporation. Biochemistry **27,** 4138–4141.

STASSI, D., DONADIO, S., STAVER, M.J., and KATZ, L. (1993) Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. J. Bacteriol. **175,** 182–189.

WALSH, J.B. (1987). Sequence-dependent gene conversion: Can duplicated genes diverge fast enough to escape conversion? Genetics **117,** 543–557.

WATERS, L.C., ZELHOF, A.C., SHAW, B.J., and CH'ANG, L.Y. (1992). Possible involvement of the long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. Proc. Natl. Acadi. Sci USA **89,** 4855–4859.

Address reprint requests to:
Dr. R. Feyereisen
Department of Entomology
University of Arizona
Forbes 410
Tucson, AZ 85721

Received for publication July 6, 1994; accepted August 31, 1994.