ELSEVIER

Gene 272 (2001) 257–266

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

CYP9E2, *CYP4C21* and related pseudogenes from German cockroaches, *Blattella germanica*: implications for molecular evolution, expression studies and nomenclature of P450s

Zhimou Wen, Christine E. Horak, Jeffrey G. Scott*

Department of Entomology, Comstock Hall, Cornell University, Ithaca, NY 14853-0901, USA

Received 2 January 2001; received in revised form 12 March 2001; accepted 14 May 2001
Received by W. Makalowski

Abstract

The cDNAs of two novel P450s (CYP9E2 and CYP4C21) were isolated from German cockroaches, *Blattella germanica*. Both CYP9E2 and CYP4C21 are typical microsomal P450s and their deduced amino acid sequences share a number of common characteristics with other members of the P450 superfamily. Northern blot analyses using a CYP9E2 or CYP4C21 probe showed that 'CYP9E2' and 'CYP4C21' were expressed at all life stages. Two pseudogenes related to *CYP9E2* and three pseudogenes related to *CYP4C21* were also isolated. These represent the first P450 pseudogenes from an insect other than *Drosophila melanogaster*. The relative number of P450 pseudogenes in *B. germanica* is apparently higher than in *D. melanogaster*. The implications of these results for the molecular evolution, expression studies and nomenclature of P450s are discussed.

Keywords: Cytochrome P450 monooxygenase; P450 pseudogenes; Evolution; *Blattella germanica*; Insecta

1. Introduction

The cytochrome P450 monooxygenases are a very important enzymatic system involved in the metabolism of structurally diverse endogenous and exogenous compounds (Mansuy, 1998). The presence of P450s in diverse organisms, from bacteria to plants and animals, implies that the P450 superfamily is an extremely ancient enzymatic system and that all the current P450s may have descended from a common ancestral gene (Nebert and Gonzalez, 1987). The current P450 superfamily is thought to have been formed by gene duplication and adaptive diversification (Gotoh, 1993). The concept of molecular drive has also been used to explain the formation of new P450 genes (Gonzalez and Nebert, 1990). Molecular drive is the processes by which the genotypic composition of a population can be changed as a result of the internal dynamics of DNA turnover independent of natural selection (Dover and Tautz, 1986). This continuous dynamic DNA turnover can be generated by several mechanisms, including gene duplication, unequal crossover, gene conversion, transposition, slippage replication and RNA-mediated transfers (Dover and Tautz, 1986; Gonzalez and Nebert, 1990; Fogleman et al., 1997).

During the processes of evolution, pseudogenes may be generated. A pseudogene is a sequence which is similar to a functional gene, but does not produce a functional product (enzymatically active protein in the case of a P450). A pseudogene can be identified by either its aberrant coding region, transcriptional silence, or both (Wilde, 1986). Pseudogenes are mainly the consequence of gene duplication arising either by abnormal genomic DNA duplication or by retrotransposition (Mighell et al., 2000). Pseudogenes can result from random mutations. P450 pseudogenes with aberrant coding regions have been described in various organisms. However, only four have been identified in insects and all are from *Drosophila melanogaster* (Adams et al., 2000). A P450 pseudogene is indicated by a 'P' suffix (Nelson et al., 1996).

We report in this paper the isolation of two novel P450s and their related pseudogenes from German cockroaches, *Blattella germanica*. The implications of the results are discussed in terms of P450 molecular evolution, expression studies and nomenclature.

Abbreviations: MCH, cDNA clones from male German cockroaches obtained with the degenerate heme primer; P450, cytochrome P450; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-PCR; SDS, sodium dodecyl sulfate

* Corresponding author. Tel.: +1-607-255-7340; fax: +1-607-255-0939.
E-mail address: jgs5@cornell.edu (J.G. Scott).

0378-1119/01/$ - see front matter
PII: S0378-1119(01)00529-7

2. Materials and methods

2.1. Insects, RNA isolation and cDNA synthesis

German cockroaches (Baygon-R strain) were reared as described previously (Siegfried and Scott, 1991). Total RNA was isolated using 5 M guanidine thiocyanate solution (Chirgwin et al., 1979). mRNA was isolated from total RNA using Oligotex suspension (QIAGEN) or directly from tissues using a QuickPrep mRNA purification kit (Pharmacia) as described by the manufacturers. Superscript II (GIBCO/BRL) was used to synthesize the first strand cDNA following the manufacturer's instructions. For the first strand cDNA synthesis, C3PT (Danielson and Fogleman, 1997) was used as primer and ~500 ng mRNA from abdomens of male adult cockroaches was used as template. After RNAase H (2 units) treatment at 37°C for 30 min, the first strand cDNA was isolated using a QIAquick PCR purification kit (QIAGEN) to remove the primers and short cDNA (eluted with 100 μl H_2O). The purified cDNA was used as template for 3′ and 5′ RACE (Frohman et al., 1988).

2.2. cDNA cloning

The procedures for cDNA isolation using 3′ and 5′ RACE techniques have been described previously (Wen and Scott, 2001). The specific points pertaining to isolation of the cDNAs for CYP9E2, CYP4C21 and their related pseudogenes are described below.

2.3. CYP9E2 and related pseudogenes

Fig. 1 outlines the approach used to clone CYP9E2 and its related pseudogenes. Briefly, 3′ RACE led to the isolation of MCH9, a clone that encoded the C-terminal amino acid sequence of a putative P450, MCHD. Using a gene-specific primer (MCHDA1, Table 1) based on the sequence of MCH9, the 5′ cDNA sequences of MCHD were cloned after three rounds of 5′ RACE. All of the adjacent sequences obtained by 3′ and 5′ RACE had at least 150 bp overlap with each other. By merging the overlapping sequences from the 3′ and 5′ RACE, MCHDMER, the putative full-length cDNA sequence for MCHD, was generated.

To ensure that the fragments used to generate MCHDMER were from the same gene, a gene-specific primer set (MCHDS4/MCHDA1, Table 1) was used to amplify the 'full-length' cDNA of MCHD in two rounds of PCR. The first round PCR used first strand cDNA as template. The products from the first round PCR (not visible following agarose electrophoresis) served as templates to perform the second round PCR. The conditions for the first round PCR were 95°C for 3 min to denature, followed by 32 cycles at 95°C for 1 min, 50°C for 1 min and 72°C for 2 min. The PCR reaction had a final extension time of 15 min at 72°C. The conditions (and primers) for the second round PCR were the same as the first, except that 35 cycles were used. Two bands (1.6 and 1.2 kb) were always detected

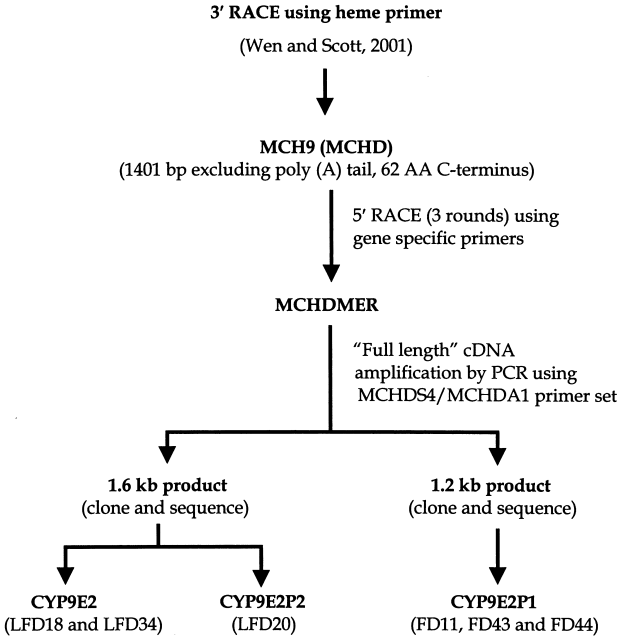


Fig. 1. Schematic representation of the cloning of *CYP9E2* and its related pseudogenes.

from the second round PCR. Both bands were cut, and the PCR products were purified and cloned into pCR® 2.1 vector (Invitrogen). Three clones for the 1.6 kb products (LFD18, LFD20 and LFD34) and three clones for the 1.2 kb products (FD11, FD43 and FD44) were identified by colony PCR (Gussow and Clackson, 1989) using MCHDS4/MCHDA1 as the primer set. The conditions for colony PCR were the same as the first round PCR except that the initial denaturation lasted for 10 min. The inserts of these clones were entirely sequenced. Based on sequence analysis LFD18 and 34 were later named CYP9E2 (= MCHDMER), LFD20 was named CYP9E2P2, and FD11, 43 and 44 were named CYP9E2P1.

2.4. CYP4C21 and related pseudogenes

Similar to the isolation of MCH9, a clone that encoded the C-terminal amino acid sequence of a putative P450, BAYC6, was isolated by 3′ RACE. Using a gene-specific primer (BAYC6A1, Table 1) based on the sequence of

Table 1
Primers used for cloning cDNAs of CYP4C21 (BAYC6), CYP9E2 (MCHD), and their related pseudogenes from German cockroaches

CYP	Primers	Sequences (5′–3′)	Positions
9E2	MCHDS4	TTCTGTGCTTTCTACAGAATA	117/137
	MCHDA1	CCAGAAACCTCCATCCACAG	1715/1734
4C21	BAYC6S5	AGTAACGAACCGAGAGCTACCG	1/22
	BAYC6A1	GAATAGGCTTGACAGGGGTG	1486/1505
	BAYC6A3	GTGATGACCTAAAAGAAAAACAG	1553/1575
	BAYC6A6	ATGCGATCAGAGAGGACATCCG	651/672
4C21P2	BAYC6S1	AGTAACAATCTAAATGCCAAAC	223/244

BAYC6, clone CHBAY (i.e. BAYC6-5′, 1408 bp) was obtained. CHBAY and BAYC6 have four nucleotide differences in the 154 bp overlapping region (between heme primer and BAYC6A1), suggesting that these two pieces might not be from the same gene. The CHBAY sequence likely represents part of a pseudogene because translation of CHBAY reveals 18 stop codons N-terminal to the heme binding region. CHBAY was later named *CYP4C21P2*. We also screened a German cockroach cDNA library (UniZAP-XR phagemid expression vector) (Arruda et al., 1995) kindly provided by Dr M. Chapman (University of Virginia Health Sciences Center) using the BAYC6A1/BAYC6S1 (Table 1 and Fig. 3B) fragment of CHBAY (~1.2 kb) as the probe. At least 30 positive plaques were obtained from the first 5×10^7 plaques that were screened. To confirm the identity of these positive plaques, the BAYC6A1/BAYC6S1 primer set was used to run plaque PCR (Gussow and Clackson, 1989) with seven randomly picked positive plaques as templates. A band of ~560 bp was generated from all the positive plaques rather than the expected ~1.2 kb product based on the CHBAY sequence. We cloned the PCR products using BAYC6A1/BAYC6S1 as the primer set and the first strand cDNA as template. The inserts of two positive clones (WSH4 and WSH5) were sequenced. These clones represent yet another pseudogene, later to be named *CYP4C21P3*.

Given our success with the 5′ RACE technique (Section 2.3), we used a similar approach to investigate if a functional BAYC6 gene existed in German cockroaches. Two rounds of 5′ RACE (using gene-specific primers BAYC6A1 and BAYC6A6 (Table 1) for the first and second round, respectively) led to the isolation of sequences beyond the translational start codon. All of the adjacent sequences by 3′ and 5′ RACE had at least 150 bp overlap with each other. BAYMER, the putative full-length cDNA sequence for BAYC6, was generated by merging the overlapping sequences.

To ensure that the fragments used to generate BAYMER were from the same gene, two sets of gene-specific primers were used to amplify the 'full-length' cDNA using two rounds of PCR. The product of first round PCR, using BAYC6S5/BAYC6A3 (Table 1) as the primer set and first strand cDNA as template, served as template for the second round PCR using BAYC6S5/BAYC6A1 as the primer set. The PCR conditions were the same as previously described (Section 2.3). Two clones with inserts of 1.5 kb (YJ1 and YJ2) and another two clones with inserts of 1.2 kb (PW1 and PW2) were obtained. The inserts of all these clones were fully sequenced. Based on the sequence analyses, clones YJ1 and YJ2 were later named CYP4C21, while PW1 and PW2 were named CYP4C21P1.

2.5. Northern blot analyses

Northern blot analysis was performed by standard methods (Sambrook et al., 1989). Briefly, 10 μg of total RNA was fractionated on 1% denaturing formaldehyde agarose gel containing ethidium bromide. After washing in distilled water for about 3–4 h with several changes, the RNA was transferred to a GeneScreen Plus® hybridization transfer membrane (NEN™ Life Science Products, Inc.). Equal loading was monitored by comparing the density of the 18S ribosomal RNA (rRNA) band (Savonet et al., 1997; Spiess and Ivell, 1998) on the agarose gel before transfer and/or on the membrane after transfer under UV. Plasmids from the LFD18 clone were used as templates with the MCHDS4/MCHDA1 primer set to PCR-amplify a 1.6 kb fragment of CYP9E2 cDNA. Similarly, a 1.5 kb fragment of CYP4C21 cDNA was PCR-amplified using plasmids from the YJ1 clone with BAYC6S5/BAYC6A1 being the primer set. These PCR products were used as probes for Northern hybridizations. The membrane was hybridized to the probe in QuickHyb solution (Stratagene) at 68°C. Washing was done at high stringency (i.e. three 15 min washes with 2× SSC + 0.1% SDS at room temperature, followed by a 30 min wash with 0.2× SSC + 0.1% SDS at 60°C). The membrane was air-dried and exposed to a BioMax MR film (Eastman Kodak). All Northern analyses were repeated at least three times with independent preparations of RNA.

2.6. Sequence analyses

The comparisons of nucleotide or amino acid sequences were carried out using the Clustal method of the MegAlign program (DNASTAR Inc., Madison, WI).

3. Results

3.1. Full-length cDNAs of CYP9E2 and CYP4C21

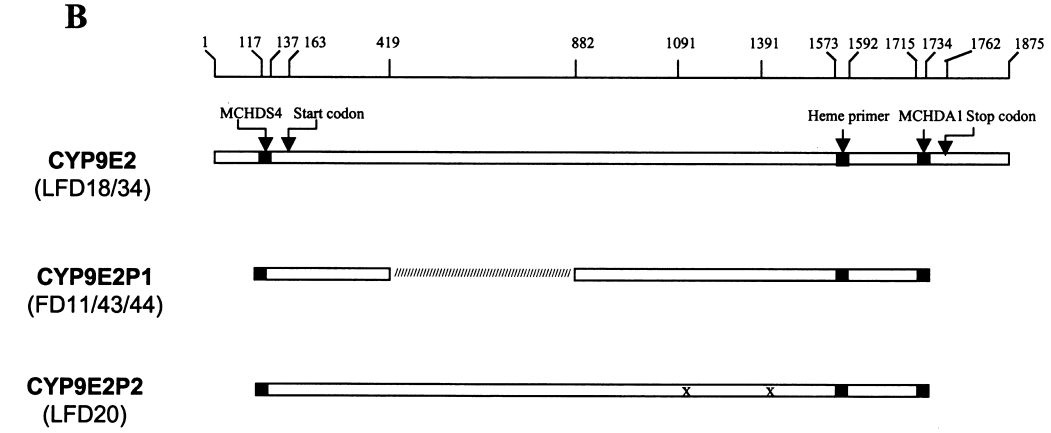
The full-length cDNA sequence for MCHD (based on clones LFD18 and LFD34 as well as 3′ and 5′ RACE results) was named *CYP9E2* (Accession number: AF275640) by the P450 Nomenclature Committee. *CYP9E2* has an open reading frame of 1599 nucleotides with a deduced protein of 533 amino acids and a molecular mass of 61.5 kDa. The CYP9E2 protein sequence has the highest percentage amino acid identity with Cyp9c1 (38.1%), Cyp9b2 (37.4%) and Cyp9f2 (35.9%) from *D. melanogaster* (Adams et al., 2000; Nelson, 2000), and CYP9A1 (34.8%) from tobacco budworm, *Heliothis virescens* (Rose et al., 1997).

The full-length cDNA sequence for BAYC6 (based on clones YJ1, YJ2 as well as BAYC6 and 5′ RACE results) was named *CYP4C21* (Accession number: AF275641). *CYP4C21* has an open reading frame of 1503 nucleotides with a deduced protein of 501 amino acids and a molecular mass of 58.3 kDa. CYP4C21 is most similar to CYP4C1 (47.7%) from *Blaberus discoidalis* (Bradfield et al., 1991), Cyp4c3 (42.1%) from *D. melanogaster* (Adams et al., 2000; Nelson, 2000) and CYP4C7 (38.0%) from *Diploptera punctata* (Sutherland et al., 1998).

A

```
9E2     TTCTGTGCTTTCTACAGAATAGCCAAACAATTATAAAGACAACAAAATGATGTCCCTGGAATCATTATGGTTATGGACTTTTGCAATCAGCTTCATCGTT  100
9E2P1   TTCTGTGCTTTCTACAGAATAGCCAAACAATTATAAAGACAACAAAATGATGTCCCTGGAATCATTATGGTTATGGACTTTTGCAATCAGCTTCATTGTT  100
9E2P2   TTCTGTGCTTTCTACAGAATAGCCAAACAATTATAAAAACAACAAAATGATGTCCCTGGAATCATTATGGTTATGGACTTTTGCAATCAGCTTCATTGTT  100

9E2     CTGATTGCTTACTTGATTGGGACATGGAACCATGATTTCTTCTCCAAGAGAAATATTCCCTCGCTTAAGACTGTACCATTTCTTGGTAATATGGGACCCC  200
9E2P1   CTGATTGCTTACTTAATTGGGACATGGAACCATGATTTCTTCTCCAAGAGAAATATTCCCTCGCTTAAGACTGTACCATTTCTTGGTAATATGGGACCCC  200
9E2P2   CTGATTGCTTACTTAATTGGGACATGGAACCATGATTTCTTCTCCAAGAGAAATATTCCCTCGCTTAAGACTGTACCATTTCTTGGTAATATGGGACCCC  200

9E2     TAGTTCTCAGGAAGGCATCCTTTGCAGAACATTCTCAAAACATTTACAACAGATTAAAAGGTTACAAATATGGTGGTATGTTTGAATTTATGAATCCAGT  300
9E2P1   TAGTTCTCAGGAAGGCATCCTTTGCAGAACATTCTCAAAACATTTACAACAGATTAAAAGGTTACAAATATGGTGGTATGTTTGAATTTATGAATCCAGT  300
9E2P2   TAGTTCTCAGGAAGGCATCCTTTGCAGAACATTCTCAAAACATTTACAACAGATTAAAAGGTTACAAATATGGTGGTATGTTTGAATTTATGAATCCAGT  300

9E2     TCTTGTGTTGCGTGATCCTGAACTTATCAAGATGGTGACTGTGAAAGATTTTGAATACTTTTTGGATCATCGAGCCCCAATATCTGAAGAAGCTGAACCT  400
9E2P1   TC----------------------------------------------------------------------------------------------  302
9E2P2   TCTTGTGTTGCGTGATCCTGAACTTATCAAGATGGTGACTGTGAAAGATTTTGAATACTTTTTGGATCGTCGAGCCCCAATATCTGAAGAAGCTGAACCT  400

9E2     ATGTTTGGCAAAAACTTGTTTAATTTAAGAGGACACAGATGGAAGGAAATGAGATCTACCTTGAGTCCTGCCTTCACATCCAGCAAGATGAAGAATATGT  500
9E2P1   ----------------------------------------------------------------------------------------------  302
9E2P2   ATGTTTGGCAAAAACTTGTTTAATTTAAGAGGACACAGATGGAAGGAAATGAGATCTACCTTGAGTCCTGCCTTCACATCCAGCAAGATGAAGAATATGT  500

9E2     TCGTGCTTGTCTCAGAATGTGGAAAACAACTTGGAGAGTTTTTGATGGAATGCAGTAGAGACAAAAATAAGAAGACTGAAGGCTGTAAAATAGAAAGAGA  600
9E2P1   ----------------------------------------------------------------------------------------------  302
9E2P2   TCGTGCTTGTCTCAGAATGTGGAAAACAACTTGGAGAGTTTTTGATGGAATGCAGTAGAGACAAAAATAAGAAGACTGAAGGCTGTAAAATAGAAAGAGA  600

9E2     GGGAGATCTACTTACAGTGGAACTGAAAGACTTATATACAAGATACACAAATGATGTCATTGCAACATCAGCATTTGGAATTGGTTGTGATTCCCTCAAG  700
9E2P1   ----------------------------------------------------------------------------------------------  302
9E2P2   GGGAGATCTACTTACAGTGGAGCTGAAAGACTTATATACAAGATACACAAATGATGTCATTGCAACATCAGCATTTGGAATTGGTTGCGATTCCCTCAAG  700

9E2     AATCCAAAGAATGAATTTTTTCAAATGGGCAAAGATGTTACAAACTTTGGTGGAATTAGACAGTTCATATTTTTGGGATACTTGTTTAGTCCTCGACTAA  800
9E2P1   ------------------------------------------------------------ATATTTTTGGGATACTTGTTTAGTCCTCGACTAA  336
9E2P2   AATCCAAAGAATGAATTTTTTCAAATGGGCAAAGATGTCACAAACTTTGGTGGAATTAGACAGTTCATATTTTTGGGATACTTGTTTAGTCCTCGACTAA  800

9E2     TGAAGTATTTGAACCTCAAATTCCTATCTTCAAAAGCTACAGAGTTCTTTCGATTCCTGGTGCATAATACAATGGACACAAGAAAAACTAAAGGCATTAT  900
9E2P1   TGAAGTATTTGAACCTCAAATTCCTATCTTCAAAAGCTACAGAGTTCTTTCGATTCCTGGTGCATAATACAATGGACACAAGAAAAACTAAAGGCATTAT  436
9E2P2   TGAAGTATTTGAACCTCAAATTCCTATCTTCAAAAGCTACAGAGTTCTTTCGATTCCTGGTGCATAATACAATGGACACAAGAAAAACTAAAGGCATTAT  900


9E2     ACGTCCAGATATGATCCATTTGTTGATGCAGGCCAAAGAGGGAACACTTAAGAGTGAAGAGAATGGAGAAACCAATGGAAAAATTGCATCCAAACCAAAA  1000
9E2P1   ACGTCCAGATATGATCCATTTGTTGATGCAGGCCAAAGAGGGAACACTTAAGAGTGAAGAGAATGGAGAAACCAATGGAAAAATTGCATCCAAACCAAAA  536
9E2P2   ACGTCCAGATATGATCCATTTGTTGATGCAGGCCAAAGAGGGAACACTTAAGAGTGAAGAGAATGGAGAAACCA-TGGAAAAATTGCATCCAAACCAAAG  999

9E2     TGGGACGATGACGATTTAACGGCACAAGCTGTTCTATTTTTCTTTGCTGGATTTGACACAGCATCTACTCTTCTCTGCTTTATGTCTCATCTGCTGGCTA  1100
9E2P1   TGGGACGATGACGATTTAACGGCACAAGCTGTTCTATTTTTCTTTGCTGGATTTGACACAGCATCTACTCTTCTCTGCTTTATGTCTCATCTGCTGGCTA  636
9E2P2   TGGGACGATGACGATTTAACGGCACAAGCTGTTCTATTTTTCTTTGCTGGATTTGACACAGCATCTACTCTTCTCTGCTTTATGTCTCATCTGCTGGCTA  1099

9E2     CAAATCCAGATGTGCAGAACAGACTCCAGGATGAGATTGATCAATCTTTGGAGGAGAATGATGGAAAGCTAACTTATGAAGCTATACACAGTATGAAATA  1200
9E2P1   CAAATCCAGATGTGCAGAACAGACTCCAGGATGAGATTGATCAATCTTTGGAGGAGAATGATGGAAAGCTAACTTATGAAGCTATACACAGTATGAAATA  736
9E2P2   CAAATCCAGATGTGCAGAACAGACTCCAGGATGAGATTGATCAATCTTTGGAGGAGAATGATGGAAAGCTAACTTATGAAGCTATACACAGTATGAAATA  1199

9E2     TCTTGATATGGTAGTCTCAGAATCTCTGAGATTATATCCTCCTGCAATTTTCACAGACCGAAAATGTGTAAAAAATTATAGGCTGCCCATGGAGCCATCA  1300
9E2P1   TCTTGATATGGTAGTCTCAGAATCTCTGAGATTATATCCTTCCTGCAATTTTCACAGACCGAAAATGTGTAAAAAATTATAGGCTGCCCATGGAGCCATCA  836
9E2P2   TCTTGATATGGTAGTCTCAGAATCTCTGAGATTATATCCTCCTGCAATTTTCACAGACCGAAAATGTGTAAAAA-TTATAGGCTGCCCATGGAGCCATCA  1298

9E2     TACACCTTAGAACCTGGAGATGCAGTTTGGATCCCCATCTATGCAATTCATCATGATCCAAAATATTATCCTAATCCAGAAAAATTTGATCCAGAGAGAT  1400
9E2P1   TACACCTTAGAACCTGGAGATGCAGTTTGGATCCCCATCTATGCAATTCATCATGATCCAAAATATTATCCTAATCCAGAAAAATTTGATCCAGAGAGAT  936
9E2P2   TACACCTTAGAACCTGGAGATGCAGTTTGGATCCCCATCTATGCAATTCATCATGATCCAAAATATTATCCTAATCCAGAAAAATTTGATCCAGAGAGAT  1398

9E2     TTAGTGATGAAAACAAGGATAACATTAAACCGTTTACTTATCTACCATTTGGATCAGGTCCTAGAAATTGCATTGGTAATAGATTTGCACTTATGGAATC  1500
9E2P1   TTAGTGATGAAAACAAGGATAACATTAAACCGTTTACTTATCTACCATTTGGATCAGGTCCTAGAAATTGCATTGGTAATAGATTTGCACTTATGGAATC  1036
9E2P2   TTAGTGATGAAAACAAGGATAACATTAAACCGTTTACTTATCTACCATTTGGATCAGGTCCTAGAAATTGCATTGGTAATAGATTTGCACTTATGGAATC  1498

9E2     CAAAATAGCTCTCGTCCATTTGTTATGTCGTTTTAATCTCAAAGTGGTCTCCAAGACACCAATACCAATAAAAATTACAAAGAAAGGCTTCAACATGACT  1600
9E2P1   CAAAATAGCTCTCGTCCATTTGTTATGTCGTTTTAATCTCAAAGTGGTCTCCAAGACACCAATACCAATAAAAATAACAAAGAAAGGCTTCAACATGACT  1136
9E2P2   CAAAATAGCTCTCGTCCATTTGTTATGTCGTTTTAATCTCAAAGTGGTCTCCAAGACACCAATACCAATAAAAATAACAAAGAAAGGCTTCAACACGACT  1598

9E2     GTGGATGGAGGTTTCTGG                                                                                  1618
9E2P1   GTGGATGGAGGTTTCTGG                                                                                  1154
9E2P2   GTGGATGGAGGTTTCTGG                                                                                  1616
```

B



Both CYP9E2 and CYP4C21 are typical microsomal P450s. Their deduced amino acid sequences share a number of common characteristics with other members of the P450 superfamily: the hydrophobic N-terminal region, the conserved signature motif FxxGxxxCxG in the heme binding region, the consensus sequence ((A/G)GxxT) within the I-helix region (EGHDT for CYP4C21), the charge pair consensus (ExxR) within the K-helix, the consensus (WxxxR) in the C-helix, P exactly 16 residues from T in the I-helix and YPDP (consensus (aromatic)xx(P/D)) consensus followed four residues later by PERF (PDNF for CYP4C21) (Nelson, 1998). In addition to the common characteristics shared with other members of the P450 family, CYP9E2 and CYP4C21 have characteristics that are unique to members of family 9 and family 4, respectively. The SR(F/I/L)(A/G)xx(Q/E) sequence immediately following the heme binding region, specific to family 9 (Fogleman et al., 1997), appears as NRFALME (position 478–484) in CYP9E2. The so called 'invariant' 13 amino acid residues (EVDTFMFEGHDTT) within the I-helix region, unique to family 4 (Bradfield et al., 1991), appear as EVHTFMFEGHDTV (position 302–314) in CYP4C21. The QK(F/Y)AxLE sequence immediately following the heme binding region, specific to family 4 (Fogleman et al., 1997), appears as QRFALLE (position 450–456) in CYP4C21.

3.2. Pseudogenes related to CYP9E2

The consensus sequence of clones FD11, FD43 and FD44 was a pseudogene, named *CYP9E2P1* (Accession number: AF276434). In addition to four nucleotide differences, the most striking difference between pseudogene *CYP9E2P1* and its homologous functional gene *CYP9E2* is that *CYP9E2P1* lacks a 464 bp fragment (Fig. 2) found in *CYP9E2*. The overall nucleotide identity between *CYP9E2P1* and *CYP9E2* within the MCHDS4/MCHDA1 region is 99.2%. When translated, CYP9E2P1 is a truncated protein of 92 amino acids with the first 85 at its N-terminus identical to the homologous N-terminal region of CYP9E2. If the 3′ region downstream of the deletion is independently translated in the proper reading frame, the resulting amino acid sequence is then 99.6% identical to the homologous C-terminal region of CYP9E2.

Clone LFD20 is another pseudogene homologous to *CYP9E2*. It was named *CYP9E2P2* (Accession number: AF276435). The overall nucleotide identity between *CYP9E2P2* and *CYP9E2* within the MCHDS4/MCHDA1 region is 98.8%. In addition to ten nucleotide differences, pseudogene *CYP9E2P2* has two deletions (Fig. 2). The lack of A^{1091} in *CYP9E2P2* results in a frame shift and leads to a truncated protein which shares a high percentage identity to the N-terminal region of CYP9E2. If the two positions with missing nucleotides in *CYP9E2P2* are filled with Xs, the translation would result in a protein that is 99.2% identical to CYP9E2. It is unlikely that the two base deletions in *CYP9E2P2* (positions A^{1091} and A^{1391} in *CYP9E2*) were due to sequencing errors because the region surrounding the deletions of LFD20 has been sequenced from both directions. It is possible that *CYP9E2P2* (LFD20) is the result of base pair slippage (i.e. *Taq* error) during PCR amplification (Tindall and Kunkel, 1988), but the probability of this seems remote based on experience with PCR in this lab over the last decade (Kasai and Scott, 2001).

3.3. Pseudogenes related to CYP4C21

The sequences of PW1 and PW2 were a pseudogene closely related to *CYP4C21* which was named *CYP4C21P1* (Accession number: AF281326). In addition to 22 nucleotide differences, the most striking difference between pseudogene *CYP4C21P1* and its homologous functional gene is that *CYP4C21P1* lacks a 316 bp fragment (Fig. 3) found in *CYP4C21*. The overall nucleotide identity between *CYP4C21P1* and *CYP4C21* within the BAYC6S5/BAYC6A1 region is 97.7%. The 316 bp deletion in *CYP4C21P1* results in not only an overall shortened nucleotide sequence, but also a frame shift leading to a truncated protein of 102 amino acids. This truncated protein is 97.0% identical to the corresponding N-terminal region of CYP4C21. If the 3′ region downstream of the deletion is independently translated in the proper reading frame, the resulting protein is 97.2% identical to the corresponding C-terminal region of CYP4C21.

The pseudogenes CHBAY and WSH were named *CYP4C21P2* (Accession number: AF284452) and *CYP4C21P3* (Accession number: AF281327), respectively (Fig. 3). The nucleotide sequences of these two pseudogenes are 85.1% identical in the BAYC6S1/BAYC6A1 region. It is noteworthy to point out that the sequences of the first 317 bp nucleotides in the 3′ region of *CYP4C21P2* (CHBAY) and *CYP4C21P3* (WSH) are highly identical to the sequences in the corresponding regions of *CYP4C21* (position 1189–1505 in *CYP4C21*) and *CYP4C21P1* (Fig. 3). Relative to the nucleotide sequences of *CYP4C21*, those of *CYP4C21P1*, *CYP4C21P2* and *CYP4C21P3* in this region are 97.2, 98.4 and 99.4% identical, respectively. When translated in the proper reading frame, the 317 bp regions of *CYP4C21P1*, *CYP4C21P2* and *CYP4C21P3* encode peptide sequences that are 97.2, 97.2 and 98.1% identical to that of CYP4C21, respectively. However, sequences upstream of the 317 bp regions in both

Fig. 2. Comparison of *CYP9E2* (AF275640) and its related pseudogenes. (A) Alignment of *CYP9E2*, *CYP9E2P1* (AF276434) and *CYP9E2P2* (AF276435) in the region between primers MCHDS4 and MCHDA1 (bold font). Nucleotides are underlined where there are differences. Gaps in the alignment are indicated by –. (B) Schematic comparison of *CYP9E2* and its related pseudogenes. Nucleotide positions are marked on the ruler and are based on *CYP9E2*. Gaps in the alignment are indicated by x or ///////.

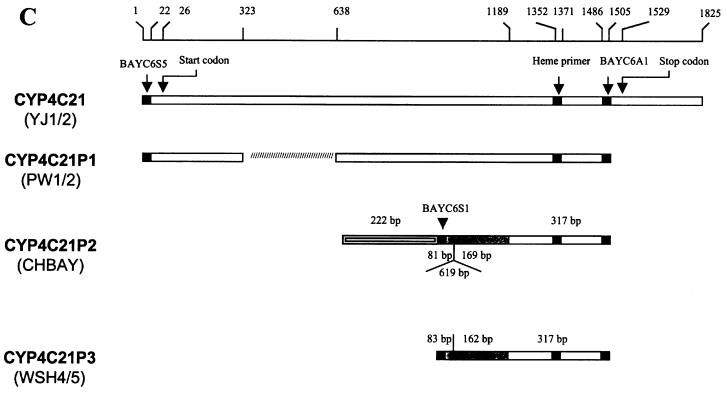
A

```
4C21     AGTAACGAACCGAGAGCTACCGAAAATGGACATAACAGTTTTTCTGTCCCTAACCGTTGTCGTGTTTCTAGCGGTTATCATTTTTTTATGGACGAAACGAA   100
4C21P1   AGTAACGAACCGAGAGCTACCGAAAATGGACATAACAATTTTTCTGTCCCTAACCGTTGTCGTATTTCTAGCGGTTATCATTTTTTTATGGACGAAACGAA   100

4C21     GAGCGAAAGCGCATGCTCCAGAAAATACCGGGTGCCCCCAAGTCTTCCTCTAATTGGAACGGCACTTCCGATACTTTACAGGCGGAAAGAAGATAGGATGA   200
4C21P1   GAGCGAAAGCGCATGCTCCAGAAAATACCGGGTCCCCCAAGTCTTCCTCTAATTGGAACGGCACTTCCGATACTTTACAGGCGGAAAGAAGATAGGATGA   200

4C21     CATGGATAGAAGAAATTCTGGAGAAATATAAACCGCTAATACTTTGGTACTTTGGAAATCGTCCGTTTGTCAATATATCTTCACCTGAGCTTATCGAGGT   300
4C21P1   TATGGATAGAAGAAATTCTGGAGAAATATAAACCGCTAATACTTTGGTACTTTGGAAATCGTCCGTTTGTCAATATATCTTCACCTGAGCTTATCGAGGT   300

4C21     TGTCTTGAGAAACACACAGCTAATTGACAAAGCATTCTTATACGATTTATTTCACTCGTGGCTTGGTACTGGGCTTCTGACTTCTTCAGGTGCTAAATGG   400
4C21P1   TGTCTTGAGAAACACACAGCTA----------                                                                    322

4C21     CATCAACACAGGAAGATCATCACTCCGACGTTCCACTTCAGCATCTTGGAGGGTTTCATCACAATCTTCGCCGAGAAATCTGAGATACTCGTGCGAAAAT   500
4C21P1   ----------                                                                                         322

4C21     TGCAGAAAGAAGTTGGAAGAGGGCCCTTCTTCATACGACAATATGTTTCAAACTGTGCCCTAGACATCATTTGTGAGACGGCAATGGGAACATCAGTAAA   600
4C21P1   ----------                                                                                         322

4C21     TGCACAAGATGAAGGATTTTCAGAATATGTAACAGCTATCAACAAAATGACGGATGTCCTCTCTGATCGCATGGCCAATCCACTCTTGTATCCAGAATTC   700
4C21P1   ----------------------------------------TCAACAAAATGACGGATGTCCTCTCTGATCGCATGGCCAATCCACTCTTGTATCCTGAATTC   384

4C21     ATATTCAAGCTCACTCCTTATTACTGGACACATAAGAAGTGCCTGAAAGTACTAAATGGCTTTGTTAATAAGATAATACAGGAACGCAAAGAAGAAAGGA   800
4C21P1   ATATTCAAGCTCACTCCTTATTACTGGACACATAAGAAATGCCTGAAAGTACTAAATGGCTTTGTTAATAAGATAATACAGGAACGCAAAGAAGAAAGGA   484

4C21     AAAAGTCTAAAGTCACACAGACAAGTGAGGATGCAGATATTGGAAAAAAGAAAAGGGTTCCTTTTTTGGACACATTATTGGATGCCAGTGAAGACGATAA   900
4C21P1   AAAAGTCTAAAGTCAAACAGAGAAGTGAGGATGCAGATATTGGTAAAAAGAAAAGGGTTCCTTTTTTGGACACATTATTGGATGCCAGAGAAGACGATAA   584

4C21     CAAGTTGACAGACACTGACATTTTGGAAGAAGTTCACACCTTCATGTTTGAGGGGCATGACACAGTATCTGCTGCCATGACCTGGTTGCTGTTTGAGCTG   1000
4C21P1   CAAGTTGACAGACACTGACATTTTGGAAGAAGTTCACACCTTCATGTTTGAGGGGCATGACACAGTATCTGCTGCCATGACCTGGTTGCTGTTTGAGCTG   684

4C21     GGCCATCATCCAGAGATTCAGGAAGAAGCCTATAAGGAAGTGCAAGATATATTTCAGGGATCAGACCGTGTTCCAACAATGGCTGACTTAAACAATATGA   1100
4C21P1   GGCCATCATCCAAAGATTCAGGAAGAAGTCTATAAGGAAGTGCAAGATATATTTCAGGGATCAGACCGTGTTCCAACAATGGCTGACTTAAACAATATGA   784

4C21     ATTATCTGGAACGAGTTATCAAAGAAAGTCTGAGACTGCATCCAAGTGTTATTTATTTCGTAAGAGAGGCCCATCAAGATTTTGAACTAGGCGGTTACAC   1200
4C21P1   ATTATCTGGAACGAGTTATCAAAGAAAGTCTGAGACTTCATCCAAGTGTTATTTATTTCGTAAGAGAGGCCCATCAAGATTTTGAACTAGGCGGTTACAC   884

4C21     CATTCCAGCTGGAACGAATATTGATTTCTCGGTGCCATTCATCCATAGGAACCCGGAAATATTCCCCAACCCTCGATGCTTCAACCCTGACAATTTCTTA   1300
4C21P1   CATTCCAGCTGGAACGAACATTGGTTTCTCGGTGCCATTCATCCATAGGAACCCGGAAATCTTCCCCAACCCTCGATGCTTCAATCCTGACAATTTCTTA   984

4C21     CCAGACAGGGTTGTTAACAGACACCCCTATGCGTACATCCCTTTCAGCGCTGGACCACGGAATTGTATTGGACAACGCTTTGCTCTCCTTGAAGAGAAAG   1400
4C21P1   CCAGAGAGGGTTGTTAACAGACACCCCTATGCGTACATCCCTTTCAGCGCTGGACCACGGAATTGTATTGGACAACGCTTTGCTCTCCTTGAAGAGAAAG   1084

4C21     TTGTCTTGTCCTATCTCTTACGTCACTACAGATTTAGGACAGTCAATAAAAGGGAGGATTCTAAATTCAAGTTGGAAATGATAAACACCCCTGTCAAGCC   1500
4C21P1   TTGTCTTGTCATACTCTTACGCCACTATAGATTTAGGACAGTCAATAAAAGGGAGGATTCTAAATTCAAGTTGGAAATGATAAACACCCCTGTCAAGCC   1184

4C21     TATTC                                                                                              1505
4C21P1   TATTC                                                                                              1189
```

B

```
4C21P2                                                                                    TTTTCAAAACATACAAACTGCT     22
         TGTCAATTAGCAAAGAATTTTTTTTTACAAATATATCTAATCACTTCTAATTTAAAGAGCAAATAATCCTACGCCAACCAACAATAGTTCAGAGCCCTGC    122
         CGCAATTAGTTACTCTTTTATAGTTATTACAATAATTACTGTAAATGGAGGAAATTTTTTATTTATTTGAAAGAAAGAAAATTCTGTAGAAATGCTAAA    222

4C21P2   AGTAACAATCTAAATGCCAAACTTGCGCTATATAT--TTGTCTTTACCTATTATTAATAATGATTCATAATAATATATTACGACATGAGGTTGGGAAGTG   320
4C21P3   AGTAACAATCTAAATGCCAAACTTGCGCTATATATATTTGTCTTTACCTACTATTAATCATGATTCATAAAAGTAGTTTGTAT----------------    83
                                                                                                  CYP4C21P3 diverges from P2

4C21P2   GTGTGTACTGAATCGAGGAGACGATTATAAAAACGAAGGAAGTGAGAGATTCTGTAATACACTTCACAAACATGTATTGTACAGTATTTTTTGCACGATA   420
         TATGTTTTGAAAATAATATGTTTAAAAAGTTTACCTCCAGAACATATGTCACTCTACGAACAGCTGACTGATTTATGGTAGCCATCTTTATTAATATCGA   520
         TTATGCTAGCAGTATGGACAACATAATTTTTCAAGTAATTTACTTGAAATGCCAACAAACTGGTATAAGATGGGGGCAATTATGAATCGTTATTGTGAAA   620
         ATTATGGAATTGAAGAAGTGTGGGAGTCCGTGTGTCCTTCTTTAGTACCAGAAAGGTCCAAACTTAGATAAAGCGAAGTTAATTTGTTTTAAATAATATG   720
         AATTTCCCTCATGGTTACAGGAAACATCCAATACATAGAAGGCTGTGGTTTTACTTATAACCCTAGAATTTACAATTAGTAATATTGTAACATACTAATC   820
         TATGTTGGCAACTCTGAATTCAGTAAAATCATCTTCCAATCATGGTGACCTTTCCCTGTAGTCATGAGAAAACACACTATCGTGCAAAAAGTAGTTTTC   920
                                                                                                  P450 reading frame for P2
4C21P2   TTAATATATTAGGGACAATAGGA--TTTTTTTTAA-TTATAAGGTGTCCATAAATTATCTGGCAACACTGTAGTTCACAGAAATATTAAAATTAAATGTT   1017
4C21P3   --AATATATTAGGGACAATAGGAATTTTTTTTTAAATTATAAGGTGTCCATAAATTATCTGGTAACACTGTAGTTCACAGAAATATTAAAATTAAATGTT   181
         P3 diverges from P2                          4C21 and P1 diverge from P2 and P3
4C21                                                                                      AGGCGGTTACACCATTCCAGCTGGAA    26
4C21P1                                                                                    AGGCGGTTACACCATTCCAGCTGGAA    26
4C21P2   GTTGAGACAATTTTTTTTACTGTCGGCTTTCCACGTGTCAATCATTTTAATCACTTGTTTCTGTGCTCTATTCCAGGCGGTTACACCATTCCAGCTGGAA   1117
4C21P3   GTTGAGACAATTTTTT-TACTGTCGGCTTTCCAC---------ATTTTACTCACTTGTTTCTGTACTCTATTCCAGGCGGTTACACCATTCCAGCTGGAA    271
                  P450 reading frame for P3

4C21     CGAATATTGATTTCTCGGTGCCATTCATCCATAGGAACCCGGAAATATTCCCCAACCCTCGATGCTTCAACCCTGACAATTTCTTACCAGACAGGGTTGT   126
4C21P1   CGAACATTGGTTTCTCGGTGCCATTCATCCATAGGAACCCGGAAATCTTCCCCAACCCTCGATGCTTCAATCCTGACAATTTCTTACCAGAGAGGGTTGT   126
4C21P2   CGAACATTGGTTTCTCGGTGCCATTCATCCATAGGAACCCGGAAATCTTCCCCAACCCTCGATGCTTCAATCCTGACAATTTCTTACCAGAGAGGGTTGT  1217
4C21P3   CGAACATTGATTTCTCGGTGCCATTCATCCATAGGAACCCGGAAATATTCCCCAACCCTCGATGCTTCAACCCTGACAATTTCTTACCAGACAGGGTTGT   371

4C21     TAACAGACACCCCTATGCGTACATCCCTTTCAGCGCTGGACCACGGAATTGTATTGGACAACGCTTTGCTCTCCTTGAAGAGAAAGTTGTCTTGTCCTAT   226
4C21P1   TAACAGACACCCCTATGCGTACATCCCTTTCAGCGCTGGACCACGGAATTGTATTGGACAACGCTTTGCTCTCCTTGAAGAGAAAGTTGTCTTGTCATAC   226
4C21P2   TAACAGACACCCCTATGCGTACATCCCTTTCAGCGCTGGACCACGGAATTGTATTGGACAACGCTTTGCTCTCCTTGAAGAGAAAGTTGTCTTGTCCTAT  1317
4C21P3   TAACAGACACCCCTATGCGTACATCCCTTTCAGCGCTGGACCACGGAATTGTATTGGACAACGCTTTGCTCTCCTTGAAGAGAAAGTTGTCTTGTCCTAT   471

4C21     CTCTTACGTCACTACAGATTTAGGACAGTCAATAAAAGGGAGGATTCTAAATTCAAGTTGGAAATGATAAACACCCCTGTCAAGCCTATTC            317
4C21P1   CTCTTACGCCACTATAGATTTAGGACAGTCAATAAAAGGGAGGATTCTAAATTCAAGTTGGAAATGATAAACACCCCTGTCAAGCCTATTC            317
4C21P2   CTCTTACGTCACTACAGATTTAGGACAGTCAATAAAAGGGAGGATTCTAAATTCAAGTTGGAAATGATAAACACCCCTGTCAAGCCTATTC           1408
4C21P3   CTCTTACGTCACTACAGACTTAGGACAGTCAATAAAAGGGAGGATTCTAAATTCAAGTTGGAAATGATAAACACCCCTGTCAAGCCTATTC            562
```

C



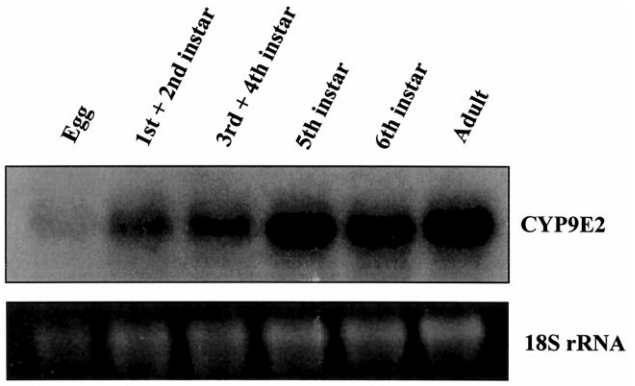


Fig. 4. Expression of 'CYP9E2' at different life stages. Total RNA was prepared from ~1 g mixed sexes for all the life stages and 10 μg of total RNA was loaded in each lane. Northern hybridization with the CYP9E2 cDNA probe is shown in the top panel. RNA loading was standardized by ethidium bromide staining of 18S rRNA (bottom).

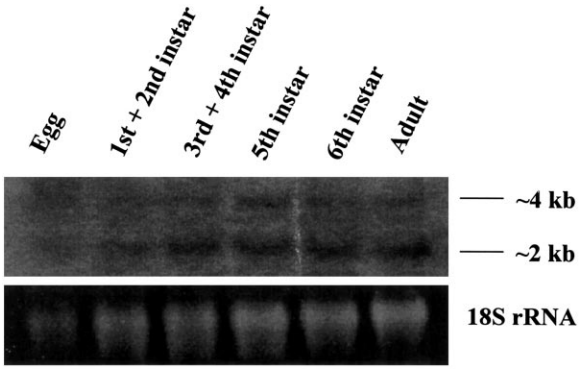


Fig. 5. Expression of 'CYP4C21' at different life stages. Total RNA was prepared from ~1 g mixed sexes for all the life stages and 10 μg of total RNA was loaded in each lane. Northern hybridization with the CYP4C21 cDNA probe is shown in the top panel. RNA loading was standardized by ethidium bromide staining of 18S rRNA (bottom).

CYP4C21P2 and *CYP4C21P3* are highly different from that of any part of *CYP4C21* (sequences were 22.6% identical for *CYP4C21P2* and 27.3% identical for *CYP4C21P3*). The overall sequences of *CYP4C21* with *CYP4C21P2* and *CYP4C21P3* are 38.4 and 66.9% identical, respectively. *CYP4C21P2* and *CYP4C21P3* are pseudogenes because there are many stop codons (18 for *CYP4C21P2* and six for *CYP4C21P3*) in the upstream regions when translated in their P450 reading frames (Fig. 3B). The overall amino acid sequence comparison of CYP4C21 (501 residues), CYP4C21P2 (139 residues) and CYP4C21P3 (125 residues) shows that CYP4C21 is 77.0 and 84.0% identical to CYP4C21P2 and CYP4C21P3, respectively, while CYP4C21P2 and CYP4C21P3 are 92% identical.

3.4. Northern blot analyses of CYP9E2 and CYP4C21

When the CYP9E2 probe was used, a single band (~2 kb) was detected in all the life stages of German cockroaches by Northern blot analyses (Fig. 4). The signal was faint in eggs, became stronger in first + second and third + fourth instar nymphs, and was the strongest in fifth and sixth instar nymphs and in adults. CYP9E2 expression was not restricted to particular body parts nor was it sex-specific. Indeed, it was detected in both the abdomens and the remaining parts of bodies of both male and female adults (data not shown).

When the CYP4C21 probe was used, two bands (~4 and ~2 kb) were detected in all the life stages (Fig. 5). The signals were weak in eggs, but of similar intensity at all the other life stages.

4. Discussion

In this paper, we report the successful isolation of two novel P450s, *CYP9E2* and *CYP4C21*, from German cockroaches. Both CYP9E2 and CYP4C21 are typical microsomal P450s having characteristics shared by other members in the P450 superfamily. mRNA signals were detected throughout all life stages by Northern blot analysis using either the CYP9E2 or CYP4C21 probe. These signals may reflect the expression of related mRNAs including pseudogenes (see below).

Multiple pseudogenes closely related to a single paralogous gene are not uncommon (Mighell et al., 2000). However, this is the first report of multiple pseudogenes closely related to functional P450s. Comparisons of the pseudogenes reported here with those from *D. melanogaster* (Adams et al., 2000) yield several interesting points. For example, the deletion of a large fragment in *CYP9E1* and *CYP4C21* pseudogenes was also found in *Cyp9f3p* from *D. melanogaster*, which was missing a fragment of 93 nucleotides that encoded a peptide of 31 amino acids in the heme binding region. Otherwise the amino acid sequences of *Cyp9f3p* and *9f2* shared a high degree of nucleotide identity (94.5%) (Adams et al., 2000; Nelson, 2000). Mosaic sequences (partial P450/partial non-P450), like those of *CYP4C21P2* and *P3*, were also found in *Cyp6a15p* from *D. melanogaster* (Adams et al., 2000; Nelson, 2000).

The genome sequence of *D. melanogaster* revealed 86 putative P450 genes and four pseudogenes (Adams et al., 2000). In contrast, there are only five P450 genes sequenced thus far from *B. germanica*, and there are already five pseu-

Fig. 3. Comparison of *CYP4C21* (AF275641) and three related pseudogenes. (A) Alignment of *CYP4C21* and *CYP4C21P1* (AF281326) in the region between primers BAYC6S5 and BAYC6A1 (bold font). (B) Alignment of *CYP4C21P2* (AF284452) and *CYP4C21P3* (AF281327). Primers (BAYC6S1, heme primer and BAYC6A1) are in bold font. The 317 bp sequences of *CYP4C21* and *CYP4C21P1* are also included for comparison. (C) Schematic comparison of *CYP4C21* and its related pseudogenes. Nucleotide positions are marked on the ruler and are based on *CYP4C21*. Regions with similar shading patterns are highly similar in sequence. Nucleotides are underlined where there are differences. Primers are in bold font. Gaps in the alignment are indicated by – or ///////.

dogenes known. Thus, the relative number of pseudogenes in *B. germanica* appears to be much higher than in *D. melanogaster*. Our results support the idea that *D. melanogaster* has a high rate of spontaneous large deletions leading to few pseudogenes, relative to other insects (Petrov and Hartl, 1998).

Nucleotide substitution in regions no longer subject to selective constraints (such as pseudogenes) undergo non-random changes that are generally biologically, but not physico-chemically neutral (Hess et al., 1994). As a result, pseudogenes have a decrease in GC content over time. This trend holds for *CYP4C21P1* (relative to *CYP4C21*), supporting our hypothesis that *CYP4C21P1* is a real pseudogene. The overlapping sequence of *CYP4C21* with *CYP4C21P2* and *CYP4C21P3* was too short (thus there were too few differences) to allow for a meaningful comparison of GC composition. Comparison of *CYP9E2* with *CYP9E2P1* and *CYP9E2P2* showed a slight overall decrease in GC content which is consistent (although not compellingly so) with the idea that these are true pseudogenes. Furthermore, the rate of deletions is greater than the rate of insertions, such that pseudogenes slowly shrink in size (Graur et al., 1989). This is consistent with our results for all pseudogenes.

Although pseudogenes have no immediate biological function (Wilde, 1986), their sequence similarity with paralogous functional genes could provide materials for the generation of new functional P450s by internal DNA turnover events such as genomic rearrangement and gene conversion (Mighell et al., 2000) as demonstrated by the somatic immunoglobulin diversity in chickens (McCormack et al., 1993). If such internal DNA turnover events do occur, we should expect to find genes that have mosaic sequences. *CYP4C21P2* and *CYP4C21P3* are sequences that fit this expectation in that the $3'$ sequences are highly identical to the corresponding region of *CYP4C21* yet their $5'$ sequences are totally unrelated. *Cyp6a15p* is also a mosaic sequence in that its $3'$ sequence is most similar to *Cyp6a13* and *Cyp6a14* while its $5'$ sequence is totally unrelated to these P450s (Adams et al., 2000; Nelson, 2000). *Cyp6a13*, *Cyp6a14* and *Cyp6a15p* are clustered in the 44D1 region of chromosome 2 (Adams et al., 2000). These results support the idea that new P450s might be formed by internal DNA turnover events (i.e. molecular drive; Dover and Tautz, 1986; Gonzalez and Nebert, 1990; Danielson et al., 1997). These events may lead to the generation of new P450s or P450 pseudogenes depending on whether the mosaic sequences are in the right reading frame or not. The deletion of large fragments in *CYP9E2P1*, *CYP4C21P1* and *Cyp9f3p* compared with their corresponding paralogous functional genes was probably also the result of such events. The high level of nucleotide identities between these pseudogenes and their paralogous genes suggests a recent occurrence of such events.

P450 sequences (~1.6 kb) with overall nucleotide identity as high as 73.3% cross-react with each other at low stringency, but do not cross-react at high stringency (Kasai and Scott, 2001). However, genes with an overall nucleotide identity of 88.4% cross-reacted with each other even at high stringency as found for *CYP6B1* and *CYP6B3* (Cohen et al., 1992; Hung et al., 1995). Due to the high sequence similarity between a gene and its related genes or pseudogenes, it is very likely that the expression of some P450 genes determined by Northern blot may have been overestimated due to detection of the mRNAs for the gene, related genes and/or pseudogenes. The chances of this happening increase when short pieces of cDNA are used as probes. For example, if the 317 bp region of *CYP4C21* was used as probe, *CYP4C21*, *CYP4C21P1*, *CYP4C21P2* and *CYP4C21P3* would all be detected. However, if the full-length cDNA sequence of CYP4C21 was used as probe under high stringency, *CYP4C21* and *CYP4C21P1* would probably be the only two species to be detected. Due to the existence of multiple pseudogenes highly similar to *CYP9E2* or *CYP4C21*, the mRNA signals we detected may reflect the expression of related mRNAs including pseudogenes. For example, the signals detected by the CYP9E2 probe are likely the mix of *CYP9E2* and *CYP9E2P2*. Given the cDNA length difference yet high level nucleotide similarity between *CYP9E2* and *CYP9E2P1*, the single band detected with the CYP9E2 probe suggests that *CYP9E2P1* was expressed at a very low level. Two bands (~4 and ~2 kb) were detected by Northern blots using the CYP4C21 probe. *CYP4C21P2* and *CYP4C21P3* do not likely contribute to the signals detected because of their low level of overall nucleotide identity with *CYP4C21*. The ~2 kb signal may be the CYP4C21 mRNA considering its poly(A) tail and possible additional $5'$ sequence. We cannot explain the 4 kb signal detected with the CYP4C21 probe. It might be the unspliced mRNA precursor. In another case, a 4.2 kb mRNA signal was detected using either CYP9A2 or CYP9A4 cDNA as probe in Northern blot analysis of P450 induction in *Manduca sexta* (Stevens et al., 2000). Like *CYP9E2P1*, *CYP4C21P1* might be expressed at low levels so that it could not be detected. In conclusion, other methods such as RT-PCR using gene-specific primers need to be employed to reinvestigate the mRNA expression of CYP9E2 and CYP4C21. In light of the above discussion, we suggest that all the P450 expression patterns examined by Northern blot using partial cDNAs as probes be reinvestigated once the full or near full-length cDNA sequences are available. If possible, the results of previous Northern blot analyses should be confirmed by other methods when multiple bands were observed or when short cDNAs were used as probes.

An important implication from the isolation of pseudogenes involves the current P450 nomenclature system. Analysis of the P450 nomenclature web site (Nelson, 2000) showed that at least 50 out of the approximately 300 named insect P450s were named based on the sequence of a fragment of deduced protein shorter than

150 amino acid residues. Some P450s have even been named based on the sequence of a fragment as short as 54 amino acid residues, which extends from the heme binding region to the C-terminus. The isolation of pseudogenes similar at the nucleotide level to their paralogous P450 genes reported herein casts doubt on the validity of assigning names based on deduced amino acids from short cDNA sequences. It is possible that some of the short sequences that have been named will later prove to be pseudogenes. For example, the 5′ sequences of *CYP4C21P2* and *CYP4C21P3* (Fig. 3B) have no similarity at all. However, the 3′ sequences of *CYP4C21P2* and *CYP4C21P3* (Fig. 3B) encode normal P450s (139 residues for CYP4C21P2 and 125 residues for CYP4C21P3) that are 77.0 and 84.0% identical to *CYP4C21*. The identities are even higher when only amino acids from the 317 bp region are compared. Similar situations exist in P450 pseudogenes identified from *D. melanogaster*. For example, the downstream region of *Cyp6a15p* encodes for a normal-looking P450 (237 amino acids) which is 48.8 and 40.0% identical with corresponding regions of CYP6A13 and CYP6A14, respectively, yet its upstream sequence is totally unrelated to any P450 sequences (nucleotide sequences are ~20.0% identical with other P450s) (Adams et al., 2000; Nelson, 2000). If the amino acid sequences of P450-coding regions of the CYP4C21P2 and CYP4C21P3 were submitted for nomenclature, it is very likely that the current P450 nomenclature system would assign names of functional genes to them. If the total known sequences of these two pseudogenes were submitted, they may well be assigned independent names as was *Cyp6a15p*. Is it justified to assign independent names to pseudogenes? One characteristic of pseudogenes is their 'close similarities to one or more paralogous genes' (Mighell et al., 2000). Thus, P450 pseudogenes should be named according to their most related functional gene whenever possible, rather than be given an independent name.

In some cases, P450 cDNAs, such as CYP9A1 (Rose et al., 1997), CYP6B2 (Xiao-Ping and Hobbs, 1995) and CYP6H1 (Winter et al., 1999), were the result of ligating two independent (i.e. partial sequences) cDNA clones. In light of our results, it is possible that one piece might be part of a pseudogene, or the two pieces might not even belong to the same gene. This possibility greatly increases when there are nucleotide differences within the overlapping regions as in the case of *CYP6B2* (Xiao-Ping and Hobbs, 1995). Thus, it may be worthwhile to obtain full-length cDNAs to confirm the sequence for P450s whose sequences resulted from two independent partial clones.

Acknowledgements

We thank Drs P. Danielson and J. Fogleman for suggestions with the 3′ and 5′ RACE techniques, Drs C. Gilbert, J. Helmann, J. Ewer, B. Danforth and R. MacIntyre for valuable discussions and Dr R. MacIntyre for helpful comments on an earlier draft of this manuscript. We thank Dr D. Nelson for naming the P450 and Dr M. Chapman for the cDNA library. We also thank J. Rice for rearing the cockroaches. This project was supported by a grant from Sigma Xi (to Z.W.) and the Cornell Hughes Scholars Program with funding from the Howard Hughes Medical Institute (to C.E.H.).

References

Adams, M.D., Celniker, S.E., Holt, R.A., et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Arruda, L., Vailes, L., Hayden, M., Benjamin, D., 1995. Cloning of cockroach allergen, Bla g 4, identifies ligand binding proteins (or calycins) as a cause of IgE antibody responses. J. Biol. Chem. 270, 31196–31201.

Bradfield, J., Lee, Y., Keeley, L., 1991. Cytochrome P450 family 4 in a cockroach: molecular cloning and regulation by hypertrehalosemic hormone. Proc. Natl. Acad. Sci. USA 88, 4558–4562.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J., 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294–5299.

Cohen, M.B., Schuler, M.A., 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.

Danielson, P.B., Fogleman, J.C., 1997. Isolation and sequence analysis of cytochrome P450 12B1: the first mitochondrial insect P450 with homology to 1alpha,25 dihydroxy-D_3 24 hydroxylase. Insect Biochem. Mol. Biol. 27, 595–604.

Danielson, P.B., MacIntyre, R.J., Fogleman, J.C., 1997. Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome P450s: evidence for involvement in host-plant allelochemical resistance. Proc. Natl. Acad. Sci. USA 94, 10797–10802.

Dover, G.A., Tautz, D., 1986. Conservation and divergence in multigene families: alternatives to selection and drift. Philos. Trans. R. Soc. Lond. B 312, 275–289.

Fogleman, J.C., Danielson, P.B., MacIntyre, R.J., 1997. The molecular basis of adaptation in *Drosophila*. The role of cytochrome P450s. Evol. Biol. 30, 15–77.

Frohman, M.A., Dush, M.K., Martin, G.R., 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85, 8998–9002.

Gonzalez, F.J., 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: Omura, T., Ishimura, Y., Fujii-Kuriyama, Y. (Eds.), Cytochrome P-450. VCH, New York, pp. 255–272.

Graur, D., Shuali, Y., Li, W.-H., 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. J. Mol. Evol. 28, 279–285.

Gussow, D., Clackson, T., 1989. Direct clone characterization from plaques and colonies by the polymerase chain reaction. Nucleic Acids Res. 17, 4000.

Hess, S.T., Blake, J.D., Blake, R.D., 1994. Wide variations in neighbor-dependent substitution rates. J. Mol. Biol. 236, 1022–1033.

Hung, C.F., Harrison, T.L., Berenbaum, M.R., Schuler, M.A., 1995. CYP6B3: a second furanocoumarin-inducible cytochrome P450 expressed in *Papilio polyxenes*. Insect Mol. Biol. 4, 149–160.

Kasai, S., Scott, J.G., 2001. Expression and regulation of CYP6D3 in the house fly, *Musca domestica* (L). Insect Biochem. Mol. Biol. (in press).

Mansuy, D., 1998. The great diversity of reactions catalyzed by cytochromes P450. Comp. Biochem. Physiol. 121C, 5–14.

McCormack, W.T., Hurley, E.A., Thompson, C.B., 1993. Germ line maintenance of the pseudogene donor pool for somatic immunoglobulin gene conversion in chickens. Mol. Cell. Biol. 13, 821–830.

Mighell, A.J., Smith, N.R., Robinson, P.A., Markham, A.F., 2000. Vertebrate pseudogenes. FEBS Lett. 468, 109–114.

Nebert, D.W., Gonzalez, F.J., 1987. P450 genes: structure, evolution and regulation. Annu. Rev. Biochem. 56, 945–993.

Nelson, D.R., 1998. Cytochrome P450 nomenclature. In: Phillips, I.R., Shepard, E.A. (Eds.), Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 15–24.

Nelson, D.R., 2000. http://drnelson.utmem.edu/nelsonhomepage.html

Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W., Gunsalus, I.C., Nebert, D.W., 1996. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics 6, 1–42.

Petrov, D.A., Hartl, D.L., 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol. Biol. Evol. 15, 293–302.

Rose, R.L., Goh, D., Thompson, D.M., Verma, K.D., Heckel, D.G., Gahan, L.J., Roe, R.M., Hodgson, E., 1997. Cytochrome P450 *(CYP)9A1* in *Heliothis virescens*: the first member of a new *CYP* family. Insect Biochem. Mol. Biol. 27, 605–615.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Savonet, V., Maenhaut, C., Miot, F., Pirson, I., 1997. Pitfalls in the use of several 'housekeeping' genes as standards for quantitation of RNA: the example of thyroid cells. Anal. Biochem. 247, 165–167.

Siegfried, B.D., Scott, J.G., 1991. Mechanisms responsible for propoxur resistance in the German cockroach, *Blattella germanica* (L). Pestic. Sci. 33, 133–146.

Spiess, A.-N., Ivell, R., 1998. Normalization of RNA hybridization signals by means of SYBR Green II-stained 28S or 18S ribosomal RNA and phosphor imager. BioTechniques 26, 46–50.

Stevens, J.L., Snyder, M.J., Koener, J.F., Feyereisen, R., 2000. Inducible P450s of the CYP9 family from larval *Manduca sexta* midgut. Insect Biochem. Mol. Biol. 30, 559–568.

Sutherland, T.D., Unnithan, G.C., Andersen, J.F., Evans, P.H., Murataliev, M.B., Szabo, L.Z., Mash, E.A., Bowers, W.S., Feyereisen, R., 1998. A cytochrome P450 terpenoid hydroxylase linked to the suppression of insect juvenile hormone synthesis. Proc. Natl. Acad. Sci. USA 95, 12884–12889.

Tindall, K.R., Kunkel, T.A., 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 27, 6008–6013.

Wen, Z., Scott, J.G., 2001. Cytochrome P450 CYP6L1 is specifically expressed in the reproductive tissues of adult male German cockroaches, *Blattella germanica* (L.). Insect Biochem. Mol. Biol. 31, 179–187.

Wilde, C.D., 1986. Pseudogenes. Crit. Rev. Biochem. 19, 323–352.

Winter, J., Bilbe, G., Richener, H., Sehringer, B., Kayser, H., 1999. Cloning of a cDNA encoding a novel cytochrome P450 from the insect *Locusta migratoria*: CYP6H1, a putative ecdysone 20-hydroxylase. Biochem. Biophys. Res. Commun. 259, 305–310.

Xiao-Ping, W., Hobbs, A.A., 1995. Isolation and sequence analysis of a cDNA clone for a pyrethroid inducible cytochrome P450 from *Helicoverpa armigera*. Insect Biochem. Mol. Biol. 25, 1001–1009.