DNA Sequence, June 2004 Vol. 15 (3), pp. 228–233

Short Communication

Cloning of *CYP9G2* from the Diamondback Moth, *Plutella xylostella* (Lepidoptera: Yponomeutidae)

BENCHANG SHEN[a], DEXIU ZHAO[b], CHUANLING QIAO[a,*] and WENSHENG LAN[a]

[a]*State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, No. 25, Bei Si Huan Xi Lu, Beijing 100080, People's Republic of China;* [b]*Institute of Botany, Chinese Academy of Sciences, Beijing 100093, People's Republic of China*

(Received 8 January 2004)

Cytochrome P450s constitute a superfamily of hemoproteins, important in the metabolism of endogenous and xenobiotic compounds. The full-length cDNA of a novel cytochrome P450, *CYP9G2*, was isolated from a cDNA library. The cDNA is 2143 bp in length and contains an open reading frame from 50 to 1615 bp, encoding a protein of 521 amino acid residues. The putative P450 protein contains a highly hydrophobic N terminus and a P450 protein signature motif, FG/S*G*R*C*G*A/G, known as the important ligand for heme binding, analysis of the NH_2-terminal sequence indicated that CYP9G2 is a microsomal P450. Using polymerase chain reaction with primers specific to *CYP9G2*, the genomic structure of *CYP9G2* was analyzed, and it was found that the gene contains seven introns and eight exons within the coding region, all the sequences of the exon–intron junctions are consistent with the AG–GT rule. Multiple alignment indicated that CYP9G2 is most similar to CYP9E2 from the *Blattella germanica* (42.7% identity), it is also similar to the insect P450s in family 9, including CYP9L1 from *Anopheles gambiae* (38.7%) and CYP9A1 from *Heliothis virescens* (39.5%).**

Keywords: *CYP9G2*; cDNA; Intron; Exon; Genomic structure

Database Accession Number: AB096739 and AB112959 are accession numbers in GenBank for cDNA and genomic DNA of *CYP9G2*, respectively

INTRODUCTION

Cytochrome P450 monooxygenases or CYPs (EC 1.14.14.1) are a superfamily of heme-containing proteins that are responsible for the oxidative metabolism of a wide variety of both xenobiotic and endogenous compounds, in insects, cytochrome P450s metabolize hormones and pheromones but are best known for their role in the metabolism of insecticides and host plant chemicals (Feyereisen, 1999; Scott and Wen, 2001). P450s act as terminal oxidase in multi-component electron transfer chains. Members of the cytochrome P450 family are found in virtually all aerobic organisms, including organisms as diverse as insects, plants, mammals, birds and bacteria (Scott, 1999; Ranson *et al.*, 2002a). The total number of cytochrome P450 was near 2500 by the end of 2002 (http://drnelson.utmem.edu/CytochromeP450.html). In insects, more than 300 P450s have been identified and are distributed throughout 27 *CYP* families of more than 70 known P450 gene superfamilies where 90 P450 genes scattered in 25 families are present in *Drosophila melanogaster* genome (Tijet *et al.*, 2001; Liu and Zhang, 2002), in *Anopheles gambiae*, 111 cytochrome P450s cDNAs were identified by a BLAST search of the genome of *A. gambiae* (Ranson *et al.*, 2002b).

Here, we report the cDNA sequence, genomic characterization of a cytochrome P450 gene, *CYP9G2*, in the diamondback moth, *Plutella xylostella*, which is the most destructive insect of cruciferous plants and resistant to almost every insecticide.

*Corresponding author. Tel./Fax: +86-10-62553369. E-mail: qiaocl@panda.ioz.ac.cn

ISSN 1042-5179 print/ISSN 1029-2365 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/10425170410001716055

MATERIALS AND METHODS

Insect

P. xylostella were obtained from a colony in southern China and maintained at the laboratory under constant chemical pressure. Each generation has been exposed to permethrin selection. Adult leaf-dip bioassay indicated that the insect exhibits a moderate level of resistance of 30.64-fold to permethrin relative to the susceptible population of our laboratory.

Reverse Transcription Polymerase Chain Reaction

Total RNA was extracted from five whole bodies of diamondback moth using TRIzol reagent as described by the manufacturer (Life Technologies/Gibco-BRL, MD, USA). The total RNA was used for RT-PCR, construction of the cDNA library and primer extension analysis. The first strand cDNA was synthesized with SuperScript II reverse transcriptase RNase H^- at 42°C for 30 min (Life Technologies/Gibco-BRL, MD, USA) and an antisense oligo$(dT)_{18}$, using diamondback moth total RNAs as templates. We designed the forward primer as 5′-ATGACITA(TC)(TC)TIGA(TC)(TA)(GT)I-3′ around a conserved amino acid region according to the alignment of CYP9 insect families reported (MTYLDC(M) (Stevens *et al.*, 2000) and the reverse primer 5′-AT(AG)CA(AG)TTIC(GT)IGGICC-3′ according to the P450 heme binding consensus region (GPRNCI) found in rat, human, and insect P450 sequences, respectively (Danielson and Fogleman, 1997; Falckh *et al.*, 1997; Liu and Zhang, 2002). The PCR cycles for the reverse-transcribed cDNA were 95°C, 5 min; 30 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 1 min, and a final extension step of 72°C for 10 min. The RT-PCR products were purified directly from the PCR reaction by bands excised from agarose gels and then cloned into pGEM®-T Easy Vector (Promega Corp., Madison, WI, USA). Positive clones were sequenced on an ABI PRISM™ 377 DNA Sequencer using the Big-Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer).

cDNA Library Construction and Isolation of the *CYP9G2* Positive Clones

A cDNA library was constructed by following the SMART™ (switching mechanism at 5′ end of RNA transcript) polymerase chain reaction cDNA library construction kit protocol (Zhu *et al.*, 2001; Shen *et al.*, 2003). Two hundred nanograms of total RNA were reverse transcribed to cDNA and second strand synthesized based on LD-PCR protocol with introducing of *Sfi* IA (5′-GGCCATTA′TGGCC-3′) and B (5′-GGCCGAGG′CGGCC-3′) sites at the 5′ and 3′ ends of the nascent cDNA, respectively. After *Sfi* digestion, proteinase K treatment and cDNA size fractionation, cDNAs >500 bp were ligated into *Sfi* I A and B-digested, dephosphorylated λ TriplEx-2 arms (Clontech Laboratories, Palo Alto, CA, USA), and packed using the Packagene Lambda DNA Packing System (Promega Corp., Madison, WI, USA). Isolation of the *CYP9G2* cDNA clones was performed using the modified method described by Michael (Lardelli and Lendahl, 1994). Briefly, XL1-Blue cells grown overnight in LB medium containing 0.2% maltose are diluted to $OD_{600} = 0.2$ in the same medium prewarmed to 37°C. Forty milliliters are infected with λ phage cDNA library of 10^6 plaque-forming units (pfu) and, after absorption of the phage for 10 min at 37°C, the culture is rapidly divided into 40 aliquots of 1 ml (containing 25,000 pfu each) that are grown at 37°C with good aeration for 2.5–3 h, when lysis occurs. Three drops of chloroform are then added to complete the lysis. Each 1-ml lysate has a titer of approximately 10^9 pfu, and, therefore, represents a 40,000-fold amplification of the original 25,000 pfu. A forward primer 5′-GCTAATGAGGCACTGAGGAAGTGGT-3′ and a reverse primer 5′-ATGCAATTCCTAGGTCCAGTT-CCAA-3′, which designed according to the information obtained after sequencing of the positive clone of the RT-PCR, were combined to screen the 40 sublibraries, the PCR program was 95°C for 5 min, followed by 30 cycles of 94°C for 30 s, 62°C for 30 s, 72°C for 1 min, and a final extension step of 72°C for 10 min. The sublibraries gave positive signals indicating cDNA clones with the desired sequences and were further subdivided until the positive clones were obtained. The positive clones were subjected to PCR screening using λTriplEx LD-Insert Screening Amplimers, and converted from λTriplEx-2 to pTriplEx-2 by Cre recombinase-mediated site-specific recombination in accordance with the manufacturer's protocol (CLONTECH Laboratories, Palo Alto, CA, USA) and fully sequenced on an ABI PRISM™ 377 DNA Sequencer (Perkin-Elmer). Using the primer walking strategy, the complete sequence was determined and internal primers were designed on the basis of the sequence information obtained. To confirm the fidelity of the sequence information, both strands were analyzed. The sequences were compared with the GenBank database using BLAST (Altschul *et al.*, 1997).

PCR Analysis of the *CYP9G2* Genomic DNA

Genomic DNA was extracted as described (Qiao and Raymond, 1995), PCRs were performed in 30 μl reactions containing 0.1 μg genomic DNA, 0.8 μM of each primer and 1.5 U of *Taq* DNA polymerase (Takara, Japan). The primer pairs for genomic DNA

PCR were designed from cDNA sequence of *CYP9G2*. The forward primer 5′-GAGTGTAAAAC-ATAAAGTTAGAGACAA-3′ (U1) and reverse primer5′-TCTTGGAGCCGGTGAACGCGGGGCT-3′ (L1) are successful primer pairs used for amplification of the first intron inside the coding region, the PCR program was 95°C for 10 min, followed by 30 cycles of 94°C for 1 min, 68°C for 3 min and a final extension step of 72°C for 10 min. The forward primer 5′-GCGGTTTGCTACTGACGACTTGTTC-AATG-3′ (U362) and reverse primer 5′-TTACTACC-ACCGCTGCTAATGTCGGTGGC-3′ (L21) are successfully used to amplify the other introns in the coding region. The reaction was denatured at 95°C for 10 min, followed by 30 cycles of 94°C for 1 min, 62°C for 1 min, 72°C for 3 min and an extension at 72°C for 10 min. PCR products were recovered from agarose gels using the UNIQ-10 gel extraction kit (Sangon, Canada), subcloned using the pGEM-T easy vector system (Promega Corp., Madison, WI, USA) and fully sequenced.

RESULTS AND DISCUSSION

cDNA Library Construction, Isolation and Characterization of *CYP9G2*

With two specific PCR primers designed according to one of the cDNA sequence obtained from RT-PCR, we screened the cDNA library, five positive clones were obtained, and three positive clones of insert size longer than 1600 bp were chosen for determination of their nucleotide sequences. After sequencing, it was found that the sequences of these three cDNAs were identical except for the length of the 5′, one of them was selected and designated c34. This cDNA is 2143 bp in length, including 49 nucleotides of 5′ untranslated region (5′ UTR) upstream of the ATG (Fig. 1), this open reading frame codes for a predicted translation product that is 521 amino acids in length. The predicted molecular mass is 59.9 kDa, which is similar to that of the most P450 proteins and the predicted isoelectric point is 7.73 (Fig. 1). A TAA stop codon is found at nucleotide 1613, this is followed by 528 nucleotides of 3′ untranslated sequence, which includes the canonical AATAAA polyadenylation signal at nucleotides 1926–1931 upstream from a 29 bp poly (A) tail. This gene was named *CYP9G2* (accession number: AB096739) by the P450 nomenclature committee (D. Nelson, personal communication).

Consensus sequences have been derived for the region surrounding the start of translation in eukaryotes (Kozak, 1986). The consensus sequence for eukaryotes is CCACC**ATG**G and in *Drosophila* is CAAAAC**ATG**. It was found that the adenine at the -3 position relative to the start of translation is the most critical nucleotide. Of 17 insect P450 cDNA sequences multialigned, all but two contain A at the critical -3 position, in concordance with the observed eukaryotic and *Drosophila* consensus sequences (Rose *et al.*, 1997). In this study, the sequence surrounding the putative start of translation of *CYP9G2* (CAAAA**ATG**) does not fit either consensus.

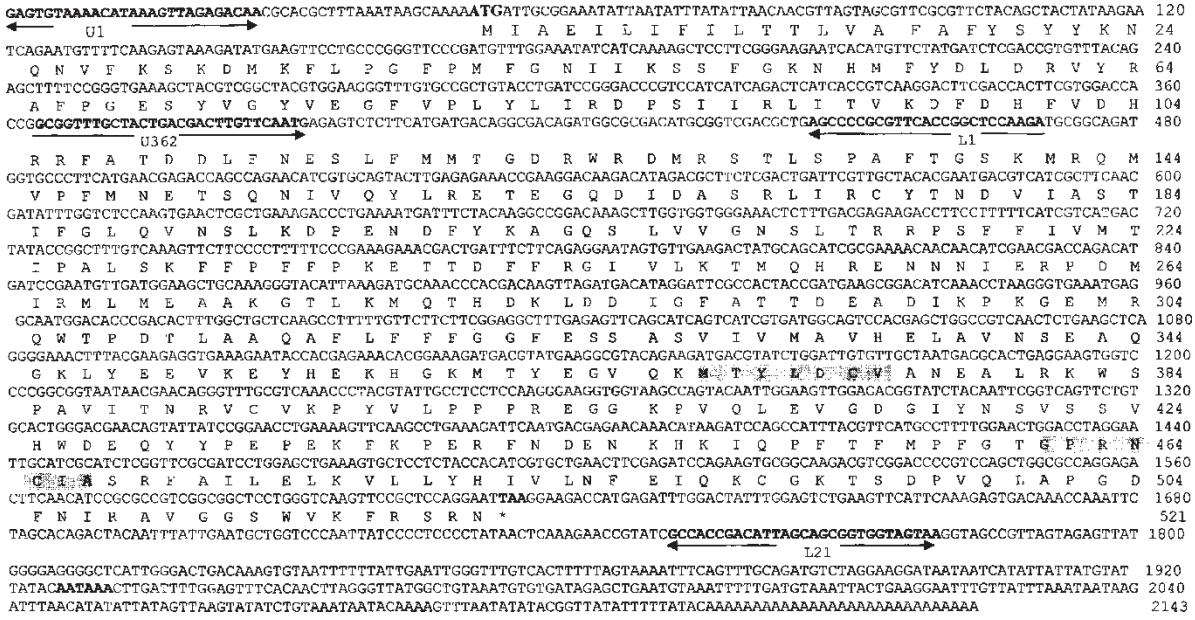

FIGURE 1 Nucleotide and amino acid sequences of diamondback moth cytovhrome P450, *CYP9G2*. The start codon, the stop codon and the polyadenylation signal is indicated in bold, the upper letters represent the nucleotide sequence, and the lower letters correspond to the amino acid sequence. The amino sequences shaded in grey are sequences used to design the degenerate primers for RT-PCR, the cDNA sequences in bold and underlined with arrow are sequences used for designing primers to characterize the genomic structure of *CYP9G2*. The GenBank accession number for *CYP9G2* is AB096739.

The nucleotide at -3 position is A, which correlates well with the general consensus sequence information from eukaryotes. The polymorphism at -3 position indicates that the translational machinery for P450s in insects may be less stringent than that required for most eukaryotic genes, or it may reflect the multiplicity of regulatory mechanisms acting upon the P450 family.

Characterization of the Predicted Protein of CYP9G2

CYP9G2 includes several sequence motifs common to all cytochrome P450 proteins. As in other microsomal proteins, in CYP9G2, a signal sequence, a sequence of 17 hydrophobic amino acids that are invariant of cytochrome P450, is found at the amino terminus of the predicted protein and it is also found that, 13 hydrophobic residues out of 17 are extremely hydrophobic. This suggested that it is a microsomal CYP. The signature heme-binding motif, FG/S*G*R*C*G***A/G (amino acids 458–471), conserved near the C-terminus of P450 enzymes (Schuler, 1996; Danielson and Fogleman, 1997), was also present in CYP9G2. The cysteine residue in this sequence is known to function as the fifth (axial) ligand of the heme iron and is responsible for the enzymes' characteristic red-shifted Soret peak at around 450 nm (Nebert and Gonzalez, 1987).

Genomic Structure of the *CYP9G2*

Comparison between CYP9G2 cDNA and its genomic sequence revealed that the CYP9G2 protein is encoded by eight exons with the size varying from 98 to 364 bp. The sizes of the introns vary from 112 to 2368 bp. All exon–intron boundaries conform to the AG...GT rules: namely GT as the 5′ donor sequence and AG as the 3′ acceptor sequence (Breathnach and Chambon, 1981) (Fig. 2). The genomic sequence of *CYP9G2* has been submitted to the DDBJ (Accession number: AB112959).

According to the splice site, the introns can be divided into three phases. A phase 0 splice site lies between two codons, while a phase I site lies one base inside the codon in the 3 prime direction and phase II intron lies two base inside a codon in the three prime direction (Tijet *et al.*, 2001). As for *CYP9G2*, of the seven introns we found, the third and the sixth intron are phase 0, the other five introns are phase I (Table I). With more genes being cloned in diamondback moth, it is possible to study the correlation between intron conservation and phylogenetic relationships between members of the P450 subfamilies.

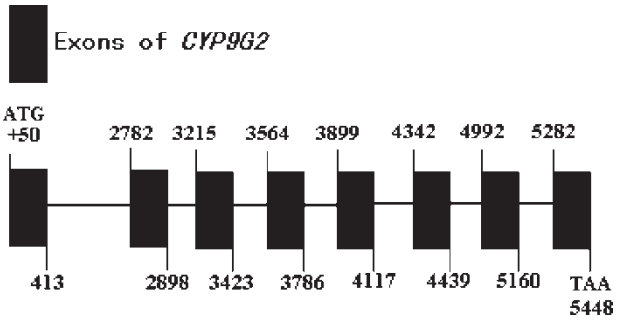

FIGURE 2 The genomic structure of *CYP9G2*. Comparison of genomic DNA sequence with cDNA of *CYP9G2* revealed that the cytochrome P450 gene spans about 5.7 kb comprising eight exons and seven introns.

Multiple Alignment of Members of Insect CYP9 Family

A BLAST search analysis indicated that this CYP9G2 exhibits the highest level of the similarity with other members of the CYP9 family. Multiple alignment of CYP9G2 with cytochrome P450 of other insect CYP9 family indicated that the degree of homology between the insect P450s is relatively low (Fig. 3). CYP9G2 has 37.5% identity to the *Blatella germanica* (German cockroach) CYP9E2, 38.7% identity to the *A. gambiae* (malaria mosquito) CYP9L1, and 39.5% identity to the *Heliothis virescens* (tobacco budworm).

In summary, we reported the molecular cloning of *CYP9G2* gene, exon and intron structure of the gene.

TABLE I Exon–intron boundary sequences of *CYP9G2* gene. Intron and exon nucleotide sequences are in lower-case and upper-case letters, respectively. Bold italics lettering indicate donor and acceptor splice sites

Exon		cDNA position			Intron		
No.	Length (bp)	CYP9G2	3′ splice acceptor	5′ splice donor	No.	Length (bp)	Phase
1	364	50–413		ATGATGACAG***gt***atgtgaga			
2	117	414–530	ctgacaac***ag***GCGACAGATG	TACTTGAGAG***gt***aattatca	1	2368	I
3	209	531–739	aatgtttc***ag***AAACCGAAGG	TTTGTCAAAG***gt***aagttgca	2	316	I
4	223	740–942	tcttttt***ag***TTCTTCCCCT	GAAATGAGGC***gt***aagtttta	3	140	0
5	219	943–1181	ttttctac***ag***AATGGACACC	GTTGCTAATG***gt***aagacatg	4	112	I
6	98	1182–1279	tgctttac***ag***AGGCACTGAG	GCCAGTACAA***gt***gagtcaa	5	224	0
7	169	1280–1448	taattttc***ag***TTGGAAGTTG	AATTGCATCG***gt***aggtttat	6	552	I
8	167	1449–1615	tcctccgc***ag***CATCTCGGTT		7	121	I

FIGURE 3 Mutiple alignment of the deduced amino acid sequence of *CYP9G2* with other insect CYP9 P450s. The alignment was performed using ClustalW (http://www.ebi.ac.uk/ClustalW.html) and modified manually. Sequences used in this figure include those from *A. gambiae* (malaria mosquito; CYP9L1; accession number: AAL96668), *B. germanica* (German cockroach; CYP9E2; accession number: AAK69411), *P. xylostella* (diamondback moth; *CYP9G2*; accession number: AB096739, this work), and *H. virescens* (tobacco budworm; CYP9A1; accession number: AAC25787). Asterisks denote a residue common to all aligned sequences.

Although the biological role of the cytochrome P450 remains unknown, the isolation of the full-length *CYP9G2* cDNA and characterization of its genomic structure should facilitate investigation of mechanism of insecticide resistance in diamondback moth.

Acknowledgements

We thank Dr Nannan Liu of Auburn University, USA, Dr Michel Raymond and Dr Myléne Weill of ISEM, CNRS, Montpellier University II, France, for their critical reading of the manuscript and helpful discussion as well as Dr David Nelson for classifying the diamondback moth P450 before its publication. We also appreciate two anonymous reviewers for their valuable comments of the manuscript. This work was supported by Grant KSCX3-IOZ-04 from CAS innovation program and Grant 2002AA601160 from National 863 project.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Research* **25**, 389–402.

Breathnach, R. and Chambon, P. (1981) "Organization and expression of eucaryotic split genes coding for proteins", *Annual Review of Biochemistry* **50**, 349–383.

Danielson, P.B. and Fogleman, J.C. (1997) "Isolation and sequence analysis of cytochrome P450 12B1: the first mitochondrial insect P450 with homology to 1 alpha, 25 dihydroxy-d3 24-hydroxylase", *Insect Biochemistry and Molecular Biology* **27**, 595–604.

Falckh, P.H., Balcombe, W., Haritos, V.S. and Ahokas, J.T. (1997) "Isolation and identification of a cytochrome P450 sequence in an Australian termite, *Mastotermes darwiniensis*", *Biochemical and Biophysical Research Communications* **241**, 579–583.

Feyereisen, R. (1999) "Insect P450 enzymes", *Annual Review of Entomology* **44**, 507–533.

Kozak, M. (1986) "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes", *Cell* **44**, 283–292.

Lardelli, M. and Lendahl, U. (1994) "Generating bacteriophage lambda sublibraries enriched for rare clones", *BioTechniques* **16**, 420–422.

Liu, N.N. and Zhang, L. (2002) "Identification of two new cytochrome P450 genes and their 5′-flanking regions from the housefly, *Musca domestica*", *Insect Biochemistry and Molecular Biology* **32**, 755–764.

Qiao, C.L. and Raymond, M. (1995) "The same esterase B1 haplotype is amplified in insecticide-resistant mosquitoes of the *Culex pipiens* complex from the Americas and China", *Heredity* **74**, 339–345.

Nebert, D.W. and Gonzalez, F.J. (1987) "P450 genes:structure, evolution, and regulation", *Annual Review of Biochemistry* **56**, 945–993.

Ranson, H., Nikou, D., Hutchinson, M., Wang, X., Roth, C.W., Hemingway, J. and Collins, F.H. (2002a) "Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*", *Insect Molecular Biology* **11**, 409–418.

Ranson, H., Clarles, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M.V., Unger, M.F., Collins, F.H. and Feyereisen, R. (2002b) "Evolution of supergene families associated with insecticide resistance", *Science* **298**, 179–181.

Rose, R.L., Goh, D., Thompson, D.M., Verma, K.D., Heckel, D.G., Gahan, L.J., Roe, R.M. and Hodgson, E. (1997) "Cytochrome P450 (CYP9A1) in *Heliothis virescens*: the first member of a new CYP family", *Insect Biochemistry and Molecular Biology* **27**, 605–615.

Schuler, M.A. (1996) "Plant cytochcrome P450 monogenases", *Critical Review of Plant Sciences* **15**, 235–284.

Scott, J.G. (1999) "Cytochrome P450 and insecticide resistance", *Insect Biochemistry and Molecular Biology* **29**, 757–777.

Scott, J.G. and Wen, Z.M. (2001) "Cytochrome P450 of insects: the tips of the iceberg", *Pest Management Science* **57**, 958–967.

Shen, B.C., Jin, Z.P., Zhao, D.X. and Qiao, C.L. (2003) "Construction of a full-length cDNA library of the diamondback moth, *Plutella xylostella*", *Zoological Research* **24**, 215–219.

Stevens, J.L., Snyder, M.J., Koener, J.F. and Feyereisen, R. (2000) "InducibleP450s of the CYP9 family from larval *Manduca sexta* midgut", *Insect Biochemistry and Molecular Biology* **30**, 559–568.

Tijet, N., Helvig, C. and Feyereisen, R. (2001) "The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron–exon organization and phylogeny", *Gene* **262**, 189–198.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. and Siebert, P.D. (2001) "Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction", *BioTechniques* **30**, 892–897.