REGULATION OF CYTOCHROME P450 GENES IN DROSOPHILA MELANOGASTER BY THE CHEMICAL ENVIRONMENT

BY

CYNTHIA MARIE MCDONNELL

DISSELTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor in Philosophy in Entomology in the Graduate College of the University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Professor May R. Berenbaum, Chair
Professor Mary A. Schuler, Director of Research
Professor Barry R. Pittendrigh
Professor Gene E. Robinson
ABSTRACT

Cytochrome P450 monooxygenases (P450s) play an important role in the adaptive response of insects and other animals to chemicals in the environment. P450s are heme-dependent enzymes that catalyze the addition of oxygen to a substrate in the presence of an electron donor, such as NADPH. In insects, some P450s are phase I detoxification enzymes, representing the first line of defense against lipophilic xenobiotics, while other P450s are involved in the biosynthesis of ecydsone, juvenile hormone and pheromones. P450s are tightly regulated throughout development, and P450s involved in detoxification have been shown to be substrate-inducible, instead of indiscriminately, constitutively expressed. Understanding constitutive and inducible expression requires knowledge about the regulatory pathways that control insect P450 expression, which is still lacking for most identified insect P450s. Drosophila melanogaster, because of its longstanding use as a genetic model organism, is a powerful tool for identifying possible regulatory mechanisms and for following expression through to function. The aim of this work is to examine the evolution of cytochrome P450 genes in D. melanogaster in response to xenobiotic compounds. First, I explore the role of constitutive and inducible expression of P450s in cross-tolerance of a methoprene-tolerant D. melanogaster strain to the fungal toxin, aflatoxin B1, a natural constituent in the diet. Next, I investigate the duplication of a xenobiotic-responsive P450 in D. melanogaster and across 11 Drosophila species and examine the changes incurred in a strain selected for DDT resistance. Finally, I address the function of a brain-specific P450 that is highly conserved across Drosophila species, by using a GAL4/UAS RNAi system to study the reproductive effects of knocking down the gene Cyp4g15. Using molecular modeling of the Cyp4g15 protein, I propose a potential substrate for this enzyme, the pheromone cis-vaccenyl acetate.
To my parents
ACKNOWLEDGMENTS

I first want to thank my main research advisers through the years, Drs. May Berenbaum and Mary Schuler, whose collaboration allows for a creative and engaging environment in which to work. In their research groups, my scientific education has been fostered by many people through the years. Drs. Zhimou Wen and Wenfu Mao were valuable for their instruction and discussions. Drs. Arthur Zangerl and Sanjeewa Rupasinghe were also available to offer instruction and answer my questions, as were Amy Dunlap, Iyoncy Rodrigo and Yinghong Pan.

Around two years ago, I began collaborating with Dr. Barry Pittendrigh, which introduced me to a whole new research group with people that I could bother with questions and discussion. Drs. Weilin Sun and Hongmei Li taught me the experimental methods of the lab, while Susan Balfe taught me the inner workings of the lab. I also appreciate the time that Jungkoo Kim spent discussing experimental design and scientific concepts with me.

In the Entomology Department, I had access to a range of expertise, in the professors, graduate students and staff, that was freely available and which I will miss. I want to thank Audra Weinstein and Karen Trame, our administrative staff, for their patience, especially as I neared completion of this dissertation. I also thank my friends in the department with whom I shared experiences that I believe we will always cherish.

I thank my parents, to whom I dedicate this dissertation, because they support me unconditionally but also pushed me to finish, when I was not always sure that I could. Dr. Cedric Alaux, also pushed me to finish, but maybe selfishly, so that we could finally reunite in France. I look forward to the next chapter of life that we write together.
<table>
<thead>
<tr>
<th>CHAPTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Introduction</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

Sensing and responding to environmental changes are key to the survival of an individual. While all five senses are translated into chemical responses via the nervous system, chemical perception also affects the molecular makeup of an organism directly. A chemical entering the body interacts with the endogenous proteins, which can start a cascade of regulatory events that change the gene expression profile of the organism. Genes that play a role in coping with chemicals in the environment are shaped by natural selection to respond to those stimuli. Detoxification represents one of the primary examples of adaptive response to environmental stimuli.

The cytochrome P450 monooxygenases (P450s) comprise a large gene family that is involved in every aspect of the sensing of a chemical stimulus, from its synthesis to its degradation. P450s are heme-dependent enzymes that catalyze the addition of oxygen to a substrate in the presence of an electron donor, such as NADPH. As detoxicative enzymes, P450s are phase I enzymes, representing the first line of defense against lipophilic xenobiotics (Ortiz de Montellano, 2005). Detoxifying P450s are often induced by xenobiotics or environmental cues that signal toxicity (Li et al., 2007).

P450s are not simply detoxicative enzymes; some act on endogenous substrates as in the biosynthesis of ecydsone, juvenile hormone and pheromones (Feyereisen, 2006; Gilbert, 2004). Xenobiotic compounds, that structurally mimic endogenous substrates of P450s, can interfere with endocrine signaling or metabolic pathways, leading to endocrine disruption. Endocrine disruptors can be novel compounds in the environment, such as synthetic byproducts of
industrialization, or naturally produced defensive compounds. To defend against endocrine
disruption, P450s are tightly regulated throughout development, and P450s involved in
detoxification have been shown to be substrate-inducible, instead of indiscriminately,
constitutively expressed. Characterizing the regulation of insects P450s, both constitutively and
inducible, is necessary for understanding their functional role in resistance to natural and
synthetic xenobiotics.

Regulation of insect P450s involved in host plant resistance

Insect P450s are known for their involvement in insecticide resistance in insect pest
species, but their role in host plant allelochemical resistance better depicts their longstanding role
in the adaptation of insects to their environment. In this evolutionary conflict, insect P450 genes
have evolved to be expressed when and where they are needed, during the feeding
developmental stages and in the relevant tissues (e.g., midgut, fat body, and Malphighian
tubules). Substrate-inducibility offers another level of regulatory control that minimizes the
costs of P450 expression.

The induction of P450s by host plant defensive compounds is associated with
allelochemical tolerance in insects. In the cactophilic Drosophila, P450s are induced by
isoquinoline alkaloids, the toxic host plant allelochemicals, and by the synthetic inducer
phenobarbital (Danielson et al., 1997). Larvae of the mosquito Aedes aegypti induce P450s when
exposed to plant allelochemicals released from leaf litter (David et al., 2006). Lepidopterans are
exceptional in their use of P450s to defend against host plant allelochemicals (Li et al., 2007).
Four Papilio species (swallowtails), Helicoverpa zea (corn earworm) and Depressaria
pastinacella (parsnip webworm), which range from a generalist that feeds on over 100 different
host plants (*H. zea*) to a specialist (*D. pastinacella*) that in North America is restricted to the closely allied genera *Pastinaca* and *Heracleum*, use P450s in the biotransformation of the toxic plant allelochemical, furanocoumarins (Li et al., 2007).

Chemical eavesdropping, the induction of P450s in by plant signaling compounds, has been proposed as a mechanism by which an insect can easily move across unrelated host plants without first evolving an inducible response to host-specific allelochemicals. In *H. zea*, P450s (*CYP6B8, CYP321A1*) are induced not only by plant allelochemicals and insecticides (Wen et al., 2009; Li et al., 2002a; Li et al., 2000) but also by the plant signaling compounds, methyl jasmonate (MeJ) and salicylic acid (SA) (Li et al., 2002b). Baculovirus-mediated expression of CYP6B8 and CYP321A1 proteins has demonstrated their ability to metabolize allelochemicals (xanthotoxin, chlorogenic acid, quercetin, and flavone) and insecticides (diazinon, cypermethrin, and aldrin) (Li et al., 2004; Sasabe et al., 2004). Chemical eavesdropping may also exact a cost, as the fungal toxin aflatoxin B$_1$ is bioactivated to a more toxic form (Zeng et al., 2006), by the P450s that are induced by MeJ and SA in this species, which results in significantly lower weight gain in larvae and pupae and significantly higher mortality in fourth instars (Zeng et al., 2009).

Studies of P450 regulation by host plant allelochemicals in *Papilio* caterpillars offer insight into the regulatory mechanisms that control expression. *Papilio* caterpillars range in feeding habits from the specialist *Papilo polyxenes* (black swallowail), which feeds entirely on furanocoumarin-containing plants in the Apiaceae, to the more generalist *Papilio glaucus* (Eastern tiger swallowtail) that occasionally feeds on furanocoumarin-containing plants in the family Rutaceae. Both species utilize P450s to metabolize furanocoumarins in the diet but the evolution of their P450s in transcriptional response and enzymatic function reflect the breadth differences in their diet.
*P. polyxenes* (black swallowtail) caterpillars, utilize CYP6B1, a substrate-inducible P450, to detoxify linear furanocoumarins (e.g. xanthotoxin) in their diet (Wen et al., 2003; Ma et al., 1994; Cohen et al., 1992). The enzyme CYP6B1 is notably efficient at metabolizing linear furanocoumarins and is less adept at metabolizing angular furanocoumarins or other plant compounds (e.g. coumarin). CYP6B4, isolated from *P. glaucus*, metabolizes both linear and angular furanocoumarins but at lower rates than CYP6B1 (Li et al., 2003; Hung et al., 1994).

*CYP6B1* is constitutively expressed in the caterpillar midgut and is induced by furanocoumarins in the diet (Petersen et al., 2001; Cohen et al., 1992), while *CYP6B4* is not constitutively expressed but is induced ~300-fold by furanocoumarins (Li et al., 2001). The promoter of *CYP6B1* contains multiple elements that are necessary for basal and inducible expression from multiple regulatory networks, including conserved xenobiotic response element to aryl hydrocarbons (XRE-AhR), an ecdysone response element (EcRE) and a novel xenobiotic response element to xanthotoxin (XRE-Xan) (Brown et al., 2005; Petersen Brown et al., 2004; Petersen et al., 2003; Prapaipong et al., 1994). The promoter of *CYP6B4* shares these response elements, underscoring the conserved nature of these regulatory cascades across *Papilio* species, even as their dietary habits have diverged (Li et al., 2002; Hung et al., 1996, 1997). In the promoter of *CYP6B4* from *P. glaucus*, XRE-Xan is necessary for xanthotoxin and benzo(α)pyrene-inducible expression, but does not confer high constitutive expression, as seen in *CYP6B1* (McDonnell et al., 2004). Thus, ecological differences between these swallowtail caterpillars are reflected by the function of their P450s, CYP6B1 and CYP6B4, and their differential expression is, in turn, mediated by changes in the conserved response elements contained within their promoters.

*Constitutive over-expression of P450s associated with insecticide resistance*
P450-mediated insecticide resistance often evolves as changes in transcriptional regulation that lead to constitutive over-expression of the gene. On a shorter evolutionary time scale, changes in the expression of a xenobiotic-metabolizing gene are more likely than changes that create a new function for an existing protein. Often the assumption is made that an over-expressed P450 gene encodes a protein that can metabolize the insecticide, although it is rarely demonstrated.

Resistance to permethrin in natural populations of *Musca domestica* (house fly) in New York has been functionally correlated with the resistant alleles of *CYP6D1* and *Vssc* (voltage-sensitive sodium channel) (Rinkevich et al., 2007). CYP6D1 protein isolated from the thoracic ganglia of the house fly metabolizes cypermethrin (Korytko and Scott, 1998). Constitutive expression of *CYP6D1v1*, an allele from the Learn Pyrethroid Resistant (LPR) strain, was 10-fold higher than in susceptible strains (Liu and Scott, 1998). Constitutive over-expression of the *CYP6D1v1* allele is due to disruption of a binding site for a repressor protein, the zinc finger protein Gfi-1, by a 15-nucleotide insertion in the 5’ noncoding region (Scott et al., 1999; Gao and Scott, 2006). *CYP6D1v1* is found in high frequency in natural populations (Rinkevich et al., 2006), although its frequency was observed to increase over summer and decline over the winter, which suggests a cost associated with overwintering (Rinkevich et al., 2007).

In *Drosophila melanogaster*, constitutive over-expression of *Cyp6g1* is associated with resistance to multiple insecticides, namely DDT and imidacloprid, in natural populations (Daborn et al. 2001) and DDT, dicyclanil, and nitenpyram in transgenic over-expressing flies (Daborn et al. 2007). Over-expression of *Cyp6g1* in natural populations relies on the Accord transposable element, a long terminal repeat retrotransposon, that seems to have spread rapidly as it was under positive selection (Daborn et al., 2002; Catania et al., 2004), with no associated
costs yet identified (McCart et al., 2005). This element, which resides 291 nucleotides upstream of the Cyp6g1 transcription site, contains enhancers that control for tissue-specific increased expression in tissues related to detoxification, i.e., the midgut, Malphigian tubules, the fat body and gastic caecum (Chung et al., 2007). The CYP6G1 protein, when expressed in a Nicotiana tabacum cell culture, dechlorinates DDT to DDD and hydroxylates imidacloprid to 4- and 5-hydroxyimidacloprid, thus demonstrating the ability of this protein to metabolize these compounds (Joußen et al., 2008). Metabolism of imidacloprid by CYP6G1 was inhibited by piperonyl butoxide (Joußen et al., 2010).

Although CYP6G1 can metabolize DDT, in some DDT-resistant strains, there is no correlation between resistance and over-expression of Cyp6g1 (Kuruganti et al., 2007; Festucci-Buselli et al., 2005). Conversely, within D. melanogaster strains that over-express Cyp6g1, resistance to DDT is variable. In Drosophila simulans, over-expression of the Cyp6g1 ortholog is associated with insertion of a Doc transposable element in the promoter, but this insertion is associated with DDT resistance in only one population of eight studied (Schlenke and Begun, 2004). While the insertion of transposable elements in the promoter of Cyp6g1 results in their over-expression, they do not alone account for DDT resistance. In the DDT resistant strain Wisconsin (Rst(2)DDTWisconsin), a region on chromosome 2 contributes to its high levels of DDT-resistance (>30-fold compared to Canton-S). Six cytochrome P450 genes map to this region and two of them, Cyp12d1 and Cyp6g1, are constitutively over-expressed in this strain, 6-fold and 4.3 fold, respectively, yet only Cyp12d1 is inducible by DDT (Brandt et al. 2002). This inducibility suggests that Cyp12d1 has evolved to respond to xenobiotics, such as DDT, in D. melanogaster, in which P450s are one piece of the puzzle of DDT resistance.
Inducible expression of P450s in Drosophila melanogaster

Constitutive over-expression, like that seen in CYP6D1 in M. domestica or Cyp6g1 in D. melanogaster, is selected to occur quickly and spread rapidly with few genetic changes. This may explain why the insertion of transposable elements is more often observed in detoxification-associated P450s than those involved in ecdysteroid biosynthesis (Chen and Li, 2007). Inducible expression may require more evolutionary time to develop as more molecular pathways contribute to the response.

In Drosophila melanogaster, Cyp6a2 and Cyp6a8 were also initially identified as P450s conferring DDT resistance due to their over-expression in the DDT-resistant strain 91-R, when compared to its control strain 91-C. Although the protein Cyp6a2 has been shown to metabolize a range of insecticides (aldrin, diazinon, and heptachlor) (Dunkov et al., 1997), the wildtype allele does not efficiently turn over DDT (Amichot et al., 2004), and Cyp6a8 better metabolizes the fatty acid lauric acid than DDT (Helzig et al., 2004).

Although their association to DDT resistance now seems tenuous, the investigation of the transcriptional regulation of Cyp6a2 and Cyp6a8 by the natural inducer caffeine has furthered our understanding of both constitutive and inducible regulation of insect P450 genes. Deletion analysis of the Cyp6a2 and Cyp6a8 promoters has localized regions needed for basal and caffeine-inducible activity (Bhaskara et al., 2006). Within this region lie putative Ap-1 sites, which, in vertebrates, are regulated by the Jun/Fos heterodimer and also respond to elevated cAMP levels (Bhaskara et al., 2006). Exogenous treatment of flies or SL-2 cells with cAMP activated the Cyp6a2 and Cyp6a8 promoters (Bhaskara et al., 2008). In SL-2 cells, caffeine treatment increased intracellular cAMP, but surprisingly also reduced the protein level of D-JUN, an AP-1 transcription factor without inhibiting transcription. Thus, the researchers
concluded that D-JUN and D-FOS act as negative regulators of the \textit{Cyp6a2} and \textit{Cyp6a8} promoters via Ap-1 sites (Bhaskara et al., 2008). This work is the first to identify the transcriptional regulatory proteins that mediate inducible expression of an insect P450.

Constitutive over-expression is a common mechanism of P450-based insecticide resistance, as the coordination of multiple molecular pathways is necessary for inducible expression. Understanding constitutive and inducible expression requires knowledge about the regulatory pathways that control insect P450 expression, which is still lacking for most identified insect P450s. \textit{D. melanogaster}, because of its longstanding use as a genetic model organism, is a powerful tool for identifying possible regulatory mechanisms and for following expression through to function. The aim of this work is to examine the response of cytochrome P450 genes in \textit{D. melanogaster} to xenobiotic compounds, in order to characterize the regulatory mechanisms of insect P450s and the functional significance of P450 expression for development and reproduction.

In Chapter 2, I investigate the role of a transcriptional regulatory protein, Methoprene-tolerant (Met), in the constitutive and inducible expression of cytochrome P450 genes in \textit{D. melanogaster}. Mutation of the Met protein in a \textit{D. melanogaster} strain confers tolerance to the insecticide methoprene, a juvenile hormone analog. In comparing the methoprene-tolerant strains with a wildtype strain, two P450 genes, \textit{Cyp6a2} and \textit{Cyp4g15}, were differentially expressed between strains in a methoprene-dependent manner whereas three genes, \textit{Cyp6d4}, \textit{Cyp6d5} and \textit{Cyp314a1} were constitutively over-expressed in the methoprene-tolerant strain, independent of methoprene. These results offer a glimpse of differential P450 expression between the two strains that may play a role in methoprene tolerance and therefore merit further investigation.
In Chapter 3, I examine the role of inducible expression of two P450s, Cyp6a2 and Cyp6d2, in the observed cross-tolerance of the methoprene-tolerant strain to the fungal toxin, aflatoxin B1, a natural constituent in the diet of *Drosophila melanogaster*. I demonstrate the differential effects of aflatoxin B1 and methoprene on larval and pupal development between the methoprene-tolerant strain Rst(1)JH1 and a wildtype strain, Oregon-R. Differences in the transcriptional expression profiles of two cytochrome P450 genes, Cyp6a2 and Cyp6d2, were associated with tolerance to aflatoxin B1 and methoprene. Cyp6a2 is expressed at negligible levels in the Rst(1)JH1 strain, suggesting that it may bioactivate aflatoxin B1 and/or methoprene, and contribute to their toxicity in *D. melanogaster*. Conversely, Cyp6d2 increases in the presence of aflatoxin B1, in a dose-dependent manner, which suggests that it may detoxify that compound.

In Chapter 4, I focus on the evolution of Cyp12d1 across *Drosophila* species and within *D. melanogaster* strains. In *D. melanogaster*, Cyp12d1 is induced by a range of xenobiotic compounds and appears to have recently duplicated into two tandemly arranged genes. This presents a unique scenario to study duplication events of cytochrome P450 genes, a gene superfamily that has evolved by duplication and divergence. In order to determine if Cyp12d1 recently duplicated in *D. melanogaster*, I searched the sequenced genomes of 11 other species of *Drosophila* for orthologous genes. I observed that the Cyp12d1 duplication, which introduced a new gene Cyp12d2, is shared across the genomes of the 6 species in the melanogaster group of *Drosophila* and the more basal lineages retain a single copy of the gene. Thus, it seems that Cyp12d1 originally duplicated between 10-20 million years ago after the melanogaster subgroup split from the basal lineages, and has diverged to different extents across the 6 species.
In order to study the Cyp12d1 region in *D. melanogaster*, I amplified the region with flanking primers based on the reference strain (*y; cn bw sp*) and discovered that the region varies between a double copy seen in the reference genome and a single copy in two lab strains, 91-C and 91-R. In order to determine the extent of this variation, I then surveyed 55 additional wildtype stocks of *D. melanogaster* and found that 47 strains have a single copy of the Cyp12d1 gene. Those strains that have the duplication do not share geographical or temporal.

In order to study the evolution of transcriptional regulation of Cyp12d1, I sequenced the Cyp12d1 region from two strains of *D. melanogaster*, 91-C and 91-R, that were derived from a common population founded from several hundred individuals collected in St. Paul, Minnesota in 1952 (Dapkus and Merrell, 1977), and one was selected for DDT resistance (91-R), while the other was never exposed to DDT (91-C). Sequencing the Cyp12d1 genomic regions of 91-C and 91-R, which show differential Cyp12d1 response to DDT, revealed that no major differences occur in the 5’ non-coding region. Instead, 91-R has acquired a splice site mutation in the coding region that may affect translation of the protein. In addition, 91-C and 91-R both differed from the reference strain in the 3’ UTR, which may affect mRNA polyadenylation. Variation in transcript length has previously been linked to differential responses to DDT between 91-C and 91-R. Selection for mRNA transcript length may have lead to differences in the translation of the protein that may be associated with differences in function between the two strains.

The findings for the Cyp12d1 region depart from commonly observed mechanisms of regulation of cytochrome P450 genes in which increased expression occurs via selection for induction or constitutive overexpression (Li et al., 2007; Feyereisen, 2005). In contrast, the Cyp12d1 region varies in gene copy number and the polyadenylation signal sequence in the 3’
UTR across strains of *D. melanogaster*, perhaps as novel mechanisms of responding to the selection pressure of xenobiotic compounds.

In Chapter 5, I address the function of *Cyp4g15*, a brain-specific P450 that is highly conserved across *Drosophila* species, by using a GAL4/UAS RNAi system to study the behavioral effects of knocking down the gene *Cyp4g15*. Based on its restricted expression in the brain of larvae and male adults, I predicted that Cyp4g15 would affect courtship behavior in males. When I tested *Cyp4g15* knockout males for changes in their courtship behavior and their reproductive success, I found that they had reduced fecundity, and even though they initiated courtship, the length of their courtship was extended. Thus, knocking out *Cyp4g15* does not abolish courtship but seems to alter some feature of the complex communication involved. I propose that *Cyp4g15* may function in the metabolism of the courtship pheromone, cis-vaccenyl acetate, which would lead to miscommunication between males and females, if knocked out. Based on molecular modeling of the Cyp4g15 protein, cis-vaccenyl acetate fits in the putative catalytic site. Comparative genomics shows that Cyp4g15 is conserved across at least 10 of the 12 *Drosophila* species sequenced, which supports a conserved role in reproductive function.

**LITERATURE CITED**


CHAPTER 2

Regulation of cytochrome P450 transcriptional expression in Drosophila melanogaster by methoprene and the Met protein

INTRODUCTION

Cytochrome P450 genes encode membrane-bound enzymes that are heme-associated, NADPH-dependent; these enzymes catalyze the addition of oxygen to lipophilic molecules, which renders them more hydrophilic for excretion in phase 2 detoxification. Of the 86 full length cytochrome P450 genes (P450s) that have been identified in the Drosophila melanogaster genome (Tijet et al., 2001), the cis and trans-regulatory elements have been characterized for only a handful, even though transcriptional regulation is considered the main mechanism of P450-based xenobiotic resistance in insects. In D. melanogaster, P450s are associated with insecticide resistance but are also crucial in the synthesis and degradation of the insect hormones ecdysone and juvenile hormone (Feyereisen, 2006). Resistance through transcriptional regulation has been shown to occur through an insertion or deletion in the non-coding regulatory region of the gene, although the precise mechanisms have rarely been elucidated. Understanding the regulation of P450s that increase toxicity of an insecticide could aid in the use of negative cross-resistance, a concept emerging as a possible tool in insecticide resistance management (Pittendrigh et al., 2008; Pittendrigh et al., 2004; Pittendrigh and Gaffney, 2001; Pittendrigh et al., 2000).

In vertebrates, increased transcription of cytochrome P450 genes by the aryl hydrocarbon receptor (AhR) and its heterodimerization partner (ARNT) is toxic, because cytochrome P450 enzymes bioactivate aryl hydrocarbons to render them more reactive (Guengerich, 2008; Nebert et al., 2004). AhR binds aryl hydrocarbons, flavones and coplanar polychlorinated biphenyls
AhR/ARNT binds XRE-AhR, the xenobiotic response element to aryl hydrocarbons (Denison et al., 1988a; Denison et al., 1988b), an upstream regulator of the cytochrome P450 genes CYP1A1 and CYP1B1 (Denison and Whitlock, 1995).

Benzo(α)pyrene, a lipophilic planar compound, induces cytochromes P450 via AhR/ARNT and is subsequently hydroxylated to a more reactive molecule. Although documented primarily in vertebrates, induction of cytochrome P450 activity by benzo(α)pyrene is correlated with toxicity in a mutant strain of Drosophila simulans, but not in a benzo(α)pyrene-resistant strain, in which P450-mediated benzo(α)pyrene hydroxylase activity cannot be induced (Fuchs et al., 1992).

The P450 genes CYP6B1 and CYP6B4 in the swallowtail caterpillars, Papilio polyxenes and P. glaucus, respectively, which encode proteins capable of metabolizing the highly reactive furanocoumarins (Cohen et al., 1992; Ma et al., 1994; Hung et al., 1995b; Chen et al., 2002; Baudry et al., 2003; Wen et al., 2003), were the first insect P450s to be linked to the aryl hydrocarbon response cascade. The CYP6B1 and CYP6B4 promoters are induced by the furanocoumarin xanthotoxin and the aryl hydrocarbon benzo(α)pyrene, which suggests a shared regulatory mechanism for these compounds in insect P450 expression (McDonnell et al., 2004; Hung et al., 1997; Cohen et al., 1992). In addition, the CYP6B1 promoter contains a xenobiotic response element to the aryl hydrocarbon receptor (XRE-AhR), similar to the vertebrate XRE-AhR (Denison et al., 1988), which is necessary for basal and inducible expression of the promoter and which can be activated by Drosophila homologs of AhR/ARNT (Brown et al., 2005). The fact that this insect P450 gene contains response elements similar to those in
mammalian genes that are the targets of aryl hydrocarbon regulatory cascades suggests that insect promoters that are inducible by aryl hydrocarbons may share features of the regulatory mechanisms found in vertebrates, including a role for bHLH-PAS proteins.

In *D. melanogaster*, orthologs of the AhR regulate embryonic nervous system development (Sonnenfeld et al., 1997; Nambu et al., 1991) and antennal/leg differentiation in larval imaginal discs (Emmons et al., 1999; Duncan et al., 1998), but, unlike AhR homologs in vertebrates, they do not bind any ligands (Hahn, 1998). The Met protein was the first bHLH-PAS in insects demonstrated to bind a ligand, methoprene, a juvenile hormone analog (JHA) (Ashok et al., 1998; Miura et al., 2005). In *D. melanogaster*, methoprene disrupts the metamorphic reorganization of the central nervous system, salivary glands and musculature in a dose-dependent manner (Restifo and Wilson, 1998). Selection studies in *D. melanogaster* show that mutations of a single gene, which reduce either transcription of the gene or ligand affinity of the protein, can lead to methoprene tolerance (Turner and Wilson, 1995; Wilson, 1996; Wilson and Ashok, 1998). The Met gene encodes a bHLH-PAS receptor, homologous to the vertebrate aryl hydrocarbon receptor (AhR), which binds ligand in the PAS domain of the protein and binds DNA to activate transcription at the basic helix-loop-helix (bHLH) domain (Ashok et al., 1998). Thus, mutations in the Met receptor reduce the toxic response to methoprene to which wildtype insects are susceptible. Overexpression of the Met gene in *D. melanogaster* resulted in higher sensitivity to methoprene application at the critical developmental time as wildtype flies, which confirms the stage-dependent effects of JH in *D. melanogaster* (Barry et al., 2008).

The presence of Met receptor in the imaginal cells of larvae and pupae that undergo transformation during metamorphosis implicates those tissues as targets of methoprene. The Met receptor is also found in the fat body (Pursley et al., 2000), a tissue that is analogous to the
human liver in some insects, which suggests a role for the Met receptor in detoxification. In the fat body, JH disrupts normal induction of programmed cell death by MET and GCE induced caspase-dependent programmed cell death that occurs during metamorphosis as a result of 20-ecdysone activation (Liu et al., 2009). At very high levels, in which overexpression is driven by an actin or tubulin promoter, larvae experience high mortality, which may occur from ectopic tissue expression (Barry et al., 2008). Met null mutant males demonstrated reduced protein accumulation in their accessory glands and diminished courtship and mating efforts with females, relative to wildtype or rescued Met males (Wilson et al., 2003). Exposure to methoprene partially rescued the courtship deficiency (Wilson et al., 2003).

From studies in *D. melanogaster*, a molecular model of Met protein action is beginning to emerge. The Met protein binds JH III at physiological concentrations and mediates JH- and JHA-dependent reporter gene activation (Miura et al., 2005). Met heterodimerizes with germ cell expressed (gce), a bHLH-PAS protein with which is shares ~70% identity (Moore et al., 2000), and Met/Met and Met/gce dimerization is reduced in the presence of JH or methoprene (Godlewski et al., 2006). An interaction of the Met protein with Broad Complex Z1 isoform has been proposed using double mutants of *Met* and *BR-C* that show more severe defects in double mutants than would be expected if the gene acted independently and that could not be rescued by complementation with other *BR-C* mutants (Wilson et al., 2006). Application of JH or a JHA disrupts the interaction between Met and BR-C (Wilson et al., 2006). Thus, a model has been proposed in which Met and BR-C bind directly, through the PAS domain of Met, to activate genes for the pupal stage, and JH, when present, disrupts this interaction by binding to Met (Wilson et al. 2006b; Wilson, 2004).
The Met protein has been also shown to homodimerize and heterodimerize with other proteins in the ecdysone response cascade, the ecdysone receptor (EcR) and Ultraspiracle (Usp) (Bitra and Palli, 2009). The Met protein interacts with EcR, Usp and two additional proteins (FKBP39 and Chd64) to bind a JH response element identified in *D. melanogaster* and *Apis mellifera* (honey bee) genes (Li et al., 2007). Different combinatorial interactions between these proteins are proposed to depend on the presence of JH or ecdysone throughout development (Li et al., 2007). These two models are not mutually exclusive, and the evidence to support them underscores the possibility that Met functions differently in the presence and absence of JH.

The potential link between the Met protein and cytochrome P450 regulation deserves exploration in model insect systems, such as *D. melanogaster*. By identifying cytochrome P450 genes co-regulated by methoprene or the Met receptor in insects, we can begin to identify common elements in the regulation of insect cytochrome P450s. The Met mutant strain Rst(1)JH\(^1\) provides a tool for determining if the Met protein can regulate cytochrome P450 expression in *D. melanogaster*. It is not known if the Met protein uses the XRE-AhR binding site in regulating gene expression, but by searching the promoters of cytochrome P450 genes regulated by Met protein, a correlation between their expression and presence of XRE-AhR binding sites may be established.

**MATERIALS AND METHODS**

*Chemicals*

Methoprene (98% mix of isomers) was purchased from Chem Service (West Chester, PA). Methanol was purchased from Fisher Scientific (Pittsburgh, PA). The RNAeasy RNA isolation kit was from Qiagen (Valencia, CA). Superscriptase III reverse transcriptase, *Taq* polymerase and deoxynucleotide triphosphates (dNTPs) were from Invitrogen (Carlsbad, CA).
Insects

*Drosophila melanogaster* strains Oregon R and Rst(1)JH1 were obtained from Bloomington *Drosophila* Stock Center (Bloomington, IN). wvMet27 was kindly provided by Dr. Thomas Wilson at The Ohio State University. Flies were reared in plastic bottles or vials on standard cornmeal-agar diet recipe from the Bloomington Stock Center. Flies were transferred to new bottles every two weeks.

Methoprene exposure assays, RNA isolations and RT-PCR Southern analyses

First instars (n=15) of *D. melanogaster* were placed on diet containing methoprene diluted in methanol at 3.2 ppm or methanol alone and were raised at room temperature (24ºC). Four days later, larvae were collected and flash frozen in liquid nitrogen. Larvae from multiple vials were pooled for RNA isolation. Total RNA was extracted using Trizol extraction method, treated with DNase I, and re-extracted with phenol-chloroform. First-strand cDNA was synthesized in a 13 µL reaction containing 1 µg of RNA, 500 ng of oligo (dT)17, and 0.5 mM each dNTP by heating to 65ºC for 5 min and incubating on ice for 1 min. To this reaction 4 µL of 5X First-Strand buffer (Invitrogen), 1 µL of 0.1 M DTT, 1 µL of rRNA sin (40 µg/µL) (Promega Corp., Madison, WI) and 200 units of SuperScript III reverse transcriptase (Invitrogen Corp., Carlsbad, CA) were added, incubated at 50ºC for 60 min, and inactivated by heating at 70ºC for 15 min.

Cytochrome P450 genes and actin mRNA were quantified by RT-PCR Southern blot analysis in separate 50 µL PCR reaction mixtures containing 2 µL of first-strand cDNA, 1X PCR buffer minus Mg (Invitrogen), 0.2 mM each dNTP, 1.5 mM MgCl2, 2.5 units Taq DNA
polymerase (Invitrogen). Cytochrome P450 and actin genes were amplified with 20 and 10 pmol, respectively, of forward primer and reverse primer (Table 2.1, 2.2). Control experiments determined that for PCR 18 cycles of actin, 20 cycles of Cyp12d1, Cyp315a1, and 25 cycles of Cyp6a2, Cyp6a8, Cyp6d4, Cyp6d5, Cyp4g15, Cyp314a1, Cyp6g1, Cyp4e2, Cyp4d1, Cyp4ac1 and Cyp12a4/5 were within the linear range of RT-PCR amplification for whole body of larvae.

For quantification of cytochrome P450 and actin mRNA, ten µL of the RT-PCR amplified products were run on 1.5% 1X Tris-borate/EDTA (TBE) gels at 150 V for 30 min. RT-PCR Southern gels were denatured by soaking in denaturing solution (0.5 M NaOH) for 15 min and in neutralizing solution (1 M Tris, pH 7.4) for 15 min. Denatured RT-PCR products were transferred to Hybond-N nylon membranes overnight by capillary action. Actin Southern blots were probed with 32P-labeled 200 bp cDNA fragment of D. melanogaster actin gene. Probes were confirmed by sequencing. Hybridization was done at high stringency in 50% formamide, 5X SSC, 50 mM Na2HPO4 (pH 7.0), 0.5% SDS, 5X Denhardt’s solution at 42°C overnight. All Southern blots were washed at low stringency twice in 2X SSC, 0.1% SDS at 42°C for 15 min. Hybridized blots were quantified by phosphoimager analysis. Phosphoimager values for cytochrome P450 genes were corrected by background readings from the blots and normalized against the actin values.

**Promoter sequence analysis for putative transcriptional elements**

Searches of promoters of cytochrome P450 genes for putative transcriptional regulation elements were conducted using the MatInspector tool of the Genomatix database (www.genomatix.de). Sequences with 75% or higher similarity to the core canonical sequence in the database were reported by the program.
RESULTS

Differential cytochrome P450 gene expression in Oregon-R, Rst(1)JH$^1$ and Met$^{27}$

A sublethal dose of methoprene (3.2 ppm) did not strongly induce or repress the cytochrome P450 genes that were tested in the wildtype D. melanogaster strain, Oregon-R or the methoprene-tolerant strains, Rst(1)JH$^1$ and Met$^{27}$. Cyp6a2, a gene associated with insecticide resistance, was slightly induced by methoprene in Oregon-R, while its constitutive and inducible expression was completely abolished in the Rst(1)JH$^1$ strain, but not the Met$^{27}$ strain. Four other xenobiotic-responsive genes, Cyp6d5, Cyp6d4, Cyp12a4/5 and Cyp4e2, were constitutively expressed at higher levels in the methoprene-tolerant strains than in Oregon-R (Fig 2.1), whereas the xenobiotic-responsive genes Cyp12d1, Cyp6g1 and Cyp6a8 showed no difference in expression between strains (Fig 2.1). Cyp314a1, the ecdysone 20-hydroxylase that is expressed in the fat body and midgut (Petryk et al., 2003), was higher in the methoprene-tolerant strains, while expression of Cyp315a1, the ecdysone 2-hydroxylase that is restricted to the ring gland (Warren et al., 2002) did not differ between Oregon-R and Rst(1)JH$^1$ but showed two-fold higher constitutive expression in Met$^{27}$ (Fig. 2.3). Cyp4g15, a P450 with expression restricted to the brain (Maibeche-Coisne et al., 2000), was constitutively lower in the methoprene-tolerant strains (Fig. 2.1), as was Cyp4d1, a gene expressed in the midgut and fat body (Chung et al., 2009) and downregulated by wheat germ agglutinin (Li et al., 2009) and an ecdysteroid agonist (Davies et al., 2006) (Fig 2.1). Expression of Cyp4ac1, which is also downregulated by the same ecdysteroid agonist (Davies et al., 2006), did not differ between strains (Fig 2.1). These expression data provide the first view of differential cytochrome P450 expression between
methoprene-tolerant strains and the wildtype Oregon-R strain. Further investigation of P450 expression and its relevance to methoprene tolerance is needed.

*Enrichment of XRE-AhR, bHLH-PAS and BrC-Z1 binding sites in P450 promoters*

Eight of the candidate P450 genes tested have putative BrC-Z1 binding sites (Table 2.3) and eleven genes contain putative XRE-AhR binding sites (Table 2.4) in the upstream regulatory region within 1 kb of their transcription start site. Binding sites with 75% or higher similarity to the canonical binding sites are reported here, yet their functionality is now known without further testing. The genes *Cyp6a2*, *Cyp6d4* and *Cyp4ac1* each have three putative BrC-Z1 binding sites, while *Cyp12d1*, *Cyp12a5*, *Cyp314a1*, *Cyp6g1* and *Cyp4g15* each have one. *Cyp4e2*, *Cyp4d1*, *Cyp6d5*, *Cyp315a1* and *Cyp6a8* contain no putative BrC-Z1 binding motifs. For XRE-AhR and similar bHLH-PAS protein binding sites, with four putative sites, *Cyp12a4* has the most while *Cyp12a5* has no putative sites in its promoter. Many of the genes have 3 putative XRE-AhR binding sites, including *Cyp4d1*, *Cyp6g1*, *Cyp314a1*, *Cyp315a1* and *Cyp12d1*. Four genes, *Cyp4g15*, *Cyp6a8*, *Cyp6d4* and *Cyp4ac1*, each have one putative XRE-AhR binding site. The genes *Cyp4d1*, *Cyp4e2* and *Cyp6d5* have no putative XRE-AhR binding sites.

Seven of the P450 genes (*Cyp4g15*, *Cyp6a2*, *Cyp6d4*, *Cyp4ac1*, *Cyp314a1*, *Cyp6g1*, *Cyp12d1*) contain both XRE-AhR and BrC-Z1 putative binding sites in the 1 kb genomic sequence upstream of the transcription start site. Two of these genes, *Cyp4g15* and *Cyp6a2*, demonstrated methoprene and Met protein-dependent expression in opposing directions. Of the other genes, *Cyp6d4* and *Cyp314a1* were constitutively higher in the methoprene-tolerant strains, while *Cyp6g1*, *Cyp12d1* and *Cyp4ac1* did not differ between strains and do not respond to
methoprene treatment. All of the genes except Cyp4e2 contain at least one XRE-AhR or BrC-Z1 element.

The putative XRE-AhR and BrC-Z1 response elements within the Cyp4g15 promoter are separated by 439 nucleotides and within the Cyp6a2 promoter are separated by 918, 682, and 259 nucleotides. Cyp4g15 contains a single BrC-Z1 site at -362/-344 region upstream of the transcription start site and a XRE-AhR sites at -817/-801. Cyp6a2 has a single XRE-AhR site at +102/+126, downstream of the transcription start site, and three BrC-Z1 sites at -834/-816, -598/-580 and -275/-257.

**DISCUSSION**

Measuring transcriptional expression profiles of 13 cytochrome P450 genes in response to a sublethal dose of methoprene and in strains with a mutated Met protein has identified two genes, Cyp4g15 and Cyp6a2, that respond to methoprene in a Met protein-dependent manner, but in different ways that can be explained by two contrasting models of methoprene regulation via the Met protein.

In the case of Cyp4g15, treatment with methoprene and mutation of the Met protein represses expression, which suggests a positive regulatory role in constitutive expression of a gene for the Met/gce heterodimer, perhaps in conjunction with the Broad protein, as proposed by Wilson et al. (2006b). In the case of Cyp6a2, methoprene treatment induces expression and that inducibility is lost with the mutation of the Met protein. This model of regulation posits that the Met protein when binding methoprene aids in the induction of constitutive expression of the gene, perhaps in different combinations of regulatory proteins, as in the model proposed in the JH-dependent regulation by the ecdysone receptor complex (Li et al., 2007). Thus, the models of
regulation for \textit{Cyp4g15} and \textit{Cyp6a2} propose roles in constitutive and inducible expression for insect cytochrome P450 genes.

\textit{Cyp6a2} and \textit{Cyp4g15} were the only genes to respond to methoprene in a Met-protein dependent manner out of seven genes that contain both putative XRE-AhR and BrC-Z1 binding sites. Thus, identifying putative binding sites \textit{a priori} does not consistently predict those genes that will respond to methoprene. Identifying putative binding sites offers a testable hypothesis for the role of the XRE-AhR and Broad complex signaling cascades in the regulation of the \textit{Cyp6a2} and \textit{Cyp4g15} promoters by methoprene.

From this survey of 13 cytochrome P450 genes in \textit{D. melanogaster}, two genes, \textit{Cyp6a2} and \textit{Cyp4g15}, were observed to respond to methoprene in opposing directions, and those responses were disrupted in the methoprene-tolerant strains with mutated Met proteins. In addition, three genes, \textit{Cyp6d4}, \textit{Cyp6d5} and \textit{Cyp314a1}, were constitutively overexpressed in methoprene-tolerant strain, independent of methoprene exposure. While their consistency needs to be determined, these results suggest that the involvement of cytochrome P450 genes in methoprene tolerance bears further examination.

\textbf{ACKNOWLEDGMENTS}

This work was funded by an EPA STAR fellowship and a University of Illinois Research Board Grant to Mary Schuler.

\textbf{LITERATURE CITED}


Wilson, T. G. 2004. The molecular site of action of juvenile hormone and juvenile hormone insecticides during metamorphosis: how these compounds kill insects. J. Insect Physiol. 50, 111-121.


Figure 2.1 Relative expression of P450 transcripts in three strains of *Drosophila melanogaster*, wildtype, Oregon-R, mutated Met strain, Rst(1)JH¹ and the null Met mutant Met²⁷, in response to a single concentration of methoprene (3.2 ppm). Methoprene was administered in the diet from first instar until larvae were collected at late third instar. Total whole body RNA was used to make cDNA that was separately PCR amplified and analyzed as described in the Materials and Methods. The fold induction for each amplification after normalization to the constitutive actin transcript and in relation to Oregon-R control expression is shown within each column.
### TABLES

Table 2.1 Subfamily primer sequences used in RT-PCR reactions and amplified gene used for final probe in Southern hybridizations

<table>
<thead>
<tr>
<th>Primer name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sequence (5’-3’)</th>
<th>Sequenced product used as probe</th>
</tr>
</thead>
</table>
| Cyp4d-740                | GCAGAAACTGCCATGGG  
                          | CGTGGCCCTCGAACAT            | *Cyp4d1*                       |
| Cyp4d-1136               |                 |                                 |
| Cyp12a-873for            | GAATCCGTATCCGTGGTG  
                          | AGCGGTACCCCAATTCCAG          | *Cyp12a4/Cyp12a5*              |
| Cyp12a-1707rev           |                 |                                 |
| Cyp4ac-147for            | GGATGGCAGTCCATTGAAAA  
                          | GCAAAGGGATGACGATTAC          | *Cyp4ac1*                      |
| Cyp4ac-1342rev           |                 |                                 |

<sup>a</sup>Number in the primer names indicate the position within the mRNA transcript
Table 2.2 Gene specific primer sequences used in RT-PCR reactions and to amplify gene probe, the exonal position and PCR product sizes for mRNA transcripts and genomic DNA

<table>
<thead>
<tr>
<th>Primer name(^a)</th>
<th>Sequence (5’-3’)</th>
<th>Within exon</th>
<th>Size of mRNA PCR product (nt)</th>
<th>Size of genomic PCR product (nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp6a2-for</td>
<td>ACGTACGAATCCATCAAGGC ATGACCTGTGTGCCCCCTTC</td>
<td>1 2</td>
<td>226</td>
<td>295</td>
</tr>
<tr>
<td>Cyp6a2-rev</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp6g1-for</td>
<td>GACCATGTCTTTTGGCCCTGT CTCAAGTCTGCTGCTGCC</td>
<td>2 2</td>
<td>228</td>
<td>228</td>
</tr>
<tr>
<td>Cyp6g1-rev</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4e2-855</td>
<td>TCAAGATGAACAGCGAAAC ATCAGAGACCCAGTTCAGG</td>
<td>4 5</td>
<td>237</td>
<td>297</td>
</tr>
<tr>
<td>Cyp4e2-1091</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4g15-995</td>
<td>AGGATGACCTGGATGGAG TAGTTGCCCGAGTTCAAGCTT</td>
<td>4 5</td>
<td>394</td>
<td>490</td>
</tr>
<tr>
<td>Cyp4g15-1388</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp314a1-37</td>
<td>TCGTACTTGGCTGCTACGTGC TGTCAGGTTGCACTATGGGC</td>
<td>1 2</td>
<td>324</td>
<td>514</td>
</tr>
<tr>
<td>Cyp314a1-545</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp6a8-1111</td>
<td>CGACTGTACACCATTTGTGCC CAAATGGATTGGAGTTGAC</td>
<td>2 2</td>
<td>380</td>
<td>380</td>
</tr>
<tr>
<td>Cyp6a8-1490</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp315a1-1211</td>
<td>AAGGAGCGAGCTCCATAATGA GCAGACTGCGGCTGATTTA</td>
<td>5 6</td>
<td>268</td>
<td>323</td>
</tr>
<tr>
<td>Cyp315a1-1478</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp12d1-772</td>
<td>ATATCGACGCCCCACCTACAG GAGGCAAGTAGAACAACAGCCG</td>
<td>4 4</td>
<td>251</td>
<td>251</td>
</tr>
<tr>
<td>Cyp12d1-1022</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp6d4-for</td>
<td>GCATTCAGAAAGACACCCGTAT AATGGTTTCTGCACAAACA</td>
<td>2 3</td>
<td>263</td>
<td>323</td>
</tr>
<tr>
<td>Cyp6d4-rev</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp6d5-for</td>
<td>TGGGACAAGCTCTGACCTGAC TGTATTTTCCTGGCTTCCG</td>
<td>2 3</td>
<td>243</td>
<td>312</td>
</tr>
<tr>
<td>Cyp6d5-rev</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp6d2-for</td>
<td>GCACCGAAGATATCCGGGTA TTGGTCACAAACACAGGGGC</td>
<td>3 4</td>
<td>331</td>
<td>401</td>
</tr>
<tr>
<td>Cyp6d2-rev</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>actin</td>
<td>TGGGYGAYATGGAGAAGATCTGG TAGATGGGBCBGTGTGBAGACAA</td>
<td>3 4</td>
<td>263</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Number in the primer names indicate the position within the mRNA transcript
Table 2.3 Locations of putative sequence motifs in the upstream DNA of selected cytochrome P450 genes that are related to the Broad Z1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position (strand)</th>
<th>Core match</th>
<th>Sequence* (+ strand)</th>
<th>Factor name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp6a2</td>
<td>-834/-816 (+)</td>
<td>1.000</td>
<td>atgtcattaACAAattcat</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td></td>
<td>-598/-580 (+)</td>
<td>0.833</td>
<td>tcataaaaGCAAaagc</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td></td>
<td>-275/-257 (-)</td>
<td>1.000</td>
<td>aaaatataACAAatgata</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp12d1</td>
<td>-249/-231 (-)</td>
<td>1.000</td>
<td>aaaaaataAACAAatgaac</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp12a5</td>
<td>-390/-372 (+)</td>
<td>0.833</td>
<td>attgattaGCAAaactaa</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp314a1</td>
<td>-145/-127 (+)</td>
<td>1.000</td>
<td>aagtcacrAACAAattagt</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp6d4</td>
<td>-863/-845 (+)</td>
<td>0.800</td>
<td>atctaaaaCCAAaagtc</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td></td>
<td>-323/-305 (+)</td>
<td>1.000</td>
<td>taggtagaacCAAgattca</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td></td>
<td>-135/-117 (-)</td>
<td>1.000</td>
<td>ttattatatCAAAatcat</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp6g1</td>
<td>-322/-304 (+)</td>
<td>1.000</td>
<td>atgaaattcACAAatgc</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp4ac1</td>
<td>-765/-747 (-)</td>
<td>1.000</td>
<td>agattaataCAAAattct</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td></td>
<td>-33/-15 (-)</td>
<td>0.833</td>
<td>taatcataCACAAAattt</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td></td>
<td>+125/+143 (+)</td>
<td>1.000</td>
<td>aacacaattaCAAAaacct</td>
<td>BrCZ-1</td>
</tr>
<tr>
<td>Cyp4g15</td>
<td>+344/+362 (-)</td>
<td>1.000</td>
<td>aaatatattCAAAaattat</td>
<td>BrCZ-1</td>
</tr>
</tbody>
</table>

*Capital letters indicate the positions in the sequence which match the core sequence of the matrix. Bold letters indicate CI value >60.
Cyp12a4, Cyp4e2, Cyp4d1, Cyp6d5, Cyp315a1 and Cyp6a8 contain no putative BrC Z-1 motifs.
Table 2.4 Locations of putative sequence motifs in the upstream DNA of selected cytochrome P450 genes that are related to the bHLH-PAS transcriptional regulator cascade

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position (strand)</th>
<th>Core match</th>
<th>Sequence* (+ strand)</th>
<th>Factor name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp6a2</td>
<td>102-126 (+)</td>
<td>1.000</td>
<td>aatgcecgccccGTGcccaacatcgtg</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td>Cyp6a8</td>
<td>-551/-527 (-)</td>
<td>1.000</td>
<td>egeccttccttgcGTGaaccttetc</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td>Cyp12d1</td>
<td>-879/-685 (+)</td>
<td>0.870</td>
<td>atgctccacCTGcagt</td>
<td>bHLH protein (Dec2, Sharp2)</td>
</tr>
<tr>
<td></td>
<td>-256/-242 (-)</td>
<td>1.000</td>
<td>aaatggaGTGacg</td>
<td>ARNT homodimers</td>
</tr>
<tr>
<td></td>
<td>+75/+89 (+)</td>
<td>0.903</td>
<td>ttggcacaCATGaaca</td>
<td>bHLH protein (Dec2, Sharp2)</td>
</tr>
<tr>
<td>Cyp12a4</td>
<td>-1042/-1018 (+)</td>
<td>1.000</td>
<td>ttctctacGCTGcctgccataaa</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td></td>
<td>-1018/-879 (+)</td>
<td>0.870</td>
<td>aattacccacGCTGctaccetact</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td></td>
<td>-935/-921 (-)</td>
<td>1.000</td>
<td>aattcgtGCCCAAgctgg</td>
<td>bHLH heterodimers (HAND2/E12)</td>
</tr>
<tr>
<td></td>
<td>-682/-668 (+)</td>
<td>1.000</td>
<td>atttecaAGCTGgagtt</td>
<td>ARNT homodimers</td>
</tr>
<tr>
<td></td>
<td>-369/-345 (+)</td>
<td>1.000</td>
<td>attacattttgGTGacgtttta</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td></td>
<td>-358/-344 (-)</td>
<td>1.000</td>
<td>tttaaaaAGCTGcacc</td>
<td>ARNT homodimers</td>
</tr>
<tr>
<td></td>
<td>-30/-6 (+)</td>
<td>0.750</td>
<td>cacactgacGCTGTgtctgccatc</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td>Cyp315a1</td>
<td>-1017/-993 (+)</td>
<td>0.750</td>
<td>ccaagctctgGTGctccctgtgg</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td></td>
<td>-342/-318 (-)</td>
<td>1.000</td>
<td>gtccecaacacGCTGtactaaagttca</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td></td>
<td>-216/-192 (+)</td>
<td>1.000</td>
<td>aagtttacGCTGgctgccaa</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td>Cyp6d4</td>
<td>-504/-490 (-)</td>
<td>0.903</td>
<td>caatcacaGTGggga</td>
<td>bHLH protein (Dec2, Sharp2)</td>
</tr>
<tr>
<td>Cyp6g1</td>
<td>-483/-459 (+)</td>
<td>0.750</td>
<td>eetcctggaGCCTGctctatct</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td></td>
<td>-264/-240 (+)</td>
<td>1.000</td>
<td>ataattatagCTGcaataacttt</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td></td>
<td>-39/-25 (+)</td>
<td>0.870</td>
<td>cacegcaCTTGgacc</td>
<td>bHLH protein (Dec2, Sharp2)</td>
</tr>
<tr>
<td>Cyp4ac1</td>
<td>-422/-398 (+)</td>
<td>1.000</td>
<td>taaaaagtttcgGTGctggctaa</td>
<td>AhR/ARNT</td>
</tr>
<tr>
<td>Cyp4d1</td>
<td>-707/-683 (+)</td>
<td>0.750</td>
<td>cctcttgaCCCCTGctgctcactaa</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td></td>
<td>-672/-658 (-)</td>
<td>1.000</td>
<td>acagtgGCCActgg</td>
<td>bHLH heterodimers (HAND2/E12)</td>
</tr>
<tr>
<td></td>
<td>-595/-571 (-)</td>
<td>0.750</td>
<td>gctccttccCCCCTGtctgctcegge</td>
<td>AhR/ARNT (fixed core)</td>
</tr>
<tr>
<td>Cyp4g15</td>
<td>-817/-801 (+)</td>
<td>1.000</td>
<td>agcactgCCAGacttcca</td>
<td>Thing1/E74 heterodimer</td>
</tr>
</tbody>
</table>

*Capital letters indicate the positions in the sequence which match the core sequence of the matrix. Bold letters indicate CI value >60.

1Cyp12a5, Cyp4e2, and Cyp6d5 contain no putative bHLH-PAS binding motifs
CHAPTER 3

Cytochrome P450 expression associated with cross-tolerance to aflatoxin B1 in a methoprene tolerant strain

INTRODUCTION

Feeding animals must strike a balance between obtaining nutritious chemicals and avoiding toxic ones. For *Drosophila melanogaster*, the consummate model organism, it has long been recognized that yeasts are foods for *D. melanogaster* larvae (Baumberger, 1917) and these are now included in most standard rearing media for *D. melanogaster*. However, in nature larvae of *D. melanogaster* feeding on the yeasts found in rotting fruit face greater complexities in their diets, due to the presence of bacteria, fungi and plant allelochemicals. In this diverse and changing chemical milieu, *D. melanogaster* larvae must adapt toxicologically in order to survive.

Cytochrome P450 monooxygenases (P450s) play important roles in chemical homeostasis, as enzymes that both produce active endogenous chemicals (*i.e.*, ecdysteroids, fatty acids) and eliminate harmful chemicals (*i.e.*, toxic allelochemicals, insecticides). Diversified in their various eukaryotic hosts, P450s are characterized as heme-dependent enzymes that partner with NADPH-dependent P450 reductases to perform a variety of reactions including hydroxylations, epoxidations, and demethylations (Ortiz de Montellano, 2005; Sigel et al., 2007).

P450s are essential to development and reproduction in insects, because they catalyze steps in ecdysone and juvenile hormone biosyntheses (Ono et al., 2006; Rewitz et al., 2006a, 2006b, 2006c, 2007; Gilbert, 2004; Helvig et al., 2004; Niwa et al., 2004; Warren et al., 2004; Petryk et al., 2003; Warren et al., 2002; Sutherland et al., 2000, Andersen et al., 1997). P450s are also documented in their mechanistic role in biochemically-based insecticide resistance as phase I detoxification enzymes, along with esterases, and phase II glutathione S-transferases (Feyereisen, 1999, 2005; Hemingway, 2000; Ranson et al., 2002; Ranson and Hemingway, 2005;
Li et al., 2007). P450-mediated detoxification is a major mechanism of resistance to the pyrethroids cypermethrin and deltamethrin in *Musca domestica* (house fly) (Korytko and Scott, 1998; Scott, 1998), to cypermethrin, diazinon, chlorpyrifos and fenvalerate in *Helicoverpa zea* (corn earworm) (Li et al., 2000, 2004; Rupasinghe et al., 2007), and to carbaryl, cypermethrin, permethrin, deltamethrin, and DDT in *Anopheles* species (Hemingway and Ranson, 2000; Corbel et al., 2007; Boonsuepsakul et al., 2008; Chiu et al., 2008; McLaughlin et al., 2008; Muller et al., 2008; Wen et al., 2010).

In addition to insecticides, insect P450s metabolize natural xenobiotic compounds that are encountered in the insect diet and environment during feeding. Many examples now exist of lepidopteran P450s essential in the detoxification of plant allelochemicals found in the diet. In *Papilio polyxenes* (black swallowtail) and *Papilio glaucus* (tiger swallowtail), CYP6B1 and CYP6B4, respectively, are able to metabolize the linear furanocoumarin xanthotoxin and are inducible by a range of linear and angular furanocoumarins (Hung et al., 1997; Li et al., 2001, 2002a, 2003). In the related species, *Papilio multicaudatus* (two-tailed swallowtail), CYP6B33 and multiple related cDNAs efficiently metabolize at least two angular furanocoumarins and four linear furanocoumarins at intermediate efficiencies (Mao et al., 2007b, 2008a). CYP6AB3, from the highly specialized *Depressaria pastinacella* (parsnip webworm), selectively metabolizes imperatorin, the most abundant furanocoumarin in its two hostplant genera (Li et al., 2004b; Mao et al., 2006, 2007a), and myristicin, a naturally occurring methylenedioxyphenyl (MDP) compound (Mao et al., 2006).

Like CYP6B1, CYP6B4 and CYP6B33, the polyphagous *H. zea* CYP6B8 metabolizes xanthotoxin (Li et al., 2004a, 2007; Rupasinghe et al., 2007), but it also metabolizes a wide variety of unrelated plant compounds (flavone, chlorogenic acid, indole-3-carbinol, and
quercetin) as well as some pyrethroid (cypermethrin), organophosphate (diazinon) and cyclodiene (aldrin) insecticides (Li et al., 2004a; Rupasinghe et al., 2007). CYP321A1, also from *H. zea*, metabolizes plant compounds (xanthotoxin, angelicin), insecticides (cypermethrin, diazinon, and aldrin) and a fungal toxin (aflatoxin B1 (AFB1)) (Sasabe et al., 2004, Rupasinghe et al., 2007; Niu et al., 2008). Thus, the substrate specificity of inducible P450 enzymes reflects the breadth of diet of the insect.

Aflatoxin B1, a toxic metabolite produced by *Aspergillus* fungi that contaminate agricultural crops (Robens and Cardwell, 2005), is both detoxified and bioactivated by P450s. The existence of these counteracting effects is currently best exemplified in the larvae of *H. zea*, a generalist insect. When these generalist larvae are treated with piperonyl butoxide (PBO), a general P450 inhibitor, their development progresses better on aflatoxin B1 (Zeng et al., 2006). When natural inducers of cytochrome P450s (xanthotoxin, coumarin, and indole 3-carbinol) are added to the diet, AFB1-metabolizing P450s are induced (Zeng et al., 2007). CYP321A1, one of the induced P450s, has been shown through heterologous expression to metabolize aflatoxin B1 to its less toxic metabolite, aflatoxin P1 (Niu et al., 2008).

AFB1 resistance varies across wild-type strains of *D. melanogaster* (Chinnici and Melone, 1985), as do the AFB1 metabolites produced (Foerster and Wurgler, 1984). Experiments with laboratory strains demonstrated that AFB1 in the diet increases larval mortality and decreases adult body length (Chinnici and Bettinger, 1984; Llewellyn and Chinnici, 1978). It also acts as a chemosterilant of *D. melanogaster* and *M. domestica* adults (Matsumura and Knight, 1976). Histological evidence indicates that AFB1 suppresses the activity of neurosecretory cells in the larval ring gland (Chinnici and Llewellyn, 1979).
AFB1 resistance levels are higher in *D. melanogaster* collected from agricultural areas contaminated with AFB1 than those taken from uncontaminated areas (Delawder and Chinnici, 1983). Genetic analysis has indicated that this naturally acquired tolerance to AFB1 is due to factors on chromosome X and chromosome 2 (Chinnici and Melone, 1985; Kirk et al., 1971). Heterologous expression in yeast has shown that CYP6A2, one of several P450s located on chromosome 2, produces a bioactivated form of AFB1 (Saner et al., 1996). Despite the ability of CYP6A2 to activate this toxin, *Cyp6a2* expression occurs in the midgut and Malphigian tubules of late third instars and continues until the wandering stage (Chung et al., 2006). Reducing *Cyp6a2* expression in the larval stage would presumably produce higher levels of AFB1 tolerance. Alternatively, the induction of an AFB1-metabolizing P450 could remediate CYP6A2-dependent AFB1 bioactivation by eliminating the substrate before it accumulates. Natural variation in the expression of CYP6A2 and other AFB1-inducible P450s may contribute to the observed variation in tolerance to AFB1.

In earlier P450 expression studies, *Cyp6a2* transcript repression was shown to occur in Rst(1)JH¹, a strain that had been selected for its high tolerance to the insecticide methoprene (Shemshedini et al., 1990). The primary mechanism of methoprene tolerance in Rst(1)JH¹ is target site inactivation of the Met protein, a transcription factor in the basic helix-loop-helix PAS family of proteins (Ashok et al., 1998; Wilson et al., 2006). While the juvenile hormone-mimicking properties of methoprene have long been documented (Staal, 1975), its molecular mode of action is not completely understood. The Met protein, which can bind methoprene, may exert its effects by interacting with other transcription factors, the Broad Complex (BRC) (Restifo and Wilson, 1998; Wilson et al., 2006) or the ecdysone receptor complex with ultraspiracle (EcR/USP) (Bitra and Palli, 2009).
The suppression of \textit{Cyp6a2} expression in Rst(1)JH\textsuperscript{1} implicates CYP6A2 in the toxicity of methoprene, in a manner similar to its role in the bioactivation of AFB1. Beyond the bioactivation of AFB1, the known catalytic functions of CYP6A2 are limited to insecticide metabolism (Dunkov et al., 1997). Because the related CYP6A1, from \textit{Musca domestica} (49% amino acid identity), has been shown to mediate the epoxidation of both methyl farnesoate and JHIII to produce active products (Andersen et al., 1997), it remains possible that CYP6A2 bioactivates juvenile hormone and/or methoprene.

To determine if CYP6A2 bioactivates AFB1, we compared the tolerance of the Rst(1)JH\textsuperscript{1} \textit{Cyp6a2} knockdown strain to the progenitor methoprene-susceptible Oregon-R strain. We also compared the expression levels of four P450 transcripts (\textit{Cyp6a2}, \textit{Cyp6d4}, \textit{Cyp6d5}, \textit{Cyp6d2}) that are normally expressed in third instar larvae (Gauhar et al., 2008). Of these, \textit{Cyp6d4} and \textit{Cyp6d5} are expressed in the midgut (Chung et al., 2009) and induced by xenobiotics (Le Goff et al., 2006; Willoughby et al., 2006, 2007). Expression of \textit{Cyp6d2}, a related P450, is restricted to the tracheae (Chung et al., 2009). We also measured their transcriptional responses to methoprene to determine whether molecular responses selected for methoprene tolerance lead to AFB1 cross-tolerance.

\textbf{MATERIALS AND METHODS}

\textit{Chemicals and Enzymes}

AFB1 was obtained from Sigma (St. Louis, MO). Analytical grade dimethyl sulfoxide (DMSO) and methanol were purchased from Fisher Scientific (Pittsburgh, PA). Methoprene (98% mix of isomers) was purchased from Chem Service (West Chester, PA). The RNAeasy RNA isolation kit was from Qiagen (Valencia, CA). Superscriptase III reverse transcriptase, \textit{Taq} polymerase and deoxynucleotide triphosphates (dNTPs) were from Invitrogen (Carlsbad, CA).
Insects

*D. melanogaster* strains, Oregon R and Rst(1)JH$^1$ were obtained from Bloomington *Drosophila* Stock Center (Bloomington, IN). Two vials of the strain Rst(1)JH$^1$ were obtained and progeny from these vials were reared independently. Flies were reared in plastic bottles or vials on standard diet (Applied Scientific, San Francisco, CA). Flies were transferred to new bottles every two weeks.

Bioassays

For AFB1 bioassays, stock solutions of AFB1 were prepared in DMSO at concentrations of 1 mg/mL and added to standard agar-cornmeal diet at final concentrations of 0, 0.5, 1, and 2 ppm (assay 1) or 0, 0.2, 0.6, 0.8 and 1 ppm (assay 2). A total volume of 10 ul of DMSO was used for each vial across treatments. To determine the toxicity of AFB1, 15 first instars of the two strains, Oregon-R and Rst(1)JH$^1$, were placed in vials containing the standard diet and varying levels of AFB1. Insects were then reared in an insectary kept at 23-26°C with a photoperiod of 12 hr light/12 hr dark. The number of larvae that pupated, days to pupation and the number of adults emerging from pupation were recorded for each vial. Two independent assays were performed with each treatment replicated 6 (assay 1) and 5 (assay 2) times.

To assess the effects of AFB$_1$ exposure at different developmental time points, 15 first instars (0-24 hr) of Oregon-R and Rst(1)JH$^1$ were placed on vials containing AFB$_1$ at concentrations of 0, 250 and 500 ppb and third instars were collected 96 hr and 104 hr later. Larvae from two independent vials were pooled for RNA isolation and two biological replicates were conducted.

For methoprene bioassays, a stock solution of methoprene was prepared in methanol at a concentration of 10 mg/ml and added to standard cornmeal-agar diet for a final concentration of
0, 5, 10, or 50 ppm. A total volume of 10 µl of methanol was used for each vial across treatments. To confirm differences in the toxicity of methoprene to two strains of *D. melanogaster*, Oregon-R and Rst(1)JH\(^1\), 15 first instars were placed in vials containing the standard diet and varying levels of methoprene. Insects were reared at a temperature of 23-26°C with a photoperiod of 16 hr light/8 hr dark. The number of larvae that pupated, days to pupation and the number of adults emerging from pupation were recorded for each vial. Each treatment was replicated seven times.

**Statistical Analysis**

Data from both assays were analyzed by nested one-way ANOVA (PASW 17.0, SPSS, Inc., Chicago, IL). AFB\(_1\) bioassays were analyzed for significance by a combined probability test.

**RNA isolations and RT-PCR gel blot analyses**

Third instar larvae were flash-frozen in liquid nitrogen. RNA was isolated using the RNAeasy RNA isolation kit and its on-column DNase treatment as described by the manufacturer. First-strand cDNA was synthesized using the SuperScript® III first-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA) by first incubating 1 µg of RNA in a 13 µL reaction containing 500 ng of oligo (dT)\(_{17}\), and 0.5 mM of each dNTP at 65°C for 5 min and then on ice for 1 min. To this reaction 4 µL of 5X first-strand buffer, 1 µL of 0.1 M DTT, 1 µL of rRNAsin (40 units/µL) (Promega) and 1 µL of SuperScript III reverse transcriptase (200 units/µL) were added and the reaction was incubated at 50°C for an additional 60 min before heat-inactivating at 70°C for 15 min.

P450 and actin mRNA levels were quantified by RT-PCR gel blot analyses in separate 50 µL PCR reactions containing 1 µL first strand cDNA, 1X PCR buffer minus Mg (Invitrogen), 0.2
mM each dNTP, 1.5 mM MgCl₂, 2.5 U Taq DNA polymerase (Invitrogen) and 20 pmol each of forward and reverse primer for individual P450 or actin genes (Table 1). Control experiments determined that 18 PCR cycles for actin, 21 cycles for Cyp6d2, 25 cycles for Cyp6a2, 28 cycles for Cyp6d4 and Cyp6d5 were within the linear range of amplification for total RNA from whole larval bodies.

For quantification of P450 and actin mRNAs, ten microliters of the RT-PCR amplified products were run on 1.5% 1X Tris-borate-EDTA (TBE) gels at 150 V for 30 min. RT-PCR Southern gels were denatured by soaking in 0.5 M NaOH denaturing solution for 15 min and in 1 M Tris (pH 7.4) neutralizing solution for 15 min. Denatured RT-PCR products were transferred to Hybond-N nylon membranes overnight by capillary action and the actin blots were probed with a ³²P-labeled 200 bp cDNA fragment of D. melanogaster actin 5C gene. Hybridizations were done at high stringency in 50% formamide, 5X SSC, 50 mM Na₂HPO₄ (pH 7.0), 0.5% SDS, 5X Denhardt’s solution at 42°C overnight. All Southern blots were washed at low stringency twice in 2X SSC, 0.1% SDS at 42°C for 15 min. Hybridized blots were quantified by Image J digital analysis (1.38X, Wayne Rasband, NIH, USA) of autoradiograms. Values for P450 transcripts were corrected by background readings from the blots and normalized against constitutively expressed actin in each sample.

**Sequencing of Cyp6a2 promoters**

Genomic DNA was isolated using a modified protocol from Sambrook et al. (1989). Fifteen adults were homogenized with micropestles in 500 µL extraction buffer (10 mM Tris-HCl (pH 8.0), 0.1 M EDTA (pH 8.0), 20 µg/mL RNase A and 0.5% SDS) containing 100 µg/mL proteinase K were incubated overnight at 50°C. DNA was then extracted using 1 volume of phenol/chloroform (1:1 v/v) and precipitated with 1/20 volume of 2 M NaCl and 2 volumes of
Upstream promoter regions were PCR amplified from genomic DNAs of the Oregon-R and Rst(1)JH1 strains using a reverse primer (CYP6A2+20BglII, 5’-GGGAGATCTCGTAGCTGCTCTTTTCG-3’) and a forward primer (CYP6A2-1340KpnI, 5’-GGGTACCATCCGAACACTGCAG-3’). PCR amplification was performed in a 50 µL reaction with 1 µL Pfu polymerase (2.5 U/µL) (Stratagene, La Jolla, CA), 5 µL 10X Pfu reaction buffer (with 20mM MgCl2), and 1 µL dNTPs (10 mM each NTP). Reactions were denatured at 94°C for 5 min and then subjected to 35 cycles of denaturation at 94°C for 5 min, annealing at 55°C for 1 min and elongation at 70°C for 10 min, followed by a final 70°C step for 30 min.

Nucleotide positions in the proximal promoter are reported relative to the transcriptional start site indicated by +1, with upstream sequences preceded by “-“ and downstream sequences preceded by “+”.

The reverse primer was engineered with a BglII restriction enzyme site and the forward primer was engineered with a KpnI restriction enzyme site for potentially cloning into the pGL3-basic plasmid (Promega, Madison, WI). PCR products were separated by gel electrophoresis and those at 1.3 kb were extracted using Qiaquick Gel Extract kit (Qiagen, Inc.) and directly sequenced using the original CYP6A2+20BglII and CYP6A2-1340KpnI primers as well as internal primers (CYP6A2-100rev, 5’-ACAGCATGTGAGCTAG-3’; CYP6A2-183rev, 5’-CGGCGATAATATGCGAGGCAC-3’, CYP6A2-1200for, 5’-TAACACTACGCGACATTGC-3’) complementary to regions starting at 100, 183 and 1200 upstream of the transcription start site.

**Alignment of CYP6A1 and CYP6A2 with putative substrate recognition sites (SRS) identified**

An amino acid alignment of CYP6A1 and CYP6A2 was performed using BLOSUM50 matrix and gap opening and extension penalties of -12 and -2 in the LALIGN program (Pearson, 1990) in Biology Workbench ([http://workbench.sdsc.edu/](http://workbench.sdsc.edu/)). Putative substrate recognition sites were identified based on domains introduced by Gotoh (1992).
Molecular modeling and substrate docking

The CYP6A1 (M. domestica) and CYP6A2 (D. melanogaster) protein structures were predicted using MOE programs (Chemical Computing Group, Inc., Montreal, Canada) as previously described in Rupasinghe et al. (2003) and Mao et al. (2006). The protein sequences of CYP6A1 and CYP6A2 were aligned with all CYP6A protein sequences from Drosophila melanogaster (CYP6A8, CYP6A9, CYP6A13, CYP6A14, CYP6A17, CYP6A18, CYP6A19, CYP6A20, CYP6A21, CYP6A22, and CYP6A23). This alignment was then aligned to 10 templates including CYP1A2 (2HI4; Sansen et al., 2007), CYP2A6 (1Z10; Yano et al., 2005), CYP2A13 (2P85; Smith et al., 2007), CYP2B4 (1SUO; Scott et al., 2004), CYP2C5 (1N6B; Wester et al., 2003), CYP2C8 (1PZ2; Schoch et al., 2004), CYP2C9 (1OG5; Williams et al., 2003), CYP2D6 (2F9Q; Rowland et al., 2006), CYP2R1 (2OJD; Strushkevich et al., 2008), and CYP3A4 (1TQN; Yano et al., 2004). Three variable regions were independently aligned to the crystal structure of CYP2C9 as described in Baudry et al. (2006), the B region that includes SRS1, the FG region that includes SRS2 and SRS3 and the Beta-4 region that includes SRS6. After energy minimization with CHARMM22 force field (MacKerell et al., 1998), the model with the best packing score and structure free of aberrations was selected for substrate docking experiments.

The energy-minimized substrate-free protein structure was docked with the substrate molecules, aflatoxin B1 and methoprene, using the Monte-Carlo docking procedure of MOE with the MMFF94s force field (Halgren, 1996) for the oxygen-free heme as distributed in MOE 2009.1. The substrate was positioned in a potential active site above the heme in preparation for Monte-Carlo simulations. Hundreds of possible conformations were generated and ranked according to the overall internal energy of the ligand. The lowest energy conformation was
selected and then included in the protein for a protein/ligand minimization in which the heme coordinates were fixed in order to avoid distortion of the heme plane due to the lack of bonded parameters for the heme in the MMFF94s force field.

RESULTS

Cross-tolerance to methoprene and aflatoxin B₁ established in Rst(1)JH₁

In preliminary analyses of the gene expression (Chapter 2), the methoprene-tolerant strain demonstrated decreased expression of Cyp6a2, compared to its wildtype progenitor strain, Oregon-R. This result led me to examine the possible benefits incurred in Rst(1)JH₁ via this knockdown of expression, because the CYP6A2 enzyme is implicated in the bioactivation of AFB1. First instars were raised on diet containing 0, 0.5, 1 and 2 ppm AFB1 or 0, 0.2, 0.6, 0.8 and 1 ppm AFB1 and larval development was monitored until pupation. When compared to the Oregon-R strain, the Rst(1)JH₁ strain demonstrated faster larval development on AFB1 (Fig. 3.1) and higher emergence rates of Rst(1)JH₁ adults on AFB1 (Fig. 3.2).

P450 expression in Oregon-R and Rst(1)JH₁

Expression of Cyp6a2 in late third instars of two populations of the methoprene-tolerant strains remained low or negligible in the presence of AFB1. In Oregon-R, Cyp6a2 expression remained constant with increasing doses of aflatoxin B₁. (Fig. 3.3) Cyp6d2 expression was induced by aflatoxin B₁, starting at the lowest dose (0.5 ppm), in both Oregon-R and Rst(1)JH₁ (Fig. 3.3). Expression of Cyp6d4 and Cyp6d5 did not change in response to aflatoxin B₁ (Fig. 3.4). I conclude that the constitutive levels of Cyp6a2 are lower in the methoprene-tolerant strain and that AFB1 induces Cyp6d2 in both strains.
To determine whether AFB1 inducibility changes during larval development, I measured Cyp6d2 expression at two time points during third instar. At late third instar, constitutive Cyp6d2 expression was lower in both strains than 8 h earlier in third instars. In Oregon-R, both concentrations, 250 and 500 ppb, of AFB1 induced Cyp6d2 expression at both time points, 32-fold in late third instars and 2-3 fold in early third instars (Fig 3.5). In Rst(1)JH¹, both concentrations of AFB1 induced Cyp6d2 expression 11-fold in late third instars, but only the highest concentration, 500 ppb, induced Cyp6d2 expression 2-fold in early third instars (Fig. 3.5).

To determine if cytochrome P450 expression levels in D. melanogaster correlate with methoprene tolerance, the levels of the Cyp6a2, Cyp6d2, Cyp6d4 and Cyp6d5 transcripts were also measured in the presence of increasing concentrations of methoprene. At 5 and 10 ppm methoprene, adult mortality is significantly higher in Oregon-R than Rst(1)JH¹, whereas at the highest dose, 50 ppm, mortality between Rst(1)JH¹ and Oregon-R does not significantly differ (Fig. 3.6). Also, at 5 and 10 ppm methoprene, larval development time did not differ between Oregon-R and Rst(1)JH¹ (Fig. 3.7). Only at the highest dose tested (50 ppm) does methoprene seem to affect development time differently between the wildtype and methoprene-tolerant strains. In Oregon-R, expression of Cyp6a2 was two-fold higher in the presence of methoprene, although a strong dose-dependent correlation was not observed (Fig 3.8). In Rst(1)JH¹, expression of Cyp6a2, as with aflatoxin B₁, was negligible in the presence of methoprene. Constitutive expression of Cyp6d2 was higher in Rst(1)JH¹ than Oregon-R, but Cyp6d2, Cyp6d4 and Cyp6d5 were not induced by methoprene in either strain (Fig. 3.8, 3.9).

The upstream regulatory regions of Cyp6a2 from the two strains were sequenced in order to identify differences that may lead to the observed differential expression patterns. In the
promoters of *Cyp6a2*, the Rst(1)JH^1^ strain shows a 15-nucleotide deletion when compared to the promoters of Oregon-R and three additional strains, rosy^{506} (AF061081.1), 91-R (AF061082.1), and the reference genome, y; *cn bw sp* (Figure 3.10).

To ascertain what functional correlations can be drawn about CYP6A2 based on CYP6A1, which shares 49% overall sequence similarity, the putative substrate recognition sites were compared in an amino acid sequence alignment. CYP6A2 and CYP6A1 sequence share 45% in SRS1, 43% in SRS2, 44% in SRS3, 76% in SRS4, 40% in SRS5, and 20% in SRS6 (Fig. 3.11). Molecular models of the two sequences were also compared for substrate binding in their catalytic site. CYP6A1 and CYP6A2 show marked differences in their potential for the epoxidation or demethylation of methoprene and juvenile hormone (Table 3.1). Based on these models, CYP6A2 does not seem capable of binding methoprene, but it does have the potential to perform the epoxidation of AFB1, which fits within 6 Angstroms of the heme (Table 3.1).

**DISCUSSION**

Selection for methoprene tolerance, which produced the Rst(1)JH^1^ strain, has resulted in cross-tolerance to aflatoxin B1. Rst(1)JH^1^, a strain of *Drosophila melanogaster* selected for methoprene tolerance, demonstrates increased tolerance of aflatoxin B1 during larval development over its progenitor strain, Oregon-R. This finding raises the question of whether there are shared genetic mechanisms between tolerance to methoprene and aflatoxin B1.

In this study, we compared cytochrome P450 expression between Rst(1)JH^1^ and Oregon-R, because cytochrome P450s are implicated in the metabolism of aflatoxin B1 (Saner et al., 1996; Niu et al., 2008) and methoprene (Quistad et al. 1974; Terriere and Yu, 1973; Hammock et al. 1977) in insects. We identified differentially expressed P450 genes that help to build a model of aflatoxin B1 and methoprene toxicity in *Drosophila melanogaster*. 

50
That $Cyp6a2$ expression is knocked out in the Rst(1)JH$^1$ strain, even in the presence of aflatoxin B$_1$, supports the conclusion by Saner et al. (1996), that CYP6A2 metabolizes aflatoxin B$_1$ into its toxic byproduct. The simultaneous knockdown of $Cyp6a2$ in Rst(1)JH$^1$ and induction in Oregon-R, in the presence of methoprene, evokes a model of bioactivation for methoprene, similar to that for aflatoxin B1. While the mode of toxicity of methoprene is not understood, a bioactivation model proposes that methoprene must be bioactivated, perhaps by CYP6A2, into a toxic form that disrupts normal endocrine function.

In adults, the CYP6A2 protein is considered a detoxification enzyme, metabolizing a range of insecticides (Dunkov et al., 1997). In larvae, the role of CYP6A2 has not been fully explored. A homologous protein (48.9% identity) in *Musca domestica*, CYP6A1, was characterized as an insecticide-metabolizing P450 (Andersen et al., 1994), but further biochemical analysis revealed that CYP6A1 epoxidizes the final two products in the biosynthesis of juvenile hormone, although not methoprene (Andersen et al., 1997). While only biochemical expression can confirm whether CYP6A2 metabolizes methoprene or juvenile hormone, this work supports the hypothesis that CYP6A2 could metabolize methoprene into a toxic form.

This model of bioactivation by CYP6A2 reveals the possible target site of toxicity for aflatoxin B$_1$ and methoprene. In order to address the natural mechanisms of detoxification available to *Drosophila melanogaster*, we compared the expression of $Cyp6d2$, $Cyp6d4$ and $Cyp6d5$ between strains. The inducibility of $Cyp6d2$ expression by the natural xenobiotic aflatoxin B1 in both strains implicates this P450 as a possible detoxification enzyme, which has evolved to respond to aflatoxin B$_1$ in the diet. While $Cyp6d4$ and $Cyp6d5$ are induced by a range of xenobiotics, black pepper extract (Jensen et al., 2006) phenobarbital (Le Goff et al., 2006; Willoughby et al., 2006; Wen et al., 2006), caffeine (Willoughby et al., 2006), DDT (Willoughby
et al., 2006), and atrazine (Le Goff et al., 2006)), they do not seem to be involved in the response to aflatoxin B₁ or methoprene.

To understand the transcriptional regulation of the differentially expressed cytochrome P450 gene Cyp6a2 between the two strains, the upstream regulatory regions of Cyp6a2 were sequenced and scanned for putative regulatory elements. At -161 upstream of the transcription start site of Cyp6a2 in the Rst(1)JH¹ strain, a 15 nucleotide deleted region contains a putative Ap-1 site. This site, along with multiple other Ap-1 sites, were identified by Bhaskara et al. (2006) as necessary for both basal activity and caffeine inducibility of the Cyp6a2 and Cyp6a8 promoters. The 15-nt deletion eliminates one of two putative Ap-1 sites found in this region, which may disrupt binding of the Ap-1 transcription factor complex or change the orientation of the flanking binding sites by a one-half of a helical turn. Given the reduced constitutive expression of Cyp6a2 in Rst(1)JH₁, the Ap-1 site seems to be necessary for normal expression.

The D-FOS/D-JUN heterodimer has been shown to activate Ap-1 binding sites, but in vivo they have been described as a generic switch, involved in both activation and repression, depending on the surrounding proteins (Kockel et al., 2001). Experimental evidence from the Cyp6a8 promoter has demonstrated that overexpression of D-JUN protein, in Drosophila cell culture, inhibits Cyp6a8 promoter expression (Bhaskara et al., 2008). Thus, further investigation of the Rst(1)JH¹ Cyp6a2 promoter is needed to confirm the activational role of the deleted Ap-1 site.

I cannot exclude the possibility that the Met protein, which in the Rst(1)JH¹ strain can bind ligand but not DNA, also plays a role in the cross-tolerance to aflatoxin B₁. It remains possible that the Met protein is involved in the regulation of the cytochrome P450 genes studied here, yet the regulatory function of the Met protein is not fully understood. Null mutants for the met protein are viable; thus, its function could be phenocopied by another related protein, such as
its heterodimer partner, gce. Genetic and biochemical analyses suggest that Met/gce interacts with the Broad protein (isoform 1), while hybrid pulldown assays suggest a secondary role in the ecdysone/ultraspiracle complex (Bitra and Palli, 2009).

Thus, I propose that Cyp6a2 and Cyp6d2 act in concert when aflatoxin B1 is found in the diet. Cyp6a2, which is constitutively expressed in the larval stage of D. melanogaster, becomes a liability when aflatoxin B1 is present, perhaps because aflatoxin B1 mimics an endogenous substrate of CYP6A2. To defend against this bioactivated toxicity, Cyp6d2, would be induced by and metabolize aflatoxin B1, thereby averting further toxicity. Molecular modeling of CYP6A2 supports its role in the epoxidation of aflatoxin B1. For methoprene, underexpression of Cyp6a2 may have contributed to methoprene tolerance observed in the Rst(1)JH1 strain, due to a similar bioactivation function for methoprene, although the molecular model of CYP6A2 does not support a direct role in the epoxidation of this compound. In this way, selection for methoprene tolerance has lead to cross-tolerance to aflatoxin B1, demonstrating how two seemingly different compounds may exert similar selection pressures on the same cytochrome P450 gene Cyp6a2, thus offering insight into the mechanism and possible endogenous substrate of the enzyme.

ACKNOWLEDGMENTS

I thank Dr. Guodong Niu for his instruction in the use of aflatoxin B1 and Dr. Art Zangerl for his assistance with statistical analysis.

LITERATURE CITED


FIGURES

A

B
Figure 3.1 Larval development time in days for *Drosophila melanogaster* strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH1. Larvae were reared from first instar on diet containing either 0.2% DMSO (0 ppm) or 0.5, 1 or 2 ppm aflatoxin B1 (AFB1) (A) and 0.2% DMSO (0 ppm) or 0.2, 0.6, 0.8 and 1 ppm AFB1. Values for days to pupation are means ± standard deviation from 6 (A) and 5 (B) replicates with 15 larvae/vial. One-way nested ANOVA showed an effect of treatment by strain in B (p=0.0134) but not in A (p=0.06). A combined probability test for these two experiments confirmed a significant (α=0.05) interaction effect of treatment by strain.
Figure 3.2 Adult emergence, presented as the number of adults that survive from those that pupate, for *Drosophila melanogaster* strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH1. Larvae were reared from first instar on diet containing either 0.2% DMSO (0 ppm) or 0.5, 1 or 2 ppm aflatoxin B1 (AFB1). Values for adult emergence are means ± standard deviation from 6 replicates with 15 larvae/vial. Univariate ANOVA showed a significant effect of treatment (p=.041) and treatment by strain (p=.024) on the number of emerging adults.
Figure 3.3 P450 transcripts (Cyp6a2 & Cyp6d2) expressed in response to aflatoxin B1. First instar larvae of *Drosophila melanogaster* strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH$^1$ were reared on diet containing either 0.2% DMSO (0 ppm) or 0.5, 1 or 2 ppm aflatoxin B$_1$ (AFB1). Total whole body RNA from each treatment was used to make cDNA that were separately amplified by RT-PCR for each gene and analyzed as described in the Materials and Methods. The fold induction for each RT-PCR amplification is shown compared to Oregon-R control (above) and to control treatment within a strain (below) after normalization to the constitutive actin transcript.
Figure 3.4 P450 transcripts (Cyp6d4 & Cyp6d5) expressed in response to aflatoxin B1. First instar larvae of Drosophila melanogaster strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH¹ were reared on diet containing either 0.2% DMSO (0 ppm) or 0.5, 1 or 2 ppm aflatoxin B₁ (AFB₁). Total whole body RNA from each treatment was used to make cDNA that were separately amplified by RT-PCR for each gene and analyzed as described in the Materials and Methods. The fold induction for each RT-PCR amplification is shown compared to Oregon-R control (above) and to control treatment within a strain (below) after normalization to the constitutive actin transcript.
Figure 3.5 P450 transcripts (*Cyp6d2*) expressed in response to aflatoxin B1. First instar larvae of *Drosophila melanogaster* strains Oregon-R and the methoprene-tolerant strain, labeled Rst3472, were reared on diet containing either 0.05% DMSO (0 ppb), 250 ppb or 500 ppb aflatoxin B1 (AFB1). Third instars were collected at two time points, separated by 8 hr. Total whole body RNA from each treatment was used to make cDNA that were separately amplified by RT-PCR for each gene and analyzed as described in the Materials and Methods. The fold induction for each RT-PCR amplification is shown compared to Oregon-R control (above) and to control treatment within a strain (below) after normalization to the constitutive actin transcript.
Figure 3.6 Adult emergence, presented as the number of adults that survive from those that pupate, for Drosophila melanogaster strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH¹. Larvae were reared from first instar on diet containing 0.5% methanol (0 ppm) or 5, 10 or 50 ppm methoprene. Values for adult emergence are means ± standard deviation from 7 replicates with 15 larvae/vial. One-way nested ANOVA showed a significant effect of treatment by strain (p<.0001).
Figure 3.7 Larval development time in days for *Drosophila melanogaster* strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH\(^1\). Larvae were reared from first instar on diet containing 0.5% methanol (0 ppm) or 5, 10 or 50 ppm methoprene. Values for larval development time are means ± standard deviation from 7 replicates with 15 larvae/vial. One-way nested ANOVA showed no significant effect of treatment by strain (p=.122).
Figure 3.8 P450 transcripts (Cyp6a2, Cyp6d2) expressed in response to methoprene. First instar larvae of *Drosophila melanogaster* strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH$^1$ were reared on diet containing either 0.5% methanol (0 ppm) or 5, 10 or 50 ppm methoprene. Total whole body RNA from each treatment was used to make cDNA that were separately amplified by RT-PCR for each gene and analyzed as described in the Materials and Methods. The fold induction for each RT-PCR amplification is shown compared to Oregon-R control (above) and to control treatment within a strain (below) after normalization to the constitutive actin transcript.
Figure 3.9 P450 transcripts (Cyp6d2, Cyp6d5) expressed in response to methoprene. First instar larvae of *Drosophila melanogaster* strains Oregon-R and two populations of the methoprene-tolerant strain, labeled Rst3472 and Rst(1)JH1 were reared on diet containing either 0.5% methanol (0 ppm) or 5, 10 or 50 ppm methoprene. Total whole body RNA from each treatment was used to make cDNA that were separately amplified by RT-PCR for each gene and analyzed as described in the Materials and Methods. The fold induction for each RT-PCR amplification is shown compared to Oregon-R control (above) and to control treatment within a strain (below) after normalization to the constitutive actin transcript.
Figure 3.10 Alignment of *Cyp6a2* upstream regulatory region from Rst(1)JH1 with the reference strain (y; cn bw sp), rosy506 and 91-R, showing a 15 nucleotide deletion from the Rst(1)JH1 promoter. Within this region lies a putative Ap-1 site, GCGTAGTCATG (highlighted in purple), within a region of the promoter necessary for basal expression according to Bhaskara et al., (2006).
Figure 3.11 Amino acid alignment of two CYP6A2 variants CYP6A2vSVL and CYP6A2vMIDT (from Rst(1)JH\textsuperscript{1} strain) with CYP6A1 protein. The predicted substrate recognition sites (SRS) are denoted in red. Conserved amino acids amongs the three variants are denoted in blue. Overall CYP6A1 and CYP6A2 share 49% sequence identity. Within the predicted SRS domains, they share the following: SRS1 (9/20), SRS2 (3/7), SRS3 (4/9), SRS4 (13/17), SRS5 (6/15) and SRS6 (2/10).
**TABLES**

Table 3.1 Minimization energies and distances to the heme for substrates based on molecular modeling of CYP6A2 and CYP6A1

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Reaction</th>
<th>CYP6A2</th>
<th>CYP6A1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Potential Energy</td>
<td>Distance to heme (Angstroms)</td>
</tr>
<tr>
<td>JHIII</td>
<td>Epoxidation</td>
<td>-59.4</td>
<td>6.93</td>
</tr>
<tr>
<td>Methyl farnesoate</td>
<td>Epoxidation</td>
<td>-47.2</td>
<td>6.73</td>
</tr>
<tr>
<td>Methoprene</td>
<td>Epoxidation</td>
<td>-45.6</td>
<td>9.23</td>
</tr>
<tr>
<td></td>
<td>Demethylation</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Aflatoxin B1</td>
<td>Epoxidation</td>
<td>-45.2</td>
<td>5.16&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Hydroxylation</td>
<td>-38.9</td>
<td>6.20</td>
</tr>
</tbody>
</table>

<sup>a</sup> Distance to the heme within range needed for reaction to proceed.
Evolutionary toxicogenomics: diversification of the *Cyp12d1* and *Cyp12d2* genes in *Drosophila melanogaster* and related species

INTRODUCTION

The process of gene duplication, by which most new genes are gained (Ohno, 1970; Zhou et al., 2008), combined with stochastic gene loss, plays an important role in speciation by creating chromosomal differences between populations that lead to postzygotic mating barriers (Hahn et al., 2007; Lynch and Conery, 2000). Although the rate of gene duplication is of the same order of magnitude as the rate of nucleotide mutation (Li, 1999), loss of a duplicate gene is more likely than its preservation (Lynch and Conery, 2000, 2003). Within *Drosophila melanogaster*, 44.4% of new genes show copy number polymorphisms within a population as evidenced by 2658 genes that are recent copy number duplications, losses, or both (Emerson et al., 2008; Zhou et al., 2008). Comparisons of the genomes of 12 *Drosophila* species have revealed that high rates of gene gain and loss are often masked by a stasis in total gene numbers among species (Hahn et al., 2007).

Gene duplication and divergence are overwhelmingly considered the primary mechanism by which the cytochrome P450 monooxygenases (P450s) have radiated into a large and diverse gene family (Nelson, 1998, 2009). Found in almost all taxa, including bacteria, plants, insects, and vertebrates, P450s perform heme-dependent oxidative reactions necessary for the biosynthesis of hormones, fatty acids and pheromones and the detoxification of drugs, plant allelochemicals and insecticides (Schuler and Werck-Reichhart et al., 2003; Werck-Reichhart and Feyereisen, 2002; Schuler et al., 2006; Feyereisen, 2005; Li et al., 2007; Guengerich, 2005; Kelly et al., 2005). In the expansive P450 superfamily found in these organisms, the rates of gene gain and loss are not necessarily equal across all P450 families and their higher-order clans. This
is evident in comparisons across ten vertebrate species where the rate of P450 gene duplication or loss is higher for P450s with xenobiotic substrates than for P450s with endogenous substrates (Thomas, 2007). It is also evident in natural populations of *D. melanogaster* where *Cyp6g1*, which is associated with resistance to the insecticides DDT and imidacloprid (Daborn et al., 2001; Brandt et al., 2002), shows copy number polymorphisms (Emerson et al., 2008).

Further analysis of the *D. melanogaster* reference genome has indicated that *Cyp12d1*, whose function is not yet defined, has recently duplicated yielding the tandem gene pair, *Cyp12d1-p* and *Cyp12d1-d*; these are 99.4% identical in the coding region, resulting in only 3 amino acid differences. Although the function of *Cyp12d1* has not yet been defined, it is known that *Cyp12d1* is differentially expressed in response to a diverse array of xenobiotics, including DDT (Festucci-Buselli et al., 2005), phenobarbital (LeGoff et al., 2006; Sun et al., 2006), atrazine (LeGoff et al., 2006), caffeine (Willoughby et al., 2006), piperonyl butoxide (Willoughby et al., 2007), hydrogen peroxide (Li et al., 2008) and pepper (*Piper nigrum*) extract (Jensen et al., 2006), suggesting that it has an important role in the general response to xenobiotics. In two DDT-resistant strains (91-R and Wisconsin), *Cyp12d1* transcript expression levels are constitutively higher than in a susceptible strain (Canton-S) (Pedra et al., 2004) and are more highly induced by DDT (Festucci-Buselli et al., 2005). The *Cyp12d1* transcripts in the 91-R strain are particularly interesting since they differ both in their overall expression level and size compared to their counterparts in the DDT-susceptible 91-C strain that arose from the same originating population collected in 1952. With the availability of (i) 12 sequenced genomes from various *Drosophila* species, (ii) a large pool of *D. melanogaster* strains of diverse origin accessible, and (ii) a number of strains with recent microevolutionary changes in *Cyp12d1*, the
Cyp12d1 gene becomes a logical candidate to investigate both macro- and microevolutionary changes of gene sequences, duplications/losses and xenobiotic responses.

In this study, I have evaluated macro and microevolutionary changes in Cyp12d1 and a tandemly duplicated gene, Cyp12d2. On the macroevolutionary scale, I used phylogenetic comparisons of the sequenced Cyp12d1 region between Drosophila species to determine when Cyp12d2 likely emerged. On a microevolutionary scale, I used variations in the copy number of Cyp12d1 (due to repeated losses and duplications) across a diversity of D. melanogaster populations to demonstrate the genomic flux associated with this xenobiotic-responsive gene. At the finest scale microevolutionary level used in this study, I identified how a single copy of the Cyp12d1 gene has diverged through intron splicing of the Cyp12d1 transcript between 91-C and 91-R strains, whose recent evolutionary histories are well defined in terms of their divergence dates (58 years ago) and their selection pressure under DDT treatment.

**MATERIALS AND METHODS**

*Sequence datasets*

Sequence data were downloaded from FlyBase ([http://flybase.org/](http://flybase.org/)) using the following genome builds: *D. ananassae* (R1.3), *D. erecta* (R1.3), *D. grimshawi* (R1.3), *D. mojavensis* (R1.3), *D. persimilis* (R1.3), *D. pseudoobscura pseudoobscura* (R2.6), *D. sechellia* (R1.3), *D. simulans* (R1.3), *D. virilis* (R1.2), *D. willistoni* (R1.3), and *D. yakuba* (R1.3). Genomic and predicted amino acid sequences of Cyp12d1 orthologs were collected from each genome using the BLASTn function and the cDNA sequence of Cyp12d1-p from *D. melanogaster*.

*Fly Lines and Rearing*
The 53 wildtype strains of *D. melanogaster* used in this study were obtained from the Bloomington *Drosophila* Stock Center (http://flystocks.bio.indiana.edu/). Strains include: Amherst 3, BER 2, Berlin K, BOG 2, BOG 3, Canton-S, Canton-S-iso2B, CO 4, CO 7, Crimea, EV, Florida-9, Harwich, Hikone-A-S, Hikone-A-W, Hikone-R, KSA 3, KSA 4, Lausanne-S, MO 1, MWA 1, NO 1, Oregon-R, Oregon-R-modENCODE, Oregon-R-P2, Oregon-R-S, Oregon-R-SNPiso2, pi2 <P>, PYR 3, RC 1, Reids 1, Reids 2, Reids 3, RVC 2, RVC 4, Samarkand, Swedish-C, TW 1, TW 2, TW 3, Urbana-S, VAG 2, VAG 3, Wild 10E, Wild 11C, Wild 11D, Wild 1A, Wild 1B, Wild 2A, Wild 3B, Wild 5A, Wild 5B, Wild 5C. The laboratory strains *w*^118^ and *y; cn bw sp* were also obtained from the Bloomington *Drosophila* Stock Center. The *Wisconsin* strain was collected in Door County, Wisconsin (Brandt et al., 2002). The 91-R and 91-C strains, derived from a common population founded from several hundred individuals collected in St. Paul, Minnesota in 1952 (Dapkus and Merrell, 1977), were selected for DDT resistance (91-R) or never exposed to DDT (91-C) and were provided to us in 2000 by Dr. Ranjan Ganguly (University of Tennessee). From 2000 to 2003, 91-R was periodically selected for DDT-resistance by collecting survivors from vials with 4000 ug DDT per vial (Festucci-Buselli et al., 2005).

**DNA extraction and polymerase chain reaction analysis**

Genomic DNA was extracted from 10-15 adult flies using the DNeasy Blood and Tissue Kit (Qiagen, Inc., Valencia, CA). DNA was quantified by spectrophotometry using Nanodrop 1000 (Thermo Scientific, Wilmington, DE). The *Cyp12d1* region, based on the reference genome, was amplified from 50 ng of genomic DNA using 2µL of primers (10 pmol/µL), 5 U TaKaRa Long Amplification Taq polymerase (Otsu, Shiga, Japan) with a final concentration of 1X LA PCR™ Mg^{2+}-free buffer II, 2.5 mM MgCl₂, 2.5 mM each dNTP in 50 µL volume,
according to manufacturer’s instructions. Thermal cycling started with 94°C for 1 min., followed by 30 cycles of 98°C for 10 sec and a combined annealing and extension step of 68°C for 12 min, and finished with a 72°C heating for 15 min.

Primers were designed to amplify the regions spanning Cyp12d1-p and Cyp12d1-d, based on the D. melanogaster reference genome (y; cn bw sp). This region included duplicated noncoding regions as well as duplicated coding regions. Additional primers were designed to amplify sequences ~1.2 kb upstream and ~0.58 and 5.8 kb downstream of the duplicated region in the reference genome (Table 4.1).

*Sequencing and alignment of Cyp12d1 genomic region from 91-C and 91-R*

All amplification products were purified using the QIAquick PCR purification kit (Qiagen, Valencia, CA). The amplification products of the Cyp12d1 genomic region for the 91-C and 91-R strains were sequenced with 2X coverage by primer walking at the Core DNA Sequencing Facility of the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois Urbana-Champaign. Nucleotide alignments were performed with Clustal W (3.2) using Biology Workbench ([http://workbench.sdsc.edu/](http://workbench.sdsc.edu/)).

*Quantitative real-time polymerase chain reaction*

Total RNA was extracted from 20-25 adult flies using the Qiagen RNeasy kit and DNase treatment set (Valencia, CA). Total RNA was quantified by spectrophotometry using Nanodrop 1000 (Thermo Scientific, Wilmington, DE). First strand cDNA was synthesized from 1 µg total RNA in a 20 µL using the iScript™ cDNA synthesis kit (Bio-Rad Laboratories, Hercules, CA). A 25-fold dilution was made of each first strand cDNA reaction for quantitative PCR. qRT-PCR
was performed with Fast SYBR Green Master mix from Applied Biosystems (Foster City, CA) with SYBR Green dye on a Applied Biosystems Step One Plus Real-Time PCR System. For each cDNA, three qRT-PCR reactions were performed. The average threshold cycle (C\textsubscript{T}) was calculated by the Step One Software (version 2.0). Relative expression levels were calculated as $2^{(C\textsubscript{T} \text{Cyp12d1} - C\textsubscript{T} \text{rp49})}$, in which the C\textsubscript{T} value for the reference gene rp49 is subtracted from the gene-specific average C\textsubscript{T} value. Statistical analysis of the relative gene expression levels in y; cn bw sp were compared using a paired t-test of the relative amounts of Cyp12d1-d and Cyp12d1-p were performed using PASW (SPSS Inc., Chicago, IL). The primers used for qRT-PCR are listed in Table 4.1.

*Polymerase chain reaction analysis of mRNA transcript variants of Cyp12d1*

Amplifications of 0.5 µL of cDNA were performed using 1.25 Units of TaKaRa Ex Taq\textsuperscript{TM} polymerase (Otsu, Shiga, Japan) with a final concentration of 1X Ex Taq\textsuperscript{TM} buffer and 2.5mM each dNTP in 50 ul volume, according to manufacturer’s instructions. Thermal cycling started with 94°C for 3 min, followed by 30 cycles of 94°C for 30 sec, 50°C for 1 min, and 72 °C for 2 min, and finished with a 72°C heating for 2 min.

*Prediction of secondary structure using RNAfold*

The RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) was used to investigate potential differences in secondary structure of the third intron, as found in the 91-R and reference strains. The program performs a minimum free energy calculation of the optimal secondary structure and partition function calculations, resulting in reliability annotations of the secondary structure.
RESULTS

Characterization of Cyp12d1 and Cyp12d2 across 12 Drosophila genomes

The 12 available Drosophila genomes demonstrate that Cyp12d1 duplicated into two genes before the divergence of D. melanogaster from other members of the melanogaster subgroup. In the six sequenced species of the melanogaster subgroup, Cyp12d1 was tandemly duplicated in the genome (Fig. 4.1) and in five of these the divergence was enough to refer to the second gene as Cyp12d2. In the D. melanogaster reference genome the two genes have 99.4% amino acid identity and were noticeably more conserved than paralogs from the other species; thus, we refer to the D. melanogaster paralogs as Cyp12d1-p and Cyp12d1-d. In D. simulans, D. erecta, D. ananassae and D. sechellia Cyp12d1 and Cyp12d2 display between 49-66% amino acid identity (Table 4.2). In D. yakuba, Cyp12d1 and Cyp12d2 display only 12.1% amino acid identity (Table 4.2), but the predicted amino acid sequence of Cyp12d1 is truncated at 169 residues and it is, therefore, most likely a pseudogene.

The three sequenced genomes of D. pseudoobscura, D. persimilis and D. willistoni, the most closely related basal species, have retained a single copy of Cyp12d1, found at the same spatial location on the chromosome, around 10 kb from the BBS4 gene. Within the more distantly related genomes of D. virilis, D. mojavensis and D. grimshawi, a single copy of Cyp12d1 resides outside the conserved location of the other 9 species (Fig. 4.1), suggesting that the Cyp12d1 gene became fixed after the divergence of the Sophophora.

Single copies of Cyp12d1 from the basal species within the Sophophora (D. pseudoobscura, D. persimilis, D. willistoni) were more closely related to Cyp12d1 than Cyp12d2, according to genetic distance analysis of the amino acid sequences (Fig. 4.2). In addition, the four full-length Cyp12d2 sequences form a monophyletic group. Cyp12d1 from D.
virilis and *D. mojavensis* create a monophyletic clade that is basal to *Cyp12d1* and *Cyp12d2* from the Sophophoran species. The *Cyp12d2* pseudogene from *D. yakuba* shows the least similarity to the other species. *Cyp12d1* from *D. grimshawi*, the most distant relative in the 12 species, was used as an outgroup for rooting the tree (Fig. 4.2).

**Copy number polymorphism in Cyp12d1 region across *D. melanogaster* strains**

According to the published sequence of the reference genome, *Cyp12d1*-p and *Cyp12d1*-d from *D. melanogaster* are 99.4% identical in their coding regions, differing by only three amino acids. The flanking regions of the coding sequence are 100% identical between the two copies, from 1867 nucleotides upstream of the transcription start site to 136 nucleotides downstream of the stop codon ([http://flybase.org/](http://flybase.org/), *D. melanogaster* (R5.23)). *Cyp12d1* region primers were designed to amplify the entire *Cyp12d1* region by flanking the duplicated sequences (Fig. 4.3, Table 4.1). In total, the *Cyp12d1* regions of 58 strains of *D. melanogaster* were analyzed by PCR. For the reference strain (*y; cn bw sp*) and nine other strains, amplification with the first set of primers produced the expected band size hereafter referred to as the 7.7 kb band. This included three of the five strains of Oregon-R (Oregon-R, Oregon-R-S, and Oregon-R-SNPiso2). However, the majority of strains, 47 out of 58, produced a smaller band corresponding to a size of ~3.5 kb with the *Cyp12d1* region-spanning primer set, suggesting that the majority of strains tested likely contained only one copy of the *Cyp12d1*. A single strain, PYR3, from the Pyrenees, produced a single band corresponding to size of ~10 kb.

To determine if the tandem duplication is restricted to those strains that produced a band of 7 kb or larger, we used a method described by Emerson et al. (2008), in which a set of primers were designed to face inversely within the *Cyp12d1* gene (Fig. 4.3). In the presence of a tandem duplication, the forward and reverse tandem duplication primers produced a 3.5 kb band and, if
only one copy of Cyp12d1 resided there, the primers did not produce an amplified product. All of the strains that produced 7.7 kb bands for the Cyp12d1 region primers also produced an amplicon using the tandem duplication primers, supporting the hypothesis that they contained tandem duplications of Cyp12d1. For the strains that produced a 3.8 kb band, all of the 47 strains showed the absence of a tandem duplication of Cyp12d1.

In order to detect additional DNA corresponding to a second Cyp12d1 gene in these strains, genomic regions, based on the reference sequence, both upstream and downstream of the original 7.7 kb region were amplified in three strains, 91-C, 91-R and the reference strain with the 571 nt downstream, 5264 nt downstream, and 1207 nt upstream primers (Table 4.1). Amplification of these regions revealed no difference between the reference strain, 91-C and 91-R. When the upstream primer (1207 nt upstream-for) was combined with either of the downstream primers, the reference strain produced the expected bands, 8 kb and 12 kb, respectively. For 91-C and 91-R, the band sizes were 5.5 kb and 10 kb, respectively, which were the expected lengths for the Cyp12d1 gene and flanking regions if one copy of Cyp12d1 genes had been lost (Table 4.3).

Sequence differences in the Cyp12d1-d and Cyp12d1-p in the D. melanogaster reference genome

A detailed analysis of the Cyp12d1-d and Cyp12d1-p gene region of the reference genome indicated that 187 nucleotides downstream of the stop codon of Cyp12d1-d contained a conserved polyadenylation signal (AATAAA), which was not found in Cyp12d1-p. In concordance with this observation, the 3’ UTR of the three full-length polyadenylated mRNA transcripts in the Genbank database (BT001433.1, AY061415.1, BT031280.1) correspond to the 3’ UTR of Cyp12d1-d, not Cyp12d1-p. Cyp12d1-p and Cyp12d1-d both contain a suboptimal polyadenylation sequence (AGTAAA) that would produce an mRNA transcript ~124 nt shorter.
than those found in Genbank (Fig. 4.4). Based on the lengths of these mRNA transcripts, the optimal polyadenylation sequence lies within 20 nucleotides of the poly (A) tail, an appropriate distance for *Drosophila* polyadenylation sequences (Lutz, 2008).

In order to determine if transcripts expressed in the *y; cn bw sp* reference strain are *Cyp12d1-d* or *Cyp12d1-p*, we designed gene-specific *Cyp12d1-p* 3’ UTR and *Cyp12d1-d* 3’ UTR primer sets in which the reverse directional primers were unique to the 3’ UTR region of either gene. In the reference strain, the *Cyp12d1-d* specific primer produced, on average, 63-fold (95% CI=29.8-131.7) higher expression than the *Cyp12d1-p* primer set (t-test, t=12.857, d.f.=8, p<.0001) (Fig. 4.5). In the 91-C and 91-R strains, only the *Cyp12d1-p* primer set produced an amplified signal, which corresponds with their genomic sequences matching *Cyp12d1-p*, not *Cyp12d1-d* (Fig. 4.4).

*Changes in the single copy Cyp12d1 in the 91-C and 91-R strains*

Based on sequencing of the genomic region of *Cyp12d1* found in 91-C and 91-R, I observed only a single copy of the gene containing the 5’ upstream region of *Cyp12d1-d* in the reference genome, that shares the three coding region SNPs and the 3’ untranslated region (UTR) of *Cyp12d1-p* of the reference genome. As a consequence, this *Cyp12d1* gene contains a suboptimal polyadenylation signal found in *Cyp12d1-p* and lacks the highly conserved polyadenylation site found in *Cyp12d1-d* (Fig. 4.4).

Within the *Cyp12d1* coding regions, introns 1 and 2 from the 91-C and 91-R strains were identical to the reference genome, but intron 3 differed in the 91-R strain in having a GT to AT change in the 5’ splice site that would prevent its recognition and cleavage. I predicted, based on the altered intron splice site junctions in 91-R that the *Cyp12d1* transcript from this strain would retain the third intron (Fig. 4.6). Amplification and sequencing of the full length coding sequence
of Cyp12d1 from 91-R confirmed the retention of intron 3 in the transcript. Inclusion of intron 3 creates a premature stop codon that would eliminate 339 amino acids from this protein (Fig. 4.6). Interestingly, the upstream noncoding regions of Cyp12d1 do not largely differ between the 91-C and 91-R strains, but they do differ from the reference genome with 2 insertions (40 and 17 nt), an 18 nt deletion (981 upstream of ATG) and the transversion of a 23 nt TA-rich region to GC-rich directly upstream of the 17 nt insertion (Fig. 4.7).

**DISCUSSION**

Describing the evolution of cytochrome P450 genes in insects informs both our understanding of microevolutionary environmental responses and macroevolutionary speciation patterns. Using the twelve sequenced Drosophila genomes, and detailed analyses of D. melanogaster strains, we were able to follow the duplication, divergence, and loss of Cyp12d1, a P450 that responds to a variety of xenobiotics in D. melanogaster. All of these species contain at least one full-length Cyp12d1 sequence, suggesting that this particular P450 did not emerge within Drosophila and that perhaps its function may be fundamental and not specific to life history of the pomace flies. Based on the conserved positioning of Cyp12d1 to the next closest gene, BBS4, Cyp12d1 became fixed in the Sophophora clade on the chromosome 2, according to those genomes that are fully assembled. Within the lineage of the melanogaster subgroup, Cyp12d1 appears to have duplicated and been retained within the 6 genomes that are currently available. Based on overall amino acid identity, Cyp12d1 and Cyp12d2, in D. simulans, erecta, and ananassae, have diverged to a similar extent, while in D. sechellia Cyp12d2 has diverged to a greater extent although it is still predicted to be a functional protein. On the other hand, in D. yakuba, Cyp12d1 shares high sequence identity with other Cyp12d1 orthologs, and only a truncated Cyp12d2-like sequence is retained in the genome.
In the genomic reference strain of *D. melanogaster*, Cyp12d1 (Cyp12d1-p) and Cyp12d2 (Cyp12d1-d) were nearly identical and were similar to Cyp12d1 from the other species. PCR amplification of genomic DNA of multiple strains of *D. melanogaster* revealed that the copy number of Cyp12d1 varies between 1 or 2 copies and may represent independent gene loss/duplication events. Mapping of duplicated regions of the reference sequence of the Cyp12d1 region shows the breakpoints for the tandem duplication of Cyp12d1 that generated a core polyadenylation sequence in the 3’ UTR of Cyp12d1-d. Analysis of relative transcript abundance of Cyp12d1-d and Cyp12d1-p in the reference strain demonstrates that Cyp12d1-d experiences significantly higher transcription than Cyp12d1-p (Fig. 4.5), which may be due to higher transcript stability. Sequencing of the entire Cyp12d1 genomic region in 91-C and 91-R revealed a mosaic of Cyp12d1-d and Cyp12d-p, in which the 3’ UTR corresponds to Cyp12d1-p, and thus lacks a conserved polyadenylation signal sequence. Based on genomic and mRNA transcript sequencing, the strain 91-R retains the third intron, which contains a premature stop codon, but previous evidence demonstrates a normal-sized protein found in 91-R (Festucci-Buselli et al., 2006). Thus, our observations suggest that Cyp12d1 in *Drosophila melanogaster* and related species has undergone multiple paths of evolution that could shed light on the role of Cyp12d1 in these insects.

Based on amino acid similarity, within the Sophophoran species, the single Cyp12d1 orthologs cluster with Cyp12d1 of the duplicated pair to forms a monophyletic clade (Fig. 4.2). In contrast, the Cyp12d2 orthologs have developed in three distinct manners, through divergence into a second, putatively functional protein (*D. simulans*, *D. sechellia*, *D. erecta*, and *D. ananassae*), gene loss through non-functionalization (*D. yakuba*), or presumable gene loss and subsequent gene duplication (*D. melanogaster*). The patterns of divergence of Cyp12d1 and
Cyp12d2 across species seem to depend more on phylogenetic history (D. simulans/D. sechellia) or the rate of genomic evolution (D. melanogaster, D. yakuba, and D. erecta) than on dietary breadth or geographic distribution. D. simulans and D. sechellia are sister species whose Cyp12d1 and Cyp12d2 are respectively 99% and 72% identical (Table 4.2), yet their diet and geographic distribution are very different. D. sechellia is restricted to the Seychelles Islands where it specializes on the toxic host plant Morinda citrifolia (R’Kha et al., 1991; McBride, 2007), while D. simulans is cosmopolitan in distribution and is a generalist feeder (Markow and O’Grady, 2007; Singh et al., 2009). The identification of Cyp12d1 and Cyp12d2 in these genomes presents a unique opportunity to study how cytochrome P450 gene duplication leads to differential functionalization across different evolutionary trajectories of these six species.

Without the additional 11 genomes, D. melanogaster offered a very different scenario, in which Cyp12d1-p and Cyp12d1-d had newly duplicated. This story still stands, in some regards, but is more complicated than would have been first assumed from the sequence of the reference genome. In D. melanogaster, Cyp12d1 copy number varies across strains, while Cyp12d2 seems to have been lost. This evidence supports the hypothesis that P450s related to xenobiotics show higher rates of gene birth/death than those with endogenous substrates (Thomas 2007). Copy number variation in D. melanogaster cytochrome P450 gene is not new; Cyp6g1, a cytochrome P450 associated with resistance to the insecticides DDT and imidacloprid (Daborn et al, 2001), shows copy number polymorphisms in natural populations (Emerson et al., 2008). What is intriguing about Cyp12d1 in D. melanogaster is the loss of Cyp12d2 and subsequent re-duplication of Cyp12d1 in some strains. The juxtaposition of these varying outcomes of the Cyp12d1 duplication across and within species raises the questions (i) is there an advantage in
maintaining two copies of the Cyp12d1 gene or (ii) is this fluctuation in gene copy observed in D. melanogaster simply an artifact of the high rate of purging of pseudogenes from its genome?

The high rate of duplication and loss observed in D. melanogaster may provide genetic variability upon which natural selection can act to generate novel regulatory mechanisms for xenobiotic response. In the three strains, y; cn bw sp, 91-C and 91-R, the evolution of Cyp12d1 follow three different trajectories. In the reference strain, Cyp12d1 is tandemly duplicated to produce two nearly identical genes that differ in their 3’ UTR and relative abundance, possibly due to a conserved polyadenylation signal sequence that would increase mRNA stability. By examining the microevolution of Cyp12d1 in two strains, 91-C and 91-R, whose life histories have been manipulated over the last 58 years, one can begin to document how this region responds to selection pressure by the xenobiotic DDT and compare it to the reference strain, y; cn bw sp. In 91-C and 91-R, two strains are derived from a single population, and one of them (91-R) has undergone intense selection pressure for DDT resistance, while the other was maintained without selection pressure. In both strains, Cyp12d1 did not tandemly duplicate, nor did the 5’ non-coding region diverge significantly in response to DDT selection pressure. Instead, in the 91-R strain, a mutation in the intron splice site junction of intron 3 occurred from standing genetic variation. While this intron contains a premature stop codon, 91-R has been shown to produce a similarly sized protein as found in 91-C and other D. melanogaster strains tested (Festucci-Buselli et al., 2005). One possible explanation for this incongruence is a phenomenon called translational readthrough. Originally described in viruses, it has been proposed that translational readthrough is not a rare occurrence in Drosophila, based on comparisons of candidate protein-coding genes in the 12 Drosophila genomes (Lin et al., 2007). Steneberg and Samakovlis (2001) reported a novel translational readthrough mechanism for the
transcription factor Headcase in *D. melanogaster* in which the formation of a stem and loop secondary structure of an 80 nucleotide region downstream of a UAA stop codon is sufficient to suppress termination of the mRNA. For *Cyp12d1* in 91-R, the insertion of 5 nucleotides into intron 3 produces a different secondary structure than that in the reference strain (Fig. 4.8) and produces a stem and loop structure near the stop codon that could allow for translational readthrough of the stop codon. Retention of the third intron may also have a role in increasing mRNA stability, in lieu of the conserved polyadenylation signal sequence that 91-R lacks.

From three *D. melanogaster* strains, we have observed complexity of the *Cyp12d1* region that surpasses the standard model of insect P450 regulation by the 5’ noncoding regulatory regions. In the study of human disease, individual variation in cytochrome P450 genes that affect mRNA stability and transcript splicing are linked to differential outcomes. Polyadenylation signal sequences that change mRNA stability in P450s are associated with harm avoidance in human males due to aromatase (Matsumoto et al., 2009) and prostate cancer due to vitamin D metabolism (Ahn et al., 2009). Nicotine C-oxidase in humans (CYP2A6) 3’ UTR polymorphism affects mRNA stability and protein expression (Wang et al., 2006). Splice site mutations are associated primarily with hereditary diseases, including congenital adrenal hyperplasia (Concolino et al., 2009; Soardi et al., 2009; Baumgartner-Parzer et al., 2001), vitamin D-dependent rickets type 1 (Kim et al., 2007) cerebrotendinous xanthomatisis (Price Evans et al., 2007), sex reversal (Bhangoo et al., 2006), Bietti crystalline corneoretinal dystrophy (Shan et al., 2005) and prostacyclin synthase-mediated hypertension (Nakayama et al., 2002).

Like individual variation in humans, inter-strain comparisons, like that of 91-C and 91-R in the *Cyp12d1* region, highlight the way in which evolutionary forces may shape differential gene expression and protein function of insect P450s. In *D. melanogaster*, divergence in the
*Cyp12d1*, in just 50 years between 91-C and 91-R, shows how even a single copy of a xenobiotic-responsive gene can be altered and constrained by selection. The natural genetic variation that existed in this wild population seemed to be selected under this stringent DDT resistance regime that produced 91-R. In 91-C, we observed a greater number of SNPs in the coding region than in 91-R, which is more similar to the reference strain (*y; cn bw sp*). Knowing that 91-C and 91-R are more closely related, it therefore seems most likely that 91-R and *cn bw sp* share the ancestral states of these SNPs, while 91-C has undergone relaxed constraint on its coding region, when it was brought into the laboratory. In other words, exposure to DDT via selection or a random mechanism such as genetic hitchhiking seems to have constrained the ancestral form of *Cyp12d1*, except for one crucial mutation in the splice site junction.

By using macro and microevolutionary approaches to study the *Cyp12d1* region, we have demonstrated considerable variability in a xenobiotic-responsive gene that reinforces the importance of gene duplication, loss and mutations outside the 5’ regulatory regions in the evolution of cytochrome P450 genes. The 12 *Drosophila* genomes offer an unprecedented view of the divergence of paralogous and orthologous genes. For this xenobiotic-responsive gene, we observed that the process of duplication and divergence is not simply linked to ecological traits of the species or one selection pressure, such as DDT resistance, but is constrained by evolutionary history and stochastic genetic events that provide the variation on which selection can act, once it is applied. Thus, limiting an observation about a gene based on one strain of one species, as was done for *Cyp12d1* in the reference strain, can result in a myopic view of the evolution of that gene. In this way, “evolutionary toxicogenomics”, comparing xenobiotic-responsive genes across species, holds potential for understanding the ways a gene can respond to selection in its environment and the extent to which it is constrained by its genomic history.
The power of evolutionary toxicogenomics relies on the generation of genomic sequences across closely related species and population within a species. As the cost of sequencing genomes decreases, the availability of these genomic sequences should dramatically increase our ability to understand the evolution of detoxification systems by identifying genes and gene families are shaped by broad evolutionary patterns and those that are shaped by local or specific evolutionary challenges, abiotic and biotic. With the wealth of genetic tools available for *D. melanogaster*, we may be able to link these micro-evolutionary changes with specific laboratory or field-based environmental challenges experienced by *D. melanogaster* populations (Daborn et al., 2001; Festucci-Buselli et al., 2005), even beyond pesticide resistance.

**ACKNOWLEDGMENTS**

I thank Dr. Hongmei Li and Dr. Weilin Sun for technical assistance with qRT-PCR analysis. I thank Susan Balfe for her guidance in the laboratory. Emilie Bess assisted in generating the genetic distance phylogeny. Research was funded by C.W. Kearns, C.L. Metcalf and W.P. Flint Endowment Funds to Dr. Barry Pittendrigh. I was funded by a UIUC Dissertation Completion Fellowship.

**LITERATURE CITED**


Figure 4.1 Phylogeny of the twelve Drosophila species with sequenced genomes showing a single duplication event of Cyp12d1 within Drosophila evolution, its chromosomal location (the second right, 2R, the third, 3, or not determined, nd), and the next closest gene on the chromosome. From this analysis we hypothesize that Cyp12d1 arose before the split between Hawaiian Drosophila and the Sophophora, and came to reside in the same chromosomal arrangement by the divergence of the willistoni group, based on the next closest gene, BBS4. Direction of the genes indicated by the arrowheads show that despite their conservation in chromosome position, large-scale chromosomal inversions have occurred in some species. After the emergence of the melanogaster group, Cyp12d1 duplicated into two genes, Cyp12d1 and Cyp12d2, which remain tandemly arranged on the chromosome. The paralogs have since diverged to different extents in each species, as indicated by the percentage of sequence identity. Based on amino acid sequence similarity between all of the species, Cyp12d1, in blue, is the ancestral gene, while Cyp12d2, in orange, is most likely the result of a duplication and divergence event. The shading of each gene, in blue or orange, correlates with the level of sequence identity between each gene and its ortholog in D. simulans. In D. yakuba, Cyp12d2, marked by a red X, has become a pseudogene that is predicted to produce a truncated protein. In the other species, Cyp12d2 encodes a full-length putatively functional protein.
Figure 4.2 A neighbor-joining tree of the *Drosophila* Cyp12d1 and Cyp12d2 genes, based on amino acid sequence distances, demonstrate two distinct clades for Cyp12d1 and Cyp12d2 from the Sophophora species. Cyp12d1 from the more basal species, *D. virilis* and *D. mojavensis*, form their own clade, while *D. grimshawi* of the Hawaiian *Drosophila*, was used to root the tree. The Cyp12d1 clade includes single genes from the more basal species, *D. persimilis*, *D. pseudoobscura* and *D. willistoni*. The Cyp12d2 clade contains the four Cyp12d2 that are predicted to encode for full-length proteins. The truncated Cyp12d2 protein sequence of *D. yakuba* does not cluster with any of the full-length sequences. Thus, based on its conservation across the Sophophora, Cyp12d1 appears to be the ancestral gene in the tandem duplication, with a divergence pattern that follows that of the *Drosophila* species, while Cyp12d2 has diverged sufficiently to form a monophyletic clade.
Figure 4.3 PCR amplification with the Cyp12d1 region primers (“Cyp12d1 region for” and “Cyp12d1 region rev”) of genomic DNA from the 91-C and 91-R strains revealed that the region was only half the length of the Cyp12d1 gene region in the reference genome (y; cn bw sp). Based on this evidence, we predicted that 91-C and 91-R contain a single copy of Cyp12d1 instead of the tandem duplication observed in the reference genome. To test this prediction, we used the tandem duplication primers (“tandem duplication for” and “tandem duplication rev”), which produce an amplification product for the reference strain, but not 91-C and 91-R. Sequencing through the Cyp12d1 gene in 91-C and 91-R revealed that in these strains, Cyp12d1 lacks the conserved polyA signal sequence of Cyp12d1-d and contains only the suboptimal polyA signal sequence of Cyp12d1-p.
Cyp12d1 3’UTR for

91C genomic  GCCGTCATTTCAAGACCATTGTCTGCAATGAAACCCCTCAGTTGGTCTCCCTTACAATTTCA
Cyp12d1-p   GCCGTCATTTCAAGACCATTGTCTGCAATGAAACCCCTCAGTTGGTCTCCCTTACAATTTCA
91R genomic  GCCGTCATTTCAAGACCATTGTCTGCAATGAAACCCCTCAGTTGGTCTCCCTTACAATTTCA
Cyp12d1-d   GCCGTCATTTCAAGACCATTGTCTGCAATGAAACCCCTCAGTTGGTCTCCCTTACAATTTCA

stop codon

91C genomic  CCGATATGACATTCTTTTTTTTTTTCTCTTATTACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT...
Figure 4.5 C<sub>T</sub> expression values from qRT-PCR analysis of Cyp12d1-<i>p</i> and Cyp12d1-<i>d</i> using the 3’ UTR primers (Cyp12d1 3’UTR for, Cyp12d1-<i>p</i> 3’UTR rev and Cyp12d1-<i>d</i> 3’UTR rev). Average threshold cycle (C<sub>T</sub>) for three technical replicates was calculated by the Step One Software (version 2.0). Relative expression levels were calculated as $2^{\Delta \Delta C_T}$, in which the C<sub>T</sub> value for the reference gene <i,rp49</i> is subtracted from the gene-specific average C<sub>T</sub> value. Statistical analysis of the relative gene expression levels (n=9) in <i>y; cn bw sp</i> were compared using a paired t-test of the relative amounts of Cyp12d1-<i>d</i> and Cyp12d1-<i>p</i>. The lower C<sub>T</sub> expression value for Cyp12d1-<i>d</i> indicates its higher transcript abundance in the samples. On average, Cyp12d1-<i>d</i> has 63-fold (95% CI=29.8-131.7) higher expression than Cyp12d1-<i>p</i> (t-test, t=12.857, d.f.=8, p<.0001).
Figure 4.6 Alignment of the sequenced Cyp12d1 genomic region from 91-C and 91-R strains with +593/+893 regions of Cyp12d1-d and Cyp12d1-p of the reference genome shows a GT to AT mutation (italicized, highlighted purple) of the 3rd intron (bolded) splice site junction in 91-R. This mutation was predicted to cause the retention of intron 3 in the Cyp12d1 transcript of the 91-R strain, which was observed when the transcript was sequenced. Retention of intron 3 introduces a premature stop codon, if functional, (bolded, underlined, highlighted yellow) that would decrease the protein length by 339 amino acids.
Figure 4.7 Alignment of the sequenced Cyp12d1 genomic region from 91-C and 91-R strains with -545/-788 upstream non-coding regions of Cyp12d1-d and Cyp12d1-p of the reference genomes shows no major sequence differences between 91-C and 91-R. Instead, 91-C and 91-R share two insertions (40 and 17 nt) and a transversion of a 23 nt TA-rich region to GC-rich (italicized) directly upstream of the 17 nt insertion. An 18 nt deletion at -981 is not shown.
Figure 4.8 The predicted secondary structure of intron 3 in the mRNA transcript from *Cyp12d1* for the reference strain (A) and 91-R (B) using RNAFold program ([http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi](http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)) shows the difference in secondary structure that results from the insertion of 5 nucleotides (UGUAU) in the genome of 91-R. Arrows on each structure designate the start of the intron. The reference strain is included for comparison and is not predicted to retain intron 3 in the *Cyp12d1* transcript. Color-coded scales from 0 to 1 denote the probability of dimerization with red and yellow being the highest and second highest. In 91-R, intron 3 is retained in the mRNA transcript, which introduces a premature stop codon (UAA), starting at the second nucleotide of the sequence. According to the predicted secondary structure, the stop codon flanks a stem and loop structure, which could disrupt normal translation and result in translational readthrough, as observed in the Headcase protein (Steneberg and Samakovlis, 2001).
### TABLES

Table 4.1. Primers used to sequence and analyze the Cyp12d1 genomic region of *Drosophila melanogaster* strains.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer direction</th>
<th>Primer sequence (5’- 3’)</th>
<th>Relative to Cyp12d1 duplicated region of the reference genome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cyp12d1 region</strong></td>
<td>for</td>
<td>CACCGCAGCAGAGGAAAGCAGTCCTC</td>
<td>112 nt upstream</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>TAGGGATGCCAAACAAAGCG</td>
<td>56 nt downstream</td>
</tr>
<tr>
<td><strong>1207 nt upstream</strong></td>
<td>for</td>
<td>TGTGTGCTCAGGCTCAGTGGTTTTGAT</td>
<td>1207 nt upstream</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>GTGCTTTTGTCTGCTGAGGTTG</td>
<td>91 nt upstream</td>
</tr>
<tr>
<td><strong>571 nt downstream</strong></td>
<td>for</td>
<td>GCAAAACCTCAGGCTTTTGTGAT</td>
<td>Overlaps the last 17 nt of the 3’ end</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>AAAAACCCCGAAAAATCGCTTA</td>
<td>571 nt downstream</td>
</tr>
<tr>
<td><strong>5264 nt downstream</strong></td>
<td>for</td>
<td>GCAAAACCTCAGGCTTTTGTGAT</td>
<td>Overlaps the last 17 nt of the 3’ end</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>TGCTTCTCTGCTGATGGTTG</td>
<td>5264 nt downstream</td>
</tr>
<tr>
<td><strong>Tandem duplication</strong></td>
<td>for</td>
<td>CATTGAGCAGTGGCGGATCT</td>
<td>+9/+29</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>ATAATCGCCGTGAAATTGCT</td>
<td>-217/-197</td>
</tr>
<tr>
<td><strong>Cyp12d1 3’-UTR</strong></td>
<td>for</td>
<td>ATTTGCATTGTTGTGCAA</td>
<td>105 nt downstream of stop codon</td>
</tr>
<tr>
<td><strong>Cyp12d1-p 3’-UTR</strong></td>
<td>rev</td>
<td>TAGGGATGCCAAACAAAGCG</td>
<td>172 nt downstream of stop codon</td>
</tr>
<tr>
<td><strong>Cyp12d1-d 3’-UTR</strong></td>
<td>rev</td>
<td>TTTGCTGCTCGAATGCTTTT</td>
<td>166 nt downstream of stop codon</td>
</tr>
<tr>
<td></td>
<td>for</td>
<td>TCCATTTTACTGTTAGCGTTAAG</td>
<td>+707/+730</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>GTCTTGGAAAGGTTCAATG</td>
<td>+868/+888</td>
</tr>
<tr>
<td><strong>intron3</strong></td>
<td>for</td>
<td>CAACGAGGGTATCTGCGAGCT</td>
<td>+384/+406</td>
</tr>
<tr>
<td></td>
<td>rev</td>
<td>GCCATCCACACCAGCGAATAGG</td>
<td>+1144/+1166</td>
</tr>
<tr>
<td><strong>internal Cyp12d1</strong></td>
<td>for</td>
<td>CGGTTCAGGTGCAAACAAGCG</td>
<td></td>
</tr>
<tr>
<td><strong>Rp49</strong></td>
<td>rev</td>
<td>TGGGCGGCGTCGAAATCT</td>
<td></td>
</tr>
</tbody>
</table>

a Primers used in qRT-PCR reactions.
# Primer spans intron 3.
Table 4.2 Amino acid identity (%) based on pairwise alignments of CYP12D1 (A) and CYP12D2 (B) between species. The species compared were all within the *melanogaster* clade. Alignments were performed using BLOSUM50 matrix and gap opening and extension penalties of -12 and -2 in the LALIGN program (Pearson, 1990) in Biology Workbench (http://workbench.sdsc.edu/). CYP12D1 was more conserved among species than was CYP12D2.

A. CYP12D1

<table>
<thead>
<tr>
<th></th>
<th><em>D. sechellia</em></th>
<th><em>D. melanogaster</em></th>
<th><em>D. yakuba</em></th>
<th><em>D. erecta</em></th>
<th><em>D. ananassae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. simulans</em></td>
<td>99.0</td>
<td>97.9</td>
<td>91.7</td>
<td>90.8</td>
<td>81.0</td>
</tr>
<tr>
<td><em>D. sechellia</em></td>
<td>98.1</td>
<td>91.9</td>
<td>91.4</td>
<td>81.0</td>
<td></td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td></td>
<td>91.7</td>
<td>91.0</td>
<td>81.4</td>
<td></td>
</tr>
<tr>
<td><em>D. yakuba</em></td>
<td></td>
<td></td>
<td>90.0</td>
<td>80.0</td>
<td></td>
</tr>
<tr>
<td><em>D. erecta</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>79.3</td>
</tr>
</tbody>
</table>

B. CYP12D2

<table>
<thead>
<tr>
<th></th>
<th><em>D. sechellia</em></th>
<th><em>D. melanogaster</em></th>
<th><em>D. yakuba</em></th>
<th><em>D. erecta</em></th>
<th><em>D. ananassae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. simulans</em></td>
<td>71.8</td>
<td>65.3</td>
<td>12.9</td>
<td>77.4</td>
<td>67.7</td>
</tr>
<tr>
<td><em>D. sechellia</em></td>
<td>49.3</td>
<td>14.0</td>
<td>56.2</td>
<td>48.6</td>
<td></td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td></td>
<td>13.4</td>
<td>66.6</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td><em>D. yakuba</em></td>
<td></td>
<td></td>
<td>11.7</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td><em>D. erecta</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71.9</td>
</tr>
</tbody>
</table>
Table 4.3 PCR-based comparison of *Cyp12d1* region from the reference genome to that of strains 91-C and 91-R.

<table>
<thead>
<tr>
<th>Strain</th>
<th><em>Cyp12d1</em> region</th>
<th>Tandem duplication</th>
<th>571 nt downstream</th>
<th>5264 nt downstream</th>
<th>1207 nt upstream</th>
<th>1207 up/571 down</th>
<th>1207 up/5264 down</th>
</tr>
</thead>
<tbody>
<tr>
<td>y; cn bw sp</td>
<td>8</td>
<td>3.8</td>
<td>0.6</td>
<td>5</td>
<td>1.2</td>
<td>9.5</td>
<td>&gt;12</td>
</tr>
<tr>
<td>91-C</td>
<td>4</td>
<td>no product</td>
<td>0.6</td>
<td>5</td>
<td>1.2</td>
<td>5.5</td>
<td>10</td>
</tr>
<tr>
<td>91-R</td>
<td>4</td>
<td>no product</td>
<td>0.6</td>
<td>5</td>
<td>1.2</td>
<td>5.5</td>
<td>10</td>
</tr>
</tbody>
</table>

1Product size is based on comparison with 1 kb plus ladder (Fisher Scientific)
CHAPTER 5

Reproductive effects of reduced expression of Cyp4g15 in the nervous system of Drosophila melanogaster

INTRODUCTION

For the majority of cytochrome P450 monooxygenase (P450s) proteins in insects, substrates (and functions) are unknown. As more genomes are sequenced, comparative genomics offers a means to determine which genes retain a shared function across species and which have evolved novel functions. The genomes of multiple insect species have revealed different patterns of evolutionary change among cytochrome P450 gene families. Within the framework of the four major clades of animal P450s, CYP2, CYP3, CYP4 and the mitochondrial CYP clades, insect genomes diverge dramatically from vertebrate genomes, possibly reflecting the unique and diverse ecologies of insect species (Feyereisen, 2006). The large numbers of insect-specific radiations of the CYP6 and CYP9 families speak to the highly idiosyncratic insect responses to xenobiotic compounds (Feyereisen, 2006). Other families, restricted to and conserved across insects, contain members that perform essential physiological functions, such as the so-called Halloween genes in the ecdysone biosynthetic pathway (Gilbert, 2004).

Perhaps the least well-characterized clade, CYP4, is well represented in most sequenced insect genomes, which suggests that these genes are also diverse in their function. The paucity of CYP4 genes in the genome of Apis mellifera, the western honey bee, with only 4 represented, compared with the hymenopteran Nasonia vitripennis with 34, also suggests that many of these genes may have become expendable, perhaps with the emergence of eusociality. Within the CYP4 clade, the CYP4G genes form a monophyletic group of genes (Maibeche-Coisne et al., 2000), with one to two genes annotated in each of the 10 sequenced insect genomes (Nelson, 2009), with the lepidopteran domesticated silkworm Bombyx mori genome, containing 3 CYP4G
genes, a notable exception. In two additional lepidopteran species, *Helicoverpa armigera* and *Spodoptera litura*, 3 CYP4G genes have also been identified, while in 15 species from other insect orders only one or two CYP4G genes have been isolated (Fig. 5.1, Table 5.1).

At first glance, a cross-species comparison of the CYP4G genes shows diversity in expression patterns, relating to sensory perception, insecticide resistance, caste differentiation, and diapause. Yet closer examination reveals that CYP4G expression often occurs in sensory organs, such as the antennae, tarsi and integument, or in the target of sensory signals, the brain. In honey bees, CYP4G11 is expressed in the brain (Whitfield et al., 2002), gut (Johnson et al., 2009) and antennae (personal communication, W. Mao). In two bee species, *Apis mellifera* and *Melipona quadrifasciata*, CYP4G expression is higher in foragers than in queens (Judice et al., 2006; Evans and Wheeler 2001). Within the Lepidoptera, CYP4G genes are expressed in the antennae and tarsi of adults. CYP4G28 was cloned from an antennal and tarsal EST library of *Papilio polyxenes* (Patch, 2005). *Mamestra brassicae* expresses CYP4G20 in the sensilla trichodea, the sensory units of the antennae (Maibeche-Coisne et al., 2005), and the silkmoth *Antheraea yamamai* expresses CYP4G25 in the integument of pharate first instars after the initiation of diapause (Yang et al., 2008).

Although the restricted expression and divergence of the CYP4G genes intimate that their role is endogenous in nature, CYP4G genes are also associated with insecticide resistance in some insect pests. CYP4G8 is overexpressed in a pyrethroid-resistant strain of *Helicoverpa armigera* derived from a natural population (Pittendrigh et al., 1997), but not in a laboratory-selected pyrethroid-resistant strain (YGF) (Yang et al., 2006). In a pyrethroid-resistant strain of *Blattella germanica* CYP4G19 expression increases through the stages from nymph to adult stages, with highest expression in the abdomen (Pridgeon et al., 2003). Association with
insecticide resistance does not preclude a sensory or endogenous function, as an insecticidal substrate resembling an endogenous substrate could lead to the selection of these genes for insecticide resistance. Moreover, some forms of resistance involve altered behavior or sensitivity, possibly involving chemosensory processes.

The dual nature of $CYP4G$ genes is exemplified by the CYP4G genes in the genetic model organism, *Drosophila melanogaster*, $Cyp4g1$ and $Cyp4g15$. $Cyp4g1$ is an $\omega$-hydroxylase that regulates triacylglycerol composition, which is necessary for oenocyte control over pupal development (Gutierrez et al. 2007). The ablation of $Cyp4g1$ causes lethality in the pupal stage, by creating an imbalance in fatty acid desaturation. These authors speculate that this imbalance may repress the activity of stearoyl-CoA desaturases, thereby inhibiting inappropriate monosaturated fatty acid synthesis during long non-feeding periods, such as the pupal stage (Gutierrez et al. 2007).

The function of $Cyp4g15$ is unknown, although its expression is restricted to the central nervous system in both larvae and adults (Maibeche-Coisne et al., 2005). This restricted expression suggests a role in *D. melanogaster* neural function or sensory perception. Knocking out $Cyp4g15$ using the UAS:RNAi strain developed by the Vienna *Drosophila* Stock Center is not lethal (Chung et al., 2009), but the behavioral and developmental effects of knockdown have not been reported.

Comparison of $Cyp4g15$ orthologs in 12 *Drosophila* species allowed us to determine if the gene is highly conserved across species, which would suggest a fundamental endogenous role. In addition, certain well-known aspects of the chemical ecology of *D. melanogaster* offer opportunities for examining the role of CYP4G15 in sensory perception. We used the $Cyp4g15$ UAS:RNAi/GAL4 system to determine if knocking down expression affects sensory perception
of and habituation to a known chemical attractant (ethyl acetate) in larvae and the courtship behavior and reproductive success of adult males, which is dependent upon pheromonally-mediated chemical communication (Benton, 2007).

**MATERIALS AND METHODS**

*Chemicals*

HPLC-grade ethyl acetate was purchased from Fisher Scientific (Pittsburgh, PA). Hexadecane was obtained from Sigma Aldrich (St. Louis, MO).

*Drosophila Cyp4g15 and Cyp4g1 ortholog sequence alignment*

Sequence data were downloaded from FlyBase (http://flybase.org/) using the following genome builds: *D. ananassae* (R1.3), *D. erecta* (R1.3), *D. grimshawi* (R1.3), *D. mojavensis* (R1.3), *D. persimilis* (R1.3), *D. pseudoobscura pseudoobscura* (R2.6), *D. sechellia* (R1.3), *D. simulans* (R1.3), *D. virilis* (R1.2), *D. willistoni* (R1.3), and *D. yakuba* (R1.3). Genomic and predicted amino acid sequences of *Cyp4g15* orthologs were collected from each genome using the BLASTn function and the cDNA sequence of *Cyp4g15* from *D. melanogaster*. Amino acid alignments were performed for the 10 full-length predicted sequences using the Clustal X (2.0.11) program.

*Insects*

*D. melanogaster* transgenic lines 8034 and 8035 were created by and purchased from the Vienna *Drosophila* RNAi Center (Dietzl et al., 2007). This line had the inducible UAS-inverted repeat inserted on chromosome 3. For Line 8034, flies are homozygous for the UAS construct, whereas for line 8035 the UAS construct is balanced by a TM3 balancer chromosome. The strain w*; P{w[+mC]=GAL4-elav.L}/TM3,Sb was obtained from Dr. Scott Kreher (Dominican
University, River Forest, IL). The green fluorescent protein (GFP)-expressing strain w*; Sb\(^{1}/TM3, P\{w[+mC]=ActGFP\}JMR2, Ser[1], obtained from the Bloomington *Drosophila* Stock Center.

Flies were reared in plastic bottles or vials on standard medium (Applied Scientific, San Francisco, CA). Flies were transferred to new bottles every two weeks.

*Fly Crosses*

For ethyl acetate larval attractant assays, each strain was re-balanced with the TM# balancer chromosome from the GFP strain, in order to confirm UAS/GAL crosses in the larval stage. The balancer chromosome contains a GFP marker gene that can be detected in larvae using a fluorescence stereomicroscope. Virgin females from the UAS:RNAi/TM3 and elav:GAL4/TM3 lines were collected and crossed with males of w*; Sb\(^{1}/TM3, P\{w[+mC]=ActGFP\}JMR2, Ser[1]. F1 progeny with the genotype GFP, Ser[1]/UAS:RNAi or GFP, Ser[1]/elav:GAL4 were chosen, based on the serrated wing of the Ser marker and the normal bristles indicating the loss of the Sb marker. F1 progeny were crossed to generate offspring in a 1:1:1 ratio of the knockout phenotype UAS:RNAi/elav:GAL4, UAS:RNAi/GFP, and elav:GAL4/GFP. An absence of GFP expression indicated a *Cyp4g15* knockout.

For habituation to ethyl acetate as well as courtship and reproductive success assays, virgin females of the UAS:RNAi strains were crossed to males of elav:GAL4 to generate the knockout phenotype for *Cyp4g15* expression and of UAS:RNAi to serve as a control. Dominant markers for the balancer chromosomes of both strains appear only in the adults. In adults, the dominant marker, *Stubble*, on the balancer chromosome was scored as non-driver control. An absence of the *Stubble* marker indicated a *Cyp4g15* knockout.
**Chemical Attractant Behavioral assay**

The chemical attractant behavioral assay was based on methods previously developed by Kreher et al. (2008). Two filter paper discs were placed diametrically opposed on 1.1% layer of agarose in an 85 mm petri dish. Third instar larvae were collected in a 15% sucrose solution and washed twice in distilled water. 40-50 larvae were placed in the center of the dish between the two filter discs. Ten ul of ethyl acetate, diluted from 10^{-2} to 10^{-5} in hexadecane, was added to one disc, 10 ul of hexadecane was added to the other disc, and the lid was replaced and covered by cardboard lid. Larvae were allowed to move for 5 minutes and then the number of larvae on either half of the plate was scored. Larvae that had not moved from their initial position were excluded from the scoring. Larvae were collected, washed twice in distilled water, and scored for GFP expression using a fluorescence stereomicroscope. The response index is calculated as ((# larvae at ethyl acetate)-( # larvae at hexadecane))/(total # larvae collected), as previously described (Kreher et al., 2008).

**Chemical attractant habituation assay**

The behavioral assay, based on methods previously developed by Boyle and Cobb (2005), comprises of three testing sections: pre-habituation, post-habituation and post-recovery. For each testing section, two filter paper discs were placed diametrically opposite on caps of 1.5 mL microfuge tubes on 1.1% layer of agarose in an 85 mm Petri dish. Third instar larvae were collected as described. Between 40-50 larvae were placed in the center of the dish between the two filter discs and 2.5 µL of the pure ethyl acetate attractant was added to one disc, 2.5 µL of hexadecane was added to the other disc, and the lid was replaced and covered by cardboard lid. Larvae were allowed to move for 5 minutes and then the number of larvae on either half of the
plate was scored. Larvae that had not moved from their initial position were excluded from the scoring. Larvae were collected, washed twice in distilled water, and replaced on diet in order to grow to adults to score for dominant markers.

Male Courtship Behavioral assay

Males of Sb/8034 and elav/8034 were 4 days old. Virgin females of w<sup>1118</sup> were less than 1 day old. The female was placed in the center arena (Fig 5.2) to which a single male was added. Timing began after the female and male showed an interaction that indicated awareness of the other fly. Once the first interaction occurred, the time of the first courtship attempt by the male, indicated by the extension of a wing perpendicular to the body (Greenspan, 1995; Hall, 1994), was recorded. In addition, the total time of courtship was measured within a 5 min period.

Male fecundity measurements

Four-day old virgin females (w<sup>1118</sup>) were allowed to mate for 24 h with a single 3-5 day old male, either elav/8034 or Sb/8034. Males were collected and flash-frozen in liquid nitrogen. Females were transferred to a new vial of diet after 10 days. The number of pupae and resulting adult progeny were each counted for 7 days after the first individual emerged.

Statistical Analysis

A Pearson Chi-square test of independence was used to determine strain differences in the chemical habituation assays; results from two replicated assays, performed on separate days, were pooled to satisfy the requirement for a minimum expected count. The nonparametric Mann-Whitney U test was used to assess whether the measurements of time to courtship and male
fecundity for elav/8034 and Sb/8034 were independent samples from the same distribution (PASW Statistics 17.0, SPSS, Inc., Chicago, IL).

**Molecular modeling and substrate docking in D. melanogaster Cyp4g15**

The *D. melanogaster* Cyp4g15 protein structure was predicted using MOE programs (Chemical Computing Group, Inc., Montreal, Canada) as previously described in Rupasinghe et al. (2003) and Mao et al. (2006) using the template CYP2R1, after showing the greatest sequence identity (21%) in an alignment with 11 templates including CYP175A1 (1N97; Yano et al., 2003), CYP2A6 (1Z10; Yano et al., 2005), CYP2A13 (2P85; Smith et al., 2007), CYP51B1 (1X8V; Podust et al., 2004) and CYP2R1 (2OJD; Strushkevich et al., 2008), CYP2C8 (1PZ2; Schoch et al., 2004), CYP2C9 (1OG5; Williams et al., 2003), CYP3A4 (1TQN; Yano et al., 2004) and CYP102 (2HPD; Ravichandran et al., 1993). After energy minimization with CHARMM22 force field (MacKerell et al., 1998), the model with the best packing score and structure free of aberrations was selected for substrate docking experiments.

The energy-minimized substrate-free protein structure was docked with the substrate molecule, (11Z)-11-octadecen-1-ol, acetate (11-*cis*-vaccenyl acetate, cVA), using the Monte-Carlo docking procedure of MOE with the MMFF94s force field (Halgren, 1996) for the oxygen-free heme as distributed in MOE 2009.1. The substrate was positioned in a potential active site above the heme in preparation for Monte-Carlo simulations. Hundreds of possible conformations were generated and ranked according to the overall internal energy of the ligand. The lowest energy conformation was selected and then included in the protein for a protein/ligand minimization in which the heme coordinates were fixed in order to avoid
distortion of the heme plane due to the lack of bonded parameters for the heme in the MMFF94s force field.

**RESULTS**

*Cyp4g15 orthologs in Drosophila*

Orthologs of *Cyp4g15* occur in the genomes of the 12 *Drosophila* species with sequenced genomes. Ten of the sequenced *Drosophila* species retain full-length *Cyp4g15* orthologs that are nearly identical except for a region between amino acids 290 to 330, for which sequence identity is shared between clades (e.g., the clade comprising *D. simulans*, *D. sechellia*, and *D. melanogaster*) (Fig. 5.3). Comparison with the *Cyp4g1* orthologs (Fig. 5.4) across these species shows greater conservation than *Cyp4g15* orthologs with a region between amino acids 290 and 310 that diverges only in the most basal species (*D. virilis*, *D. mojavensis*). This region is located within exon 4 and does not demarcate an intron splice site junction. The properties of amino acids found within the region vary across species, in which species within the melanogaster subgroup share motifs that differ from those shared by the more basal species. The region is also characterized by tandem repeats, which may represent microsatellites in the genome.

Based on a molecular model of Cyp4g15 this 85-amino acid region extends the inter-helix loops in three positions, between helices E/F, G/H and H/I (Fig. 5.5A). The catalytic site of this model demonstrates that a long-chain fatty acid derivative, such as (11Z)-11-octadecen-1-ol, acetate, sits withing 6.7 Å of the heme (E=-90 kcal/mol) (Fig 5.5B).

*Chemical attractants differ in their repellancy*

There was no significant difference in the response of *Cyp4g15* knockout larvae from non-driver controls to ethyl acetate (Figure 5.6; ANOVA, d.f.=3, F=0.491, p=0.707).
Determining that the perception of the attractant was not altered, we then tested if habituation to ethyl acetate would be changed by knocking out Cyp4g15. Larvae have been shown to habituate to chemical attractants when previously exposed for periods of time ranging from 15-25 minutes (Boyle and Cobb, 2005), but the larval gender was not a factor in those studies. We found that control males were consistently repelled by ethyl acetate after 25 minutes of habituation, whereas knockout males lacked this strong repellent response, which was was marginally significant (Figure 5.7; Pearson Chi-square, 4.026, d.f.=1, p=0.067). Females, control and knockout, were not consistent in their habituation response.

Differences in courtship and reproductive success between male Cyp4g15 knockouts and non-driver controls

There was no significant difference between strains in the time after the first encounter until the first courtship attempt (Table 5.2). Rather, the total time of the courtship was significantly longer in the Cyp4g15 knockout males than in the non-driver controls (Table 5.2). When a single male was allowed to mate with a single female (w1118) for 24 hours, the number of pupae produced (d.f.=1, F=6.345, p<.025) and adult offspring was significantly lower for the Cyp4g15 knockout males than for the non-driver controls (Figure 5.8) (d.f.=1, F=6.582, p<.022).

DISCUSSION

RNAi-driven knockout mutants of D. melanogaster are guiding compasses in the search for the endogenous functions of the large number of genes identified through genome sequences, such as the cytochrome P450 monooxygenases. Because knocking out Cyp4g15 in D. melanogaster is not acutely lethal to the development of the fly, we decided to test for the role of Cyp4g15 in general fitness or reproductive success. Given that Cyp4g15 is expressed in the
brains of larvae and adults (Maibeche-Coisne et al., 2000), and expression is restricted to adult males (Gauhar et al., 2008), I hypothesized that knocking out Cyp4g15 would affect adult male behavior. To address the role of Cyp4g15 in reproductive success, I compared Cyp4g15 UAS:GAL4 knockout flies with their non-driver siblings. I found that in courtship knockout males courted longer than non-driver males with no delay in starting courtship. Yet longer courtship did not translate into higher reproductive success. In a separate assay in which males were housed with females for 24 hours, matings with knockout males produced significantly lower numbers of offspring. This was not due to a decrease in adult eclosion as the number of pupae was also significantly lower for knockout males. Thus, Cyp4g15 potentially plays a role in the mating behavior of D. melanogaster males by modifying their response to females during courtship. Although the absence of Cyp4g15 does not disrupt reproductive success entirely, its function may have evolved in conjunction with the complex mating behavior observed in D. melanogaster.

Courtship and mating behavior in insects requires the integration of multiple sensory inputs to produce a behavior that is then interpreted to elicit a response in the potential partner. Cytochrome P450s in the sensory organs and brain, such as the CYP4G genes, are likely involved in the complicated interplay between perception and reaction by contributing to pheromone clearance, as in the scarab beetle, Phyllopertha diversa, in which a CYP4AW1 and CYP4AW2 P450s in the male antenna metabolize the sex pheromone, an alkaloid, 1,3-dimethyla-2,4-(1H, 3H)-quinazolinedione (Wojtasek and Leal, 1999; Maibeche-Coisne et al. 2004 PNAS). In D. melanogaster, a similar function has been proposed for the cytochrome P450 Cyp6a20, based on male aggression assays (Wang et al., 2008; Dierick and Greenspan, 2006)
and expression in the male antennae (Wang et al., 2008). Although adult brains have been tested, neither male nor female antennae have been screened for expression of Cyp4g15.

Cyp4g15 expression has been detected in the adult brain, but pinpointing its expression to a sexually dimorphic region would help to establish its role in courtship. The antennal lobe, a primary olfactory center in the brain, has been linked to courtship behavior, but initially no sexual dimorphism in size was observed in the glomeruli (Stocker et al., 1990; Laissue et al., 1999). A cross-species comparison revealed that glomeruli (DA1, DL3) in the antennal lobe exhibit sex-specific size differences in D. melanogaster and other Hawaiian Drosophila species (Kondoh et al., 2003).

Surprisingly, one region in the antennal lobe, DA1, shows no sexual dimorphism in its electrophysiological or chemical responses to the D. melanogaster pheromone, cis-vaccenyl acetate (cVA) (Rideout and Goodwin, 2008), even though it is the sole target of the Or67d neurons that mediate the response (Kurtovic et al., 2007). Given that females and males respond differently to cVA, the neurological basis for these differences must depend on a factor other than neuronal response. 11-cis-vaccenyl acetate (cVA) (Fig. 8), the only known volatile pheromone in D. melanogaster, mediates aggregation, mate recognition, and sexual behavior (Jin et al., 2008). cVA, a male-specific volatile pheromone, acts as both a male-male aggression enhancer and a repressor of male courtship of females and an aggregation pheromone (Wang and Anderson, 2010). The way in which it mediates these different outcomes is through courtship and then subsequent oviposition. During courtship, cVA emitted by males attracts females to mate and inhibits male-male courtship. After copulation, cVA is passed to the female where it serves as a repellent to male courtship. cVA is then deposited onto the eggs, which are laid on a food source, and in this capacity it serves potentially as a long-range aggregation pheromone that
may work in tandem with food odors (reviewed in Benton 2007). High densities of males can promote aggression by emitting cVA (Wang and Anderson, 2010).

Molecular modeling of the Cyp4g15 protein shows that cVA, a fatty acid derivative, fits within the predicted catalytic region (Fig. 5.5B). Without heterologous expression data, it is unknown if Cyp4g15 catalyzes the production or breakdown of cVA. Given that cVA production occurs in the male ejaculatory bulb (Briether and Butterworth, 1970), it is unlikely that Cyp4g15 is involved in its synthesis. If Cyp4g15 catalyzes the breakdown of cVA, then Cyp4g15 knock-out males may become over-stimulated by cVA and not respond normally to the repellent properties that signal a mated female.

Molecular modeling also identifies an 85-amino acid region that is unique among all cytochrome P450 structures studied (pers. comm., Sanjeewa Rupasinghe). This region is predicted to extend three interhelical loops on the same side of the protein. The function of this extended region is unknown but could be involved in protein-protein interactions, such as those observed in metabolon formation of plant P450s (Ralston and Yu, 2006).

Using comparative genomics of the 12 related Drosophila species, I have identified Cyp4g15 to be highly conserved among 10 of them. Given this high degree of conservation in sequence among the Cyp4g15 orthologs, these proteins likely share a similar substrate. Demonstration that Cyp4g15 shows similar expression in the male adult brain across species would support the prediction of conserved function based on conserved amino acid sequence for this P450. While parallel RNAi experiments could not be conducted in the other 11 species, the combined power of D. melanogaster genetics and 12 closely related sequenced genomes offers a new avenue for predicting P450 protein function.
ACKNOWLEDGMENTS

I thank Henry Pollock for performing the fly courtship assays and Dr. Sanjeeva Rupasinghe for creating the molecular model and docking of the Cyp4g15 protein. Dr. Arthur Zangerl assisted with statistical analysis. Jungkoo Kim participated in discussions on the methods and results. Use of the fluorescence stereomicroscope was kindly provided by Dr. Lisa Stubbs in the Institute for Genomic Biology (University of Illinois Urbana-Champaign). This work was funded by a USDA NRI grant to Dr. May Berenbaum.

LITERATURE CITED


Figure 5.1 Genetic distance tree produced by neighbor-joining analysis of the amino acid sequences. Bootstrap supports (1000 replicates, reported in %) for the clade are indicated at each node. All CYP4G full length sequences were included in the analysis. Protein name and species source are indicated at the terminal nodes. Two monophyletic clades were identified and named for the *D. melanogaster* genes within each clade. Genes from the orders Lepidoptera (blue), Diptera (red), Hymenoptera (green) and Coleoptera (blue) are delineated by colored boxes. Species names are abbreviated as follows: *Manduca sexta* (Msex), *Antheraea yamamai* (Ayam), *Antheraea pernyi* (Aper), *Bombyx mori* (Bmor), *Zygaena filipendulae* (Zfil), *Helicoverpa armigera* (Harm), *Mamestra brassicaceae* (Mbra), *Drosophila melanogaster* (Dmel), *Drosophila pseudoobscura* (Dpse), *Aedes aegypti* (Aaeg), *Trichogramma cacoeciae* (Tcac), *Nasonia vitripennis* (Nvit), *Apis mellifera* (Amel), *Acythosiphon pisum* (Apis), *Diabrotica virgifera* (Dvir), *Leptinotarsa decemlineata* (Ldec), *Ips paraconfusus* (Ipar), *Tribolium castaneum* (Tcas), *Blatella germanica* (Bger), *Pediculus humanus humanus* (Phum), *Chironomus tentans* (Cten), *Musca domestica* (Mdom), *Anopheles gambiae* (Agam), *Anopheles funestus* (Afun).
Figure 5.2 Configuration of the courtship arena. Circular arena in center is 6.2 cm. A Plexiglas lid covers the entire set-up. Holes in the lid on the left and right circular arenas are 0.2 cm. A female was placed in the center arena and covered with the lid. A male was inserted through one of the holes in the lid that entered a side arena. This arena was separated by a revolving door that can be turned to open into the center arena.
Figure 5.3 Alignment of Cyp4g15 amino acid sequences from 10 *Drosophila* species. Genomic and predicted amino acid sequences of Cyp4g15 orthologs were identified in each genome using the BLASTn function and the cDNA sequence of Cyp4g15 from *D. melanogaster*. Amino acid alignments were performed for the 10 full-length predicted sequences using the Clustal X (2.0.11) program. The 10 species with full length Cyp4g15 protein sequences are *D. mojavensis* (Dmoj), *D. virilis* (Dvir), *D. grimshawi* (Dgri), *D. sechellia* (Dsec), *D. simulans* (Dsim), *D. melanogaster* (Dmel), *D. yakuba* (Dyak), *D. erecta* (Dere), *D. pseudoobscura* (Dpse), and *D. willistoni* (Dwil). The level of sequence similarity across species is indicated in gray below the sequences. Sequence similarity is high throughout the protein sequence except for a region between amino acid positions 290-330.
Figure 5.4 Alignment of Cyp4g1 amino acid sequences from 12 Drosophila species. Genomic and predicted amino acid sequences of Cyp4g1 orthologs were identified in each genome using the BLASTn function and the cDNA sequence of Cyp4g1 from D. melanogaster. Amino acid alignments were performed for the 12 full-length predicted sequences using the Clustal X (2.0.11) program. The 12 species are D. mojavensis (Dmoj), D. virilis (Dvir), D. grimshawi (Dgri), D. sechellia (Dsec), D. simulans (Dsim), D. melanogaster (Dmel), D. yakuba (Dyak), D. ananassae (Dana), D. erecta (Dere), D. persimilis (Dper), D. pseudoobscura (Dpse), and D. willistoni (Dwil).
Figure 5.5 Molecular model of Cyp4g15. (A) The molecular model of Cyp4g15 using the crystal structure of CYP2R1 (Strushkevich et al., 2008) as the template. Cyp4g15 contains 45 additional amino acids that are not found in any of the known P450 crystal structures. These amino acids were successfully incorporated into the model and shown to extend interhelical loops in three regions indicated, the Helix E-Helix F loops, the Helix G-Helix H loop and the Helix H-Helix I loop. These loops are found on the opposite face of the protein from the P450 reductase interacting domain.

(B) Cyp4g15 P450 model docked with cis-vaccenyl acetate. A ball-and-stick model of cis-vaccenyl acetate sits above the heme moiety (gray space-filling model) in the predicted catalytic site of Cyp4g15. Three substrate recognition sites (SRS1, SRS5 and SRS4) are shown. The substrate fits within the putative active site of the enzyme, as designated by blue dots, and sits 6.7 Å (≈90 kcal/mol) to double-bonded hydrogen of the substrate.
Figure 5.6 Larval response index of *Cyp4g15* knockouts and GFP controls to increasing doses of ethyl acetate. Between 40-50 larvae were tested for attraction to ethyl acetate, diluted 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2} in hexadecane compared to hexadecane alone. After 5 min, larvae that migrated either side were collected and scored for GFP expression using a fluorescence stereomicroscope. The response index is calculated as ((# larvae at ethyl acetate)-(# larvae at hexadecane))/(total # larvae collected) as previously described (Kreher et al., 2008).
Figure 5.7 Larval response index of \textit{Cyp4g15} knockouts and non-driver controls, for males and females, that were scored as adults. For the habituation procedure, 40-50 larvae were placed in a Petri dish with 40 μL ethyl acetate for 25 min as previously described (Boyle and Cobb, 2005). Larvae were immediately tested for attraction to 2.5 μL ethyl acetate (pure) or hexadecane. Larvae were collected and replaced on diet in order to grow to adulthood to score for dominant markers. The response index is calculated as \[ \frac{(# \text{ larvae at ethyl acetate}) - (# \text{ larvae at hexadecane})}{\text{total # larvae collected}} \] as previously described (Kreher et al., 2008). Non-driver control males showed a more negative response to ethyl acetate after habituation than \textit{Cyp4g15} knockout males. Female response was not consistent across trials.
Figure 5.8 Mean number of adult offspring (± s.d.) of Cyp4g15 knockout (N=8) and non-driver control (N=8) males paired with a 3-day old virgin w^{1118} females. Number of adult offspring and the number of pupae (not shown) were recorded after 7 days of egg-laying by mated females. Mean number of adult offspring are indicated was significantly lower for the Cyp4g15 knockout male than for the non-driver control (d.f.=1, F=6.582, p<.022).
# TABLES

Table 5.1 CYP4G genes in the P450 Database curated by Nelson (2009). Species in bold have sequenced genomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>CYP4 genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DIPTERA</strong></td>
<td></td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>Cyp4g1, Cyp4g15</td>
</tr>
<tr>
<td><em>D. pseudoobscura</em></td>
<td>Cyp4g1, Cyp4g15</td>
</tr>
<tr>
<td><em>Musca domestica</em></td>
<td>CYP4G2v1, v2, CYP4G13v1, v2</td>
</tr>
<tr>
<td><em>Anopheles gambiae</em></td>
<td>CYP4G16, CYP4G17</td>
</tr>
<tr>
<td><em>Aedes aegypti</em></td>
<td>CYP4G35, CYP4G36</td>
</tr>
<tr>
<td><em>Anopheles funestus</em></td>
<td>CYP4G21</td>
</tr>
<tr>
<td><em>Chironomus tentans</em></td>
<td>CYP4G33</td>
</tr>
<tr>
<td><strong>COLEOPTERA</strong></td>
<td></td>
</tr>
<tr>
<td><em>Tribolium castaneum</em></td>
<td>CYP4G7, CYP4G14</td>
</tr>
<tr>
<td><em>Diabrotica virgifera</em></td>
<td>CYP4G18</td>
</tr>
<tr>
<td><em>Ips paraconfusus</em></td>
<td>CYP4G27</td>
</tr>
<tr>
<td><em>Leptinotarsa decemlineata</em></td>
<td>CYP4G29, CYP4G34</td>
</tr>
<tr>
<td><strong>LEPIDOPTERA</strong></td>
<td></td>
</tr>
<tr>
<td><em>Manduca sexta</em></td>
<td>CYP4G4, CYP4G49</td>
</tr>
<tr>
<td><em>Helicoverpa armigera</em></td>
<td>CYP4G8v1, CYP4G9, CYP4G10</td>
</tr>
<tr>
<td><em>Mamestra brassicae</em></td>
<td>CYP4G20</td>
</tr>
<tr>
<td><em>Bombyx mori</em></td>
<td>CYP4G22, CYP4G23, CYP4G24</td>
</tr>
<tr>
<td><em>Spodoptera litura</em></td>
<td>CYP4G30, CYP4G31, CYP4G32</td>
</tr>
<tr>
<td><em>Spodoptera exigua</em></td>
<td>CYP4G37</td>
</tr>
<tr>
<td><em>Antheraea yamamai</em></td>
<td>CYP4G25</td>
</tr>
<tr>
<td><em>Antheraea pernyi</em></td>
<td>CYP4G25</td>
</tr>
<tr>
<td><em>Zygaena filipendulae</em></td>
<td>CYP4G47, CYP4G48</td>
</tr>
<tr>
<td><strong>HYMENOPTERA</strong></td>
<td></td>
</tr>
<tr>
<td><em>Apis mellifera</em></td>
<td>CYP4G11</td>
</tr>
<tr>
<td><em>Trichogramma cacoeciae</em></td>
<td>CYP4G12</td>
</tr>
<tr>
<td><em>Nasonia vitripennis</em></td>
<td>CYP4G43, CYP4G44</td>
</tr>
<tr>
<td><em>Melipona quadri fasciata</em></td>
<td>CYP4G11</td>
</tr>
<tr>
<td><strong>ANOPLURA</strong></td>
<td></td>
</tr>
<tr>
<td><em>Pediculus humanus</em></td>
<td>CYP4G38, CYP4G39</td>
</tr>
<tr>
<td><strong>BLATTODEA</strong></td>
<td></td>
</tr>
<tr>
<td><em>Blattella germanica</em></td>
<td>CYP4G19</td>
</tr>
<tr>
<td><strong>HEMIPTERA</strong> (HOMOPTERA)</td>
<td></td>
</tr>
<tr>
<td><em>Acyrthosiphon pisum</em></td>
<td>CYP4G51</td>
</tr>
</tbody>
</table>
Table 5.2 Mann-Whitney U analysis of mean ranks and significance difference between strains in male courtship measurements

<table>
<thead>
<tr>
<th>Strain</th>
<th>N</th>
<th>Time to first courtship (s)</th>
<th>Total length of time of courtship (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Ranks</td>
<td>Significance</td>
</tr>
<tr>
<td>Cyp4g15 knockout</td>
<td>5</td>
<td>5.90</td>
<td>0.283</td>
</tr>
<tr>
<td>Non-driver control</td>
<td>9</td>
<td>8.39</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX A

DDT toxicity assays in *Drosophila melanogaster* (y; cn bw sp) and 8 *Drosophila* species

**INTRODUCTION**

In this section, I describe the results of a DDT toxicity assays that I performed for the reference strain of *Drosophila melanogaster* and 8 additional sequenced strains of *Drosophila* species. I then conducted DDT toxicity analysis to determine the lethal concentrations (LC) for 25% and 3% of individuals tested within each species.

**MATERIALS AND METHODS**

*Fly Lines*

*Drosophila simulans*, *D. sechellia*, *D. erecta*, *D. yakuba*, *D. ananassae*, *D. willistoni*, *D. pseudoobscura*, *D. persimilis*, *D. mojavensis*, and *D. virilis* were obtained from the UC San Diego Drosophila Stock Center (San Diego, CA). *Drosophila melanogaster* strain *y; cn bw sp* was obtained from the Bloomington Drosophila Stock Center (Bloomington, IN).

*D. simulans*, *D. erecta*, *D. yakuba*, *D. ananassae*, *D. willistoni*, and *D. virilis* were maintained on standard medium. *D. sechellia* was raised on standard medium with a small piece of noni leather (Hawaiian Health Ohana, Anahola, HI) in each vial. *D. pseudoobscura*, *D. persimilis* and *D. mojavensis* were reared on the banana-opuntia cactus diet recommended by the UC San Diego Drosophila Stock Center. *Drosophila melanogaster* strains were reared in plastic bottles or vials on standard medium (Applied Scientific, San Francisco, CA). Flies were transferred to new bottles every 2 weeks.

*Bioassays*

Adult bioassays for the 12 sequenced species were performed as described by Brandt et al. (2002). Scintillation vials (Fisher Scientific) were coated on the inside surface with DDT in the following manner: a respective quantity of insecticide was placed in a vial in a constant volume of 200 µL acetone. The vials were continuously rolled and dried overnight until the acetone had evaporated. Ten to thirty flies (1-3 days old) were placed in vials, which were sealed with cotton balls moistened with 2 mL of 5% sucrose solution to provide the insects with food and water. Twenty-four hours after placement in the vials, insects that did not move were scored as dead. Raw bioassay data were analyzed by probit analysis in SAS 9.1 (SAS Institute, 2003).

**RESULTS**

*Analysis of DDT toxicity amongst Drosophila species*

The most susceptible species tested were *D. mojavensis* and *D. yakuba*, but their level of susceptibility did not differ significantly from *D. persimilis*, *D. willistoni*, *D. ananassae*, *D. pseudoobscura* (Table A.2). *D. melanogaster* (y; cn bw sp) and *D. simulans* were the least susceptible species tested (Table A.2). Susceptibility to DDT did not correlate with the presence of both *Cyp12d1/Cyp12d1* genes, but based on the three least susceptibility, seemed to depend on larger body size (*D. virilis*) and cosmopolitan distribution (*D. melanogaster*, *D. simulans*).
**TABLES**

Table A.1  Probit analysis results of DDT toxicity in 9 sequenced species of *Drosophila*

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>Slope (± SE)</th>
<th>LC₃ (95% CI)</th>
<th>LC₂₅ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. mojavensis</em></td>
<td>1005</td>
<td>1.66 (0.36)</td>
<td>0.2 (0.02-0.43)</td>
<td>0.9 (0.3-1.5)</td>
</tr>
<tr>
<td><em>D. yakuba</em></td>
<td>2112</td>
<td>1.56 (0.33)</td>
<td>0.2 (0.02-0.42)</td>
<td>1.0 (0.4-1.6)</td>
</tr>
<tr>
<td><em>D. persimilis</em></td>
<td>877</td>
<td>2.48 (0.45)</td>
<td>0.4 (0.12-0.68)</td>
<td>1.2 (0.6-1.6)</td>
</tr>
<tr>
<td><em>D. willistoni</em></td>
<td>1440</td>
<td>2.73 (0.48)</td>
<td>0.5 (0.17-0.76)</td>
<td>1.3 (0.8-1.7)</td>
</tr>
<tr>
<td><em>D. ananassae</em></td>
<td>961</td>
<td>2.93 (0.43)</td>
<td>0.5 (0.25-0.79)</td>
<td>1.3 (0.9-1.7)</td>
</tr>
<tr>
<td><em>D. pseudoobscura</em></td>
<td>974</td>
<td>3.05 (0.46)</td>
<td>0.6 (0.28-0.86)</td>
<td>1.4 (1.0-1.8)</td>
</tr>
<tr>
<td><em>D. virilis</em></td>
<td>391</td>
<td>2.16 (0.46)</td>
<td>0.7 (0.17-1.33)</td>
<td>2.7 (1.5-3.6)</td>
</tr>
<tr>
<td><em>D. simulans</em></td>
<td>2057</td>
<td>2.26 (0.40)</td>
<td>0.9 (0.38-1.46)</td>
<td>3.2 (2.3-4.1)</td>
</tr>
<tr>
<td><em>D. melanogaster (y;cn bw sp)</em></td>
<td>1710</td>
<td>2.55 (0.43)</td>
<td>1.2 (0.59-1.76)</td>
<td>3.6 (2.7-4.5)</td>
</tr>
</tbody>
</table>
Appendix B

Cloning of Cyp12d1, Adrenodoxin Reductase and Adrenodoxin transcripts from Drosophila melanogaster for protein expression

This section reports the results of the cloning of three genes, Cyp12d1, adrenodoxin reductase and adrenodoxin, from Drosophila melanogaster for bacterial expression of Cyp12d1. For Cyp12d1, primers were designed for the 5' and 3' end of the coding region using the sequence of Cyp12d1 from the reference genome (strain y; cn bw sp). For directional cloning into the pCWori expression plasmid, the forward and reverse primers (Table B.1) were designed with an Nde I and Hind III restriction enzyme recognition sites, respectively. The forward primer also contained the first 10 codons of the Cyp12d1 gene for which codon bias was adjusted to favor bacterial expression (Fig. B.1). Amplification products were cloned into the pGEM t-easy vector (Promega, Madison, WI) for which the standard primers T7 promoter (forward) and M13 (reverse) were used for sequencing. Additional primers were created to sequence the Cyp12d1 transcripts (Table B.2). Two full-length clones from the D. melanogaster strains Canton-S (CanS2, CanS3) and Oregon-R (OreR3, OreR4) were sequenced (Fig. B.1), and their predicted amino acid sequences showed that they were more similar to CYP12D1 than CYP12D2 from the reference strain (Fig. B.2). Five full-length sequences of Adrenodoxin from the D. melanogaster strains Rst(1)JH1 (Rst2, Rst3) and br1 (br2, br3, br4) were cloned (Fig. B.3). A single full-length sequence of Adrenodoxin Reductase was cloned from the D. melanogaster strain Oregon-R (Fig. B.4).
### TABLES

Table B.1 Primers used to amplify *Cyp12d1*, Adrenodoxin Reductase (AdR) and Adrenodoxin (Ad) from *D. melanogaster*

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp12d1-NdeI-for</td>
<td>AAGGGAATTCCTATGGCTAACACCTGTGACGCACGCCTGTCTGTG</td>
</tr>
<tr>
<td>Cyp12d1-HindII-rev</td>
<td>GGGGAAGCTTTTATGATGATGATGTCGATACCCGTGAAT</td>
</tr>
<tr>
<td>AdR-EcoRI-for</td>
<td>GGAATTCTTAAGAAGGAGATATGCCATGCACTCATCAACTCTGCCHGCGTCAATCCCA</td>
</tr>
<tr>
<td>AdR-BamHI-rev</td>
<td>CGGGATCCCTATGATGACGCGACGCCAAGCGGCGCGCGATCCGAATCTCGT</td>
</tr>
<tr>
<td>Ad-EcoRI-for</td>
<td>GGAATTCTTTAAGAAGGAGATATCCATGCGATCAACTCTGCCACCGCTCGT</td>
</tr>
<tr>
<td>Ad-BamHI-rev</td>
<td>CGGGATCCCTAGCGACGTCGTCATGAT</td>
</tr>
</tbody>
</table>

1Restriction enzyme sites indicated in primer name are underlined

Table B.2 Primers created to sequence *Cyp12d1* clones

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CanS4-Cyp12d1 for</td>
<td>TATCAATATGAGTTTATAGAG</td>
</tr>
<tr>
<td>Cyp12d1-CanS rev</td>
<td>CTCTATAACTCTATTATGATA</td>
</tr>
<tr>
<td>Cyp12d1-OreR rev</td>
<td>CTGATTTTCCTACGACATTTTCTG</td>
</tr>
</tbody>
</table>
Figure B.1 Alignment of nucleotide sequences of Cyp12d1 cDNA cloned from the D. melanogaster strains Canton-S (CanS2, CanS3) and Oregon-R (OreR3, OreR4) with Cyp12d1 from the reference strain (y; cn bw sp; AE003827). Sequences are color-coded and follow the order of the sequences names in the first line. The cloned sequences differ from the Cyp12d1 sequence of the reference strain (underlined) due to base changes inserted into the Cyp12d1 cloning primers that optimize the amino acid sequence for bacterial expression.
ACAAAATGAGGCATGGGGAAAACTACGATCAGCCATTAATCCCATCTTTATGCAACCCAGGGGCTTGAGAATGTATTATGAACCATTGTCTAATATCAATAATGCGT
ACAAAATGAGGCATGGGGAAAACTACGATCAGCCATTAATCCCATCTTTATGCAACCCAGGGGCTTGAGAATGTATTATGAACCATTGTCTAATATCAATAATGAGT
ACAAAATGAGGCATGGGGAAAAACTACGATCAGCCATTAATCCCATCTTTATGCAACCCAGGGGCTTGAGAATGTATTATGAACCATTGTCTAATATCAATAATGAGT
ACAAAATGAGGCATGGGGAAAAACTACGATCAGCCATTAATCCCATCTTTATGCAACCCAGGGGCTTGAGAATGTATTATGAACCATTGTCTAATATCAATAATGAGT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT
TTATAGACCCGATTTAGGAAATTCGGCGATTTAAAAACTCTAGATCATCCCCGAAGATTTTTCGGAATAGGAAATACCCGAGCTGTTTTTCGAGTCACTCGGCCTGGTGCT

Fig. B.1 (cont.)
Alignment of predicted amino acid sequences of CYP12D1 from the *D. melanogaster* strains, Canton-S (CanS2) and Oregon-R (OreR3, OreR4) with CYP12D1 and CYP12D2 from the reference strain (*y; cn bw sp*). Differences between the cloned sequences and the reference sequences are underlined. Strain-specific amino acid differences are in bold. The cloned sequences share amino acid similarity with CYP12D1 in three positions (underlined, bolded).

Figure B.2
OreR3_12d1  NSMLERLMEIDPKVAVIMSLDILFAGVDATATLLSAVLLCLSKHPDKQAK
OreR4_12d1  NSMLERLMEIDPKVAVIMSLDILFAGVDATATLLSAVLLCLSKHPDKQAK
CYP12D2  NSMLERLMEIDPKVAVIMSLDILFAGVDATATLLSAVLLCLSKHPDKQAK
CYP12D1  NSMLERLMEIDPKVAVIMSLDILFAGVDATATLLSAVLLCLSKHPDKQAK
CanS2_12d1  NSMLERLMEIDPKVAVIMSLDILFAGVDATATLLSAVLLCLSKHPDKQAK

OreR3_12d1  LREELLSIMPTKDSLLEENMKDMPYLRAVIKETLRYYPNGLGTMRTCQN
OreR4_12d1  LREELLSIMPTKDSLLEENMKDMPYLRAVIKETLRYYPNGLGTMRTCQN
CYP12D2  LREELLSIMPTKDSLLEENMKDMPYLRAVIKETLRYYPNGLGTMRTCQN
CYP12D1  LREELLSIMPTKDSLLEENMKDMPYLRAVIKETLRYYPNGLGTMRTCQN
CanS2_12d1  LREELLSIMPTKDSLLEENMKDMPYLRAVIKETLRYYPNGLGTMRTCQN

OreR3_12d1  DVILSGYRVPKKTVLLGSNVLMKEATYYPRPDEFLPERWLDPETGKKM
OreR4_12d1  DVILSGYRVPKKTVLLGSNVLMKEATYYPRPDEFLPERWLDPETGKKM
CYP12D2  DVILSGYRVPKKTVLLGSNVLMKEATYYPRPDEFLPERWLDPETGKKM
CYP12D1  DVILSGYRVPKKTVLLGSNVLMKEATYYPRPDEFLPERWLDPETGKKM
CanS2_12d1  DVILSGYRVPKKTVLLGSNVLMKEATYYPRPDEFLPERWLDPETGKKM

OreR3_12d1  QVSPFTFLPGFGPRMCIGKRVVDLEMETTVAKLIRNFHVEFNRDASRF
OreR4_12d1  QVSPFTFLPGFGPRMCIGKRVVDLEMETTVAKLIRNFHVEFNRDASRF
CYP12D2  QVSPFTFLPGFGPRMCIGKRVVDLEMETTVAKLIRNFHVEFNRDASRF
CYP12D1  QVSPFTFLPGFGPRMCIGKRVVDLEMETTVAKLIRNFHVEFNRDASRF
CanS2_12d1  QVSPFTFLPGFGPRMCIGKRVVDLEMETTVAKLIRNFHVEFNRDASRF

OreR3_12d1  KTMFVMEPAITFFKFTDIEQHHHAAPA
OreR4_12d1  KTMFVMEPAITFFKFTDIEQHHHAAPA
CYP12D2  KTMFVMEPAITFFKFTDIEQ--
CYP12D1  KTMFVMEPAITFFKFTDIEQ--
CanS2_12d1  KTMFVMEPAITFFKFTDINNTTRRRP

Fig B.2 (cont.)
<table>
<thead>
<tr>
<th>Strain</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rst3-Ad</td>
<td>ATGCTAGTCATAAAGCTCATGCGGATGGTGCTGCTCCAGATGTTGCGTACGAGCTCAATTTC</td>
</tr>
<tr>
<td>Rst2-Ad</td>
<td>ATGCCATCATAAAGCTCATGCGGATGGTGCTGCTCCAGATGTTGCGTACGAGCTCAATTTC</td>
</tr>
<tr>
<td>br3-Ad</td>
<td>ATGCCATCATAAAGCTCATGCGGATGGTGCTGCTCCAGATGTTGCGTACGAGCTCAATTTC</td>
</tr>
<tr>
<td>br2-Ad</td>
<td>ATGCCATCATAAAGCTCATGCGGATGGTGCTGCTCCAGATGTTGCGTACGAGCTCAATTTC</td>
</tr>
<tr>
<td>br4-Ad</td>
<td>ATGCCATCATAAAGCTCATGCGGATGGTGCTGCTCCAGATGTTGCGTACGAGCTCAATTTC</td>
</tr>
<tr>
<td>Rst3-Ad</td>
<td>CGATCGCCGATCGCAACGCGTACTTTCTCGACGGGATTGGCCCTGAAAACGAAAGATGTT</td>
</tr>
<tr>
<td>Rst2-Ad</td>
<td>CGATCGCCGATCGCAACGCGTACTTTCTCGACGGGATTGGCCCTGAAAACGAAAGATGTT</td>
</tr>
<tr>
<td>br3-Ad</td>
<td>CGATCGCCGATCGCAACGCGTACTTTCTCGACGGGATTGGCCCTGAAAACGAAAGATGTT</td>
</tr>
<tr>
<td>br2-Ad</td>
<td>CGATCGCCGATCGCAACGCGTACTTTCTCGACGGGATTGGCCCTGAAAACGAAAGATGTT</td>
</tr>
<tr>
<td>br4-Ad</td>
<td>CGATCGCCGATCGCAACGCGTACTTTCTCGACGGGATTGGCCCTGAAAACGAAAGATGTT</td>
</tr>
<tr>
<td>Rst3-Ad</td>
<td>GTTAACATAACCTTCGTGCGTGCCAATGGGGACAAGATGTAAGACGCTCCGAAAAGTGGGT</td>
</tr>
<tr>
<td>Rst2-Ad</td>
<td>GTTAACATAACCTTCGTGCGTGCCAATGGGGACAAGATGTAAGACGCTCCGAAAAGTGGGT</td>
</tr>
<tr>
<td>br3-Ad</td>
<td>GTTAACATAACCTTCGTGCGTGCCAATGGGGACAAGATGTAAGACGCTCCGAAAAGTGGGT</td>
</tr>
<tr>
<td>br2-Ad</td>
<td>GTTAACATAACCTTCGTGCGTGCCAATGGGGACAAGATGTAAGACGCTCCGAAAAGTGGGT</td>
</tr>
<tr>
<td>br4-Ad</td>
<td>GTTAACATAACCTTCGTGCGTGCCAATGGGGACAAGATGTAAGACGCTCCGAAAAGTGGGT</td>
</tr>
<tr>
<td>Rst3-Ad</td>
<td>GACTCCCTGCTGGACCTGGTGTGGTCAACAACAAATGTGAGATTGATGTTGCTGCTGCTGCTGGGCTGCT</td>
</tr>
<tr>
<td>Rst2-Ad</td>
<td>GACTCCCTGCTGGACCTGGTGTGGTCAACAACAAATGTGAGATTGATGTTGCTGCTGCTGCTGGGCTGCT</td>
</tr>
<tr>
<td>br3-Ad</td>
<td>GACTCCCTGCTGGACCTGGTGTGGTCAACAACAAATGTGAGATTGATGTTGCTGCTGCTGCTGGGCTGCT</td>
</tr>
<tr>
<td>br2-Ad</td>
<td>GACTCCCTGCTGGACCTGGTGTGGTCAACAACAAATGTGAGATTGATGTTGCTGCTGCTGCTGGGCTGCT</td>
</tr>
<tr>
<td>br4-Ad</td>
<td>GACTCCCTGCTGGACCTGGTGTGGTCAACAACAAATGTGAGATTGATGTTGCTGCTGCTGCTGGGCTGCT</td>
</tr>
<tr>
<td>Rst3-Ad</td>
<td>GAGGGCACGCTGACCTGCTCCACCTGCCACCTGATCTTCAAGACCAGCCGATTTGAGAAA</td>
</tr>
<tr>
<td>Rst2-Ad</td>
<td>GAGGGCACGCTGACCTGCTCCACCTGCCACCTGATCTTCAAGACCAGCCGATTTGAGAAA</td>
</tr>
<tr>
<td>br3-Ad</td>
<td>GAGGGCACGCTGACCTGCTCCACCTGCCACCTGATCTTCAAGACCAGCCGATTTGAGAAA</td>
</tr>
<tr>
<td>br2-Ad</td>
<td>GAGGGCACGCTGACCTGCTCCACCTGCCACCTGATCTTCAAGACCAGCCGATTTGAGAAA</td>
</tr>
<tr>
<td>br4-Ad</td>
<td>GAGGGCACGCTGACCTGCTCCACCTGCCACCTGATCTTCAAGACCAGCCGATTTGAGAAA</td>
</tr>
</tbody>
</table>

Figure B.3: Alignment of nucleotide sequences of Adrenodoxin cDNA cloned from the *D. melanogaster* strains Rst(1)JH<sup>1</sup> (Rst2, Rst3) and br<sup>1</sup> (br2, br3, br4) with Adrenodoxin from the reference strain (*y; cn bw sp*; AE003480). Sequences are in blue and those that differ are bolded. The cloned sequences differ from the 5’ end of *Cyp12d1* sequence of the reference strain (underlined) due to base changes inserted into the *Cyp12d1* cloning primers that optimize the amino acid sequence for bacterial expression. At the 3’ end of the clones sequences, a histidine tag (bold, underlined) was added in the primer.
Adrenodoxin_cDNA CTGCCCGACAAACCCTGGATGAGGAGCTGGACATGCTGGATTTGGCGTACGAACTGACG
Rst3-Ad CTGCCCGACAAACCCTGGATGAGGAGCTGGACATGCTGGATTTGGCGTACGAACTGACG
Rst2-Ad CTGCCCGACAAACCCTGGATGAGGAGCTGGACATGCTGGATTTGGCGTACGAACTGACG
br3-Ad CTGCCCGACAAACCCTGGATGAGGAGCTGGACATGCTGGATTTGGCGTACGAACTGACG
br2-Ad CTGCCCGACAAACCCTGGATGAGGAGCTGGACATGCTGGATTTGGCGTACGAACTGACG
br4-Ad CTGCCCGACAAACCCTGGATGAGGAGCTGGACATGCTGGATTTGGCGTACGAACTGACG

Adrenodoxin_cDNA GACACCTCGCGCTGGGCTGCCAGATCACCCTGTCCAAGGATATGGAGGGCCTGGAGGTA
Rst3-Ad GACACCTCGCGCTGGGCTGCCAGATCACCCTGTCCAAGGATATGGAGGGCCTGGAGGTA
Rst2-Ad GACACCTCGCGCTGGGCTGCCAGATCACCCTGTCCAAGGATATGGAGGGCCTGGAGGTA
br3-Ad GACACCTCGCGCTGGGCTGCCAGATCACCCTGTCCAAGGATATGGAGGGCCTGGAGGTA
br2-Ad GACACCTCGCGCTGGGCTGCCAGATCACCCTGTCCAAGGATATGGAGGGCCTGGAGGTA
br4-Ad GACACCTCGCGCTGGGCTGCCAGATCACCCTGTCCAAGGATATGGAGGGCCTGGAGGTA

Adrenodoxin_cDNA CATGTGCCCTCCACCATCAATGACGCACGTGCCGCG
Rst3-Ad CATGTGCCCTCCACCATCAATGACGCACGTGCCGCG
Rst2-Ad CATGTGCCCTCCACCATCAATGACGCACGTGCCGCG
br3-Ad CATGTGCCCTCCACCATCAATGACGCACGTGCCGCG
br2-Ad CATGTGCCCTCCACCATCAATGACGCACGTGCCGCG
br4-Ad CATGTGCCCTCCACCATCAATGACGCACGTGCCGCG

Fig. B.3 (cont.)
Figure B.4 Alignment of nucleotide sequences of Adrenodoxin cDNA cloned from the *D. melanogaster* strains Oregon-R (OreR) with Adrenodoxin reductase from the reference strain (y; cn bw sp; AE003826). Sequences are in blue are conserved across strains, and those that differ are bolded. The cloned sequences differ from the 5’ end of *Cyp12d1* sequence of the reference strain (underlined) due to base changes inserted into the *Cyp12d1* cloning primers that optimize the amino acid sequence for bacterial expression.
Fig. B.4 (cont.)
OreR-AdR
Adrenodoxin reductase
ACCCTCTGGCATTGACATAACGAAATGCTCTGGACACCAGTTCTGTCAA

OreR-AdR
Adrenodoxin reductase
ACCATCTGGCATTGACATAACGAAATGCTCTGGACACCAGTTCTGTCAA

OreR-AdR
Adrenodoxin reductase
ACCAGGATACGATGGGAAACGAGTGTTACTTGGGATGGCTGGC

OreR-AdR
Adrenodoxin reductase
AGAGAATCAATGATTTCGAGAGCGCAGCGGGAAAAGCCAAGGGAAAGCCG

OreR-AdR
Adrenodoxin reductase
CGCGAGAAGATTGTTAGCATTGAGGAAATGTACGGGTGGCTGGCGTC

Fig. B.4 (cont.)
AUTHOR’S BIOGRAPHY

Cynthia Marie McDonnell was born and raised in Illinois, where she grew up in a suburb of Chicago called Libertyville. She left Illinois for college at Cornell University in New York, where she first developed her interest using insects to study evolution. In college she undertook an independent project to study an ant-homopteran mutualism. It was at Cornell University that she first heard about Dr. May Berenbaum’s work in chemical ecology and it was Tom Eisner, one of the pioneers in the field of chemical ecology, who told her if she wanted to study chemical ecology, go to University of Illinois. This was a bit disappointing to hear for an Illinois native with a case of wanderlust, so before applying to graduate school, she took a 6 month internship at the Bermuda Biological Station for Research, where she studied nitric oxide synthase in the intracellular symbiotic partners of sea anemones and also indulged her love of diving.

In 2000, she came to the University of Illinois Urbana Champaign, where she joined May Berenbaum’s lab to study chemical ecology and started working on the swallowtail caterpillar project with Dr. Berenbaum’s collaborator Dr. Mary Schuler, in cell and developmental biology. For this work, she studied the regulation of cytochrome P450 genes in *Papilio polyxenes* and *Papilio glaucus*. For this work, she was funded by the University of Illinois Urbana-Champaign Toxicology Scholarship and the NIH Cell and Molecular Biology Training Grant. She won the ESA President’s Prize in Section B for her poster on this work (2003).

For her PhD, she continued studying the regulation of cytochrome P450 genes with Drs. May Berenbaum and Mary Schuler, but switched to *Drosophila melanogaster*, to take advantage of the genetic and genomic tools available. In fall of 2008, she began collaborating with Dr. Barry Pittendrigh on one particular P450 across multiple species of *Drosophila*. She was funded
by an EPA STAR fellowship and the Dissertation Completion fellowship. During this time, she received the Midwest Regional Chapter of the Society of Toxicology Young Investigator’s Award (2005) and later won 2nd prize for her presentation of part of this work at ESA (2009).

At the end of her time as a doctoral student, she began working to educate the public about pollinator decline and conservation. Along with her colleagues, she has created and organized events for National Pollinator Week, which she hopes will continue after she leaves. She soon leaves for Avignon, France, where she will work as a postdoctoral researcher in the lab of Dr. Yves Le Conte, on a project addressing honey bee health. As she stumbles through their language, she hopes that the French will take pity on her.