Graduate Programme

“Molecular Biology & Biomedicine”

Master thesis

“Elucidation of the role of CYP4G genes in insect cuticular hydrocarbon biosynthesis”

Parasyris Konstantinos

(R.N. 403)

Committee	**Supervisors**
Prof. John Vontas	Prof. John Vontas
Prof. George Chalepakis	Dr. Vasileia Balabanidou
Dr. Inga Siden-Kiamos	PhD candidate Mary Kefi

Heraklion

October 2020

FORTH

INSTITUTE OF MOLECULAR BIOLOGY & BIOTECHNOLOGY

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Table of Contents

Abstract

Hydrocarbons (HCs) act as cuticular waterproofing agents in insects. A pathway of many enzymes facilitates their production: synthases, elongases, desaturases, reductases and CYP4Gs. CYP4Gs catalyze the last step of HC biosynthesis, the decarbonylation reaction that converts the aldehyde-substrate to the hydrocarbon-product. CYP4Gs are members of the cytochrome P450 proteins (P450s) and their genes not only have orthologues distributed across the Insecta, but also there is no CYP4G-like sequence in other organisms; they possess a unique +44 residue insertion. Two CYP4G enzymes are found in *Drosophila melanogaster*, CYP4G1 and CYP4G15. Despite the fact that CYP4G1 has been studied extensively (catalyzes the insect-specific oxidative decarbonylation step for cuticular HCs production and has significant role in desiccation resistance), no information about the functional contribution of this insertion has been suggested. That is why two transgenic flies were created, with a UAS-REGal4 system and the CRISPR-Cas9 genome editing technique, to express deleted forms of the insertion of CYP4G1 under a null-CYP4G1wt background. In addition, little details about CYP4G15 are available; its localization is different than CYP4G1 (brain vs oenocytes) and its function is unknown. Again, a UAS-REGal4 system was developed for the purpose of creating flies with simultaneous oenocyte-specific knock down of *Cyp4g1* and CYP4G15 expression. Lastly, the malaria vector *Anopheles gambiae*, like *Drosophila*, has two CYP4G enzymes; CYP4G16 and CYP4G17. Both enzymes are localized in the oenocytes, having though distinct sub-cellular localizations at the adult stage, and both act as oxidative decarbonylases. However, an in depth biochemical characterization of CYP4G enzymes (substrate specificities, enzyme kinetics, catalysis) has never been conducted so far. For this reason, expression of recombinant CYP4Gs was tested under various parameters (culture conditions, expression vectors, gene sequence etc.) in the heterologous system of *Escherichia coli*.

Περίληψη

Οι υδρογονάνθρακες δρουν ως στεγανοποιητικοί παράγοντες στα έντομα. Ένα μονοπάτι πολλών ενζύμων είναι υπεύθυνο για την παραγωγή τους: συνθάσες (synthases), ελονγκάσες (elongases), δεσατουράσες (desaturases), αναγωγάσες (reductases) και CYP4Gs (κυτοχρώματα P450 της οικογένειας 4 και υπο-οικογένειας G). Τα CYP4Gs καταλύουν το τελευταίο βήμα της βιοσύνθεσης υδρογονανθράκων, και συγκεκριμένα την αντίδραση αποκαρβονυλίωσης που μετατρέπει το υπόστρωμα-αλδεΰδη σε προϊόν-υδρογονάνθρακα. Τα CYP4Gs είναι μέλη των πρωτεϊνών του κυτοχρώματος P450 και τα γονίδια τους όχι μόνο έχουν ορθόλογα σε όλα τα έντομα, αλλά επίσης δεν υπάρχει αλληλουχία τύπου CYP4G σε άλλους οργανισμούς. Μάλιστα, διαθέτουν μια μοναδική «προσθήκη» 44 αμινοξέων (residue insertion). Δύο ένζυμα CYP4G συναντώνται στη *Drosophila melanogaster*, το CYP4G1 και το CYP4G15. Παρά το γεγονός ότι το CYP4G1 έχει μελετηθεί εκτεταμένα (καταλύει την εντομοειδική οξειδωτική αποκαρβονυλίωση για την παραγωγή επιδερμικών υδρογονανθράκων και έχει σημαντικό ρόλο στην αντοχή στην αποξήρανση), δεν υπάρχουν στοιχεία σχετικά με τη λειτουργική συμβολή αυτής της αμινοξικής αλληλουχίας. Αυτός είναι ο λόγος και για τον οποίο δημιουργήθηκαν δύο διαγονιδιακές μύγες, μία με ένα σύστημα UAS-REGal4 και μία μέσω της τεχνικής γενετικής τροποποίησης CRISPR-Cas9, με σκοπό την έκφραση μορφών της CYP4G1 όπου θα φέρουν την έλλειψη για αυτή τη μοναδική αμινοξική αλληλουχία (απουσία της ενδογενούς πρωτεΐνης). Επιπλέον, υπάρχουν λίγες μόνο πληροφορίες για τη CYP4G15. Εντοπίζεται σε διαφορετικό σημείο σε σχέση με τη CYP4G1 (εγκέφαλος έναντι οινοκύτταρα) και η λειτουργία της είναι μέχρι στιγμής άγνωστη. Αντιστοίχως, ένα σύστημα UAS-REGal4 αναπτύχθηκε με σκοπό τη δημιουργία μυγών με ταυτόχρονη έκφραση του CYP4G15 και αποσιώπηση του *Cyp4g1*, ειδικά στα οινοκύτταρα. Τέλος, ο φορέας της ελονοσίας *Anopheles gambiae*, όπως και η *Drosophila*, έχει δύο ένζυμα CYP4G, τα CYP4G16 και CYP4G17. Και τα δύο ένζυμα συναντώνται στα οινοκύτταρα, με διαφορετικούς όμως υποκυτταρικούς εντοπισμούς στα ενήλικα άτομα. Επίσης, και τα δύο δρουν ως οξειδωτικές αποκαρβονυλάσες. Ωστόσο, μέχρι σήμερα, δεν έχει πραγματοποιηθεί σε βάθος ένας χαρακτηρισμός των CYP4Gs (ειδικότητα υποστρώματος, κινητική ενζύμου, κατάλυση). Για το λόγο αυτό, η επαγωγή της έκφρασης των CYP4Gs του κουνουπιού (*Anopheles gambiae*) δοκιμάστηκε υπό διάφορες παραμέτρους (συνθήκες καλλιέργειας, φορείς έκφρασης, αλληλουχία γονιδίων κ.λπ.) στο ετερόλογο σύστημα του *Escherichia coli*.

Acknowledgements

First of all, I would like to thank professor John Vontas for giving me the opportunity to be a member of his lab and work on these exciting projects and specially for offering his scientific knowledge and guidance. In addition, I would like to express my gratitude to Dr. Vasileia Balabanidou who guided me during the whole year and provided me with helpful advice about the experiments. Moreover, I am most grateful to PhD candidate Mary Kefi for the friendly collaboration, her willingness to help me and guide me whenever it was necessary, her assistance and her participation in the experimental procedures; without her implication, the outcome of this post-graduate thesis would not be the same in many ways. Next, I owe my thanks to Dr. Vasilis Douris for his valuable guidelines around the CRISPR technique and to Yannis Livadaras (IMBB/FORTH) for performing all the injections to the *Drosophila* embryos. Furthermore, I really appreciate Aris Ilias, Sofia Kaforou, George Samantsidis, Sofia Balaska and generally all the members of Molecular Entomology / Pesticide Science Lab (Vontas Group) for their kindness, hospitality and assistance from the moment I started my work in the lab. Finally, and most importantly, I am grateful to my family and friends for always supporting me.

1. Introduction

1.1 Cytochrome P450 (CYP450) proteins

The cytochrome P450 proteins (CYP450s) are encoded by CYP genes and constitute one of the largest gene families in virtually all living organisms (bacteria, protists, plants, fungi, and animals) because of the high diversity resulting from successive gene duplications succeeded by sequence divergence[1–3]. P450 enzymes are heme-thiolate proteins, comprising a Fe^{II}-CO complex that has a ~450 nm absorbance peak[1–4]. The molecular weight usually varies around 55±10 kDa[1,4]. They show activity as monooxygenases, oxidases, desaturases, reductases, isomerases, etc. and have been associated with the catalysis of at least 60 distinct chemical reactions[1–6]. On one hand, many of them are associated with the metabolism (synthesis and degradation) of endogenous substrates (signaling molecules[1,2,7], hormones[1,8], lipids[1] etc.)[1,3]. On the other hand, much of their notoriety has been involved in the metabolism or detoxification of xenobiotics (natural products, pesticides, drugs, insecticides etc.)[1,3].

Furthermore, the enzymes nomenclature is the following: P450 root is indicated by the CYP prefix, the P450 family is designated by an Arabic number, the subfamily is denoted by a capital letter and each gene (italics) or protein is designated by an Arabic number[1] (Figure 1). In insects, four clans of P450s can be found: CYP2 Clan, Mitochondrial CYP Clan, CYP3 Clan and CYP4 Clan[1].

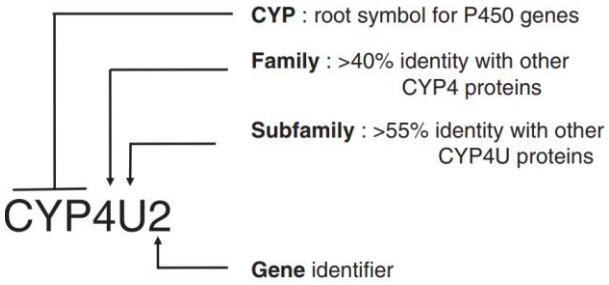


Figure 1: Scheme of the P450 nomenclature. Adapted from Feyereisen 2012[1].

1.2 Insect CYP4Gs

1.2.1 CYP4Gs in general

This work will focus on the CYP4G enzymes, because they have been investigated quite less opposed to the rest insect P450 enzymes[1] and their genes not only have orthologues distributed across the Insecta, but also there is no CYP4G-like sequence in other organisms[9]. It is believed that the first *CYP4G* gene may was the result of retroposition of the *CYP4AC1* gene and the subsequent duplication of this ancestor into the *CYP4G1* and *CYP4G15* genes, around 400 million years ago[3,10]. The majority of insect orders have at least one *CYP4G* gene (e.g. honey bees and pea aphids), with an average of two (flies, mosquitos etc.) and some have a number of *CYP4G* genes (e.g. Lepidoptera)[3]. In Table 1

some examples of *CYP4G* genes are shown. In addition, a notable finding came from the first global P450 transcriptome analysis of *Drosophila melanogaster* which showed that CYP4G1 expression is the highest among all P450 genes, indicating the greatly expression level of this subfamily genes[3,11].

Table 1: CYP4G genes found across different insect species.

Insect	Gene(s)
Drosophila melanogaster	***CYP4G1*** & ***CYP4G15***
Locusta migratoria	***CYP4G62*** & ***CYP4G102***
Acyrthosiphon pisum	***CYP4G51***
Blattella germanica	***CYP4G19***
Tenebrio molitor	***CYP4G122*** & ***CYP4G123***
Nilaparvata lugens	***CYP4G76*** & ***CYP4G115***
Rhodnius prolixus	***CYP4G106*** & ***CYP4G107***
Musca domestica	***CYP4G2***
Apis mellifera	***CYP4G11***
Dendroctonus ponderosae	***CYP4G55*** & ***CYP4G56***
Anopheles gambiae	***CYP4G16*** & ***CYP4G17***
Helicoverpa armigera	***CYP4G8***
Mamestra brassicae	***CYP4G20***

1.2.2 CYP4Gs structure and the CYP4G-specific residue insertion

CYP4Gs, as P450s, are characterized by five conserved motifs: The WxxxR motif, the GxE/ DTT/S motif, the ExLR motif, the PxxFxPE/DRF motif, and the PFxxGxRxCxG/A motif (Figure 2A)[1]. All motifs are important for the stability of the overall enzyme structure and possibly some protein interactions, but the last one also bears the cysteine (thiolate) ligand to the heme iron that is accountable for the typical 450 nm absorbance peak of the Fe^{II}–CO complex of P450s[1]. This loop is the most conserved part of the protein, making it the signature characteristic of P450 enzymes[1]. It is important to mention here the N-terminal sequence of P450 proteins, because it is responsible for the anchoring through the membrane (about 20 hydrophobic amino acids) of membrane-bound P450s and generally their correct folding for the formation of a functional enzyme[1]. Notably, all CYP4Gs are distinguished from the rest P450s due to the existence of a +44 residue insertion between the G and H helices (Figure 2B)[3]. The insertion protrudes from the globular structure of the enzyme, on the cytoplasmic side distal from the membrane surface in which the N-terminal is anchored and the loop between helices F and G is dipping[3]. Furthermore, the CYP4G insertion is quite enriched in acidic amino acids[3,12], rousing two hypotheses; it is responsible for ionic interactions which might either enhance the P450-CPR complex stabilization or assist the formation of a metabolon with CYP4G, CPR and yet unidentified proteins. A potential candidate could be the enzymes providing CYP4Gs with their substrates, i.e. fatty acyl-CoA reductases (FAR)[3]. Nevertheless, it is shown by computational work that there are no interference issues between the insertion and other necessary protein interactions, such as CPR[1,3].

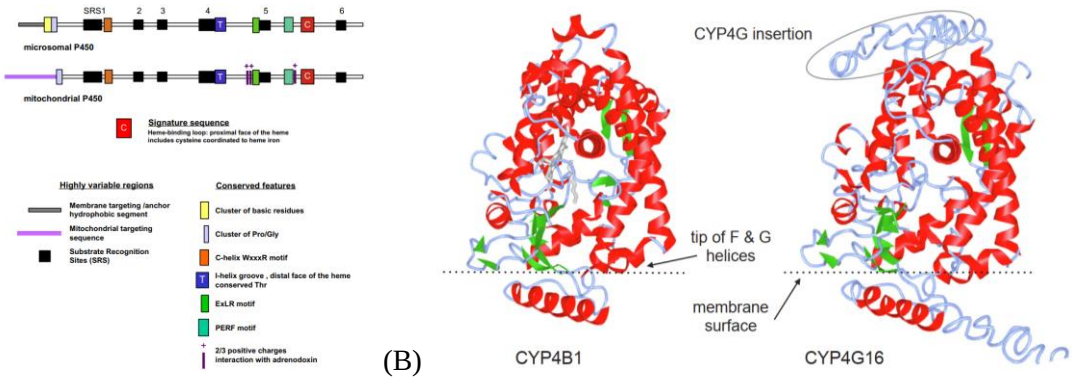


Figure 2: (A) Conserved and variable regions of P450 proteins illustrated over their primary structure (sequence). Adapted from Feyereisen, 2012[1]. (B) Comparison of the model of *Anopheles* CYP4G16 and the structure of rabbit CYP4B1. Helices are in red and sheets in green. The view is through the I helix, with the N-terminal transmembrane helix and the tips of the F and G helices at the bottom. The CYP4G insertion is visible on top of the model, after the G helix. The approximate position of the membrane surface is shown as stippled line. Adapted from Feyereisen, 2020[3].

1.2.3 CYP4Gs function

The first information about the function of CYP4G enzymes were brought to light not so many years ago, showing roles in alkanes and alkenes synthesis; those hydrocarbons were thought to have dual functions as cuticular waterproofing agents and pheromonal signals[13,14]. This work focuses on the cuticular part and the contribution of hydrocarbons has been validated by many following researches. The fundamental study by Qiu *et al.* 2012[9] proved that knock-down (KD) of *Cyp4g1* in *Drosophila* leads to increased mortality rates and defective cuticular hydrocarbons (CHC) composition, which in turn provokes increased sensitivity to desiccation stress[3,15,16]. Similar results were obtained by RNA silencing of the corresponding genes[3] in *Locusta migratoria*[17], *Acyrthosiphon pisum*[18], *Blattella germanica*[19], *Tenebrio molitor*[20], *Nilaparvata lugens*[21] and *Rhodnius prolixus*[22]. In addition, this study indicated that the recombinant *Musca domestica* CYP4G2-P450 reductase fusion protein had oxidative decarbonylation activity, with long chain fatty aldehydes as substrates, thus catalyzing the last step of CHC biosynthesis (see below)[3,9]. Again, this function was proved by a number of groups in other insects. CYP4Gs in honey bees[23], in mountain pine beetles[24], in mosquitos[25,26] and in triatomines[22] were functional oxidative decarbonylases. More specifically, the latest *Anopheles gambiae* studies showed that CYP4G16 could act as a decarbonylase *in vitro* (in *Sf9* cells)[25] and that both mosquito *Cyp4g* genes, alone or in combination, could rescue the lethal phenotype of *Cyp4g1* KD in transgenic flies[26]. Lastly, the expression of the *Cyp4g* genes takes place mainly into the oenocytes[9,17,21,25]. In other words, CYP4G enzymes share a common biochemical function in hydrocarbon biosynthesis, operating as oxidative decarbonylases[3].

1.2.4 CYP4G1 of *Drosophila melanogaster*

The *Cyp4g1* gene is one of the not too many intron-less *Drosophila* P450 genes[3,9,27], is the most highly expressed P450 gene in *Drosophila melanogaster*[11] and its expression is particularly restricted to the oenocytes[3,9,27]. Moreover, it was proved that CYP4G1 acts as an oxidative decarbonylase for cuticular hydrocarbon biosynthesis both *in vitro* (*Sf9* insect cells) and *in vivo*; the conversion of aldehydes to CHCs was feasible *in vitro* and the RNAi-mediated KD of *Cyp4g1* resulted to outcomes implicating CHC synthesis[9]. More specifically, the silenced flies showed high mortality at the time of adult emergence, significant differences in the CHC content (both qualitative and quantitative), which lead to the undoubted correlation with much lower desiccation resistance (Figure 3A)[3,9]. In addition, CYP4G1 mutants having both alleles as protein-null proteins due to the removal either of the transcription start site (*Cyp4g1*$^{\Delta 14\text{-}9}$) or all of the *Cyp4g1* coding sequence (*Cyp4g1*$^{\Delta 4}$) failed to emerge as adults; animals homozygous for both deletions, despite developing normally through larval and early pupal stages, they arrest during mid-to-late pupal stages and the majority of them die during adult eclosion (Figure 3B,C)[28].

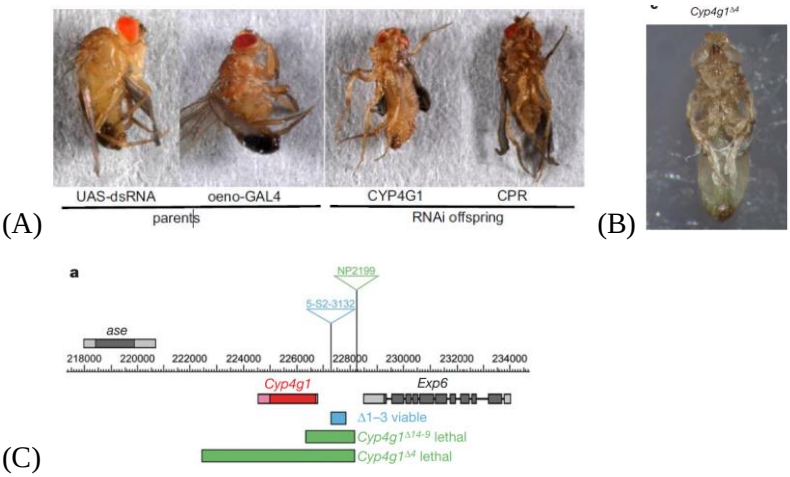


Figure 3: Cyp4g1 is an essential oenocyte-specific gene regulating CHC biosynthesis. (A) Phenotype of severe CYP4G1 and CPR suppression by RNAi at adult emergence. Oenocyte-GAL4 and UAS-dsRNA parents (left) have a normal phenotype but the F1 offspring of their crosses (right) are characterized by high mortality at emergence. Adapted from Qiu *et al.* 2012[9]. (B) *Cyp4g1*$^{\Delta 4}$ hemizygote arrested during adult eclosion, shown dissected from the puparial case. (C) Cyp4g1 locus depicting P-element insertions and viable (blue) and lethal (green) deletions. Predicted translated (dark fill) and untranslated (light fill) regions of *Cyp4g1*, asense (ase) and *Exportin 6* (Exp6). Adapted from Gutierrez *et al.* 2007 [28].

1.2.5 CYP4G15 of *Drosophila melanogaster*

In parallel, much less is known about *Drosophila*'s CYP4G15; there is only one publication at the beginning of the century[29]. *Cyp4g15* gene is located on chromosome X, like *Cyp4g1*, at position 10C1 (*Cyp4g1* at 1B). On the contrary of *Cyp4g1*, which is intron-less, *Cyp4g15* is composed of five introns (590, 430, 2330, 80 and 90 bp in succession and length). It is

expressed from the third instar larval stage highly in the cortex cells of the brain, i.e. the central nervous system, and possibly at very low levels in the digestive system and not at the oenocytes (Figure 4). There is no information about its function yet; only that *Cyp4g15*-RNA silenced flies are viable[3,16].

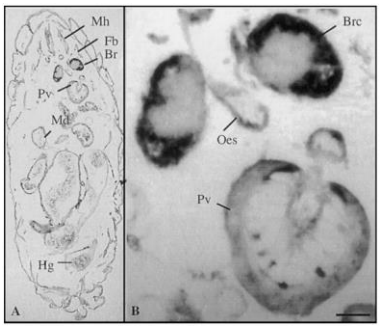


Figure 4: (A) *In situ* hybridization on frontal section of wandering third-instar larvae. (B) Detail of the anterior region (B). Scale bar stands for 100 mm. Br, brain; Brc, brain cortex; Fb, fat body; Hg, hindgut; Mg, midgut; Mh, mouth hook; Pv, proventriculus; Oes, oesophagus. Adapted from Maıbeche-Coisne *et al.* 2000 [35].

1.2.6 CYP4G16 and CYP4G17 of *Anopheles gambiae*

As discussed above, *Anopheles gambiae*, like *Drosophila*, has two *Cyp4g* genes; *Cyp4g16* and *Cyp4g17*. *Cyp4g16* is the *Cyp4g15* homolog and *Cyp4g17* is the *Cyp4g1* homolog[3]. Both enzymes are localized at the oenocytes, but have distinct sub-cellular localizations in the adult oenocytes; CYP4G16 is associated with the intracellular side of the plasma membrane, whereas CYP4G17 is dispersed throughout the cytoplasm (Figure 5)[25]. Moreover, both proteins act as decarbonylases (CYP4G17 *in vivo*, CYP4G16 both *in vivo* and *in vitro* in *Sf9* insect cells)[25,26]. In addition, the conditional expression of CYP4G16 and CYP4G17 in oenocyte-specific *Cyp4g1* KD *Drosophilae* produced the same blend of hydrocarbons but with qualitative differences with three very long CHCs being overexpressed when CYP4G17 was present[26]. In other words, it is suggested that CYP4G17 is showing preference for longer chain CHCs and CYP4G16 for shorter CHCs[26]. Similar preference is also demonstrated in *Rhodnius prolixus* (triatomine) where molecular docking analysis predicted better fit of straight chain aldehydes with CYP4G106 and methyl-branched aldehydes with CYP4G107[22]. Furthermore, it is possible that longer CHC chains increase the melting temperature of the insect cuticle and likely affect desiccation resistance, and methyl branching augments the chemical information of the cuticle[26].

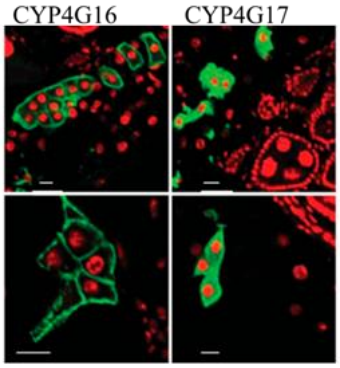


Figure 5: Merged immune-histochemical images focusing on oenocytes, showing the subcellular localization of CYP4G17 in the cytoplasm (presumably bound to ER) and CYP4G16 associated with PM. (Scale bars, 10 μm.). Adapted from Balabanidou *et al.* 2016 [20].

1.3 CYP4Gs and cuticular hydrocarbons

1.3.1 Insect cuticle

Generally, chitin, proteins and lipids, such as CHCs, are the main components of the insect cuticle and are arranged into two discrete horizontal layers based on their physiological and biochemical constitution; the outer chitin-free epicuticle and the inner chitinous procuticle[30]. As far as their ultrastructural contexture is concerned, the epicuticle consists an inner and an outer sublayer (the cuticulin layer). The procuticle comprises an upper exocuticle and a lower endocuticle (Figure 6). On one hand, the non-chitinous region, the epicuticle, incorporates different quinones, lipids (CHCs) and proteins and is characterized by a hydrophobic nature, making it the first barrier against dehydration and bulging[30–32]. On the other hand, the chitinous procuticle, as its name implies, is formed by bundles (microfibrils) of the polysaccharide chitin that attribute to the elasticity and rigidity of the cuticle[30].

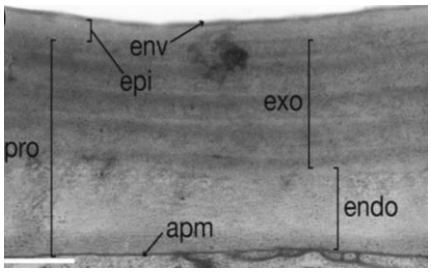


Figure 6: The three layers of the *D. melanogaster* cuticle: The procuticle (pro) is the inner chitin-protein matrix that is attached to the apical surface (apical plasma membrane, apm) of the epithelial cells and is subdivided into an upper and a lower zone called the exo- (exo) and the endo-cuticle (endo), respectively. The exocuticle is electron-denser than the endocuticle due to a higher degree of sclerotization. The epicuticle is a uniform layer. The envelope (env), a wax and cement layer to the surface of the cuticle, often collapses due to the fixation procedure. Adapted from Moussian 2010 [22].

1.3.2 Insect cuticular hydrocarbons (CHCs) and oenocytes

The insect cuticular hydrocarbons (CHCs) were firstly identified more than 80 years ago, as a principal component of the waxy layer discussed above[27,31,32]. It is also shown that their eradication by organic solvents provokes a rise in desiccation rate[25,26] and generally CHCs have a dual role, as suggested above, in attributing to the hydrophobic nature of the cuticle (maintain the water balance and prevent desiccation) and serving as signaling molecules in chemical communication and mate recognition[27]. Insect CHCs are a complex blend of long-chain, straight-chain or methyl-branched saturated or unsaturated molecules (alkanes and alkenes), ranging around from 21 to 50 carbons in chain length, and are synthesized in specialized cells named oenocytes[23,25].

Much is known now about the biosynthesis pathway of CHCs in the oenocytes[13,27,28], which vary among insect species in size, number and anatomical locations[34,35] (Figure 7). Oenocytes could be characterized as specialized hepatocyte-like cells that regulate lipid metabolism in general[28], since they are implicated with various physiological functions such as regulation of respiration, detoxification, tissue histolysis, dietary related longevity, hormone production and cuticle synthesis[36,37]. Their implication in lipid metabolism is pointed by many indications; the existence of waxes or other lipids interior of the oenocytes (shown by organic extractions and histological stains)[36], the

abundant smooth endoplasmic reticulum (roles in synthesis, processing, and/or secretion of lipids)[36,38] and the expression of a large number of lipid-synthesizing and -catabolizing enzymes and other proteins like lipophorin receptors(*1) [28,36,39–44].

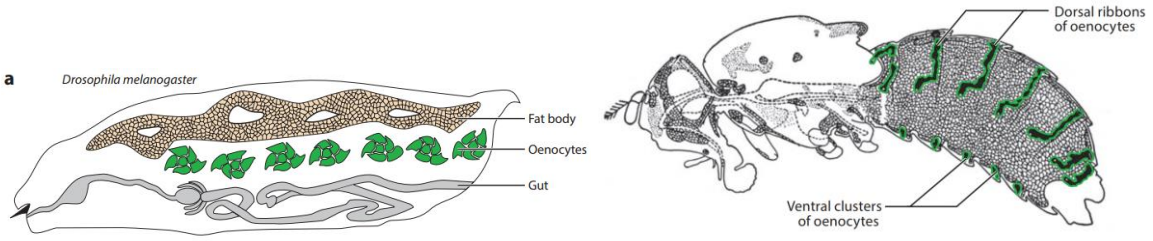


Figure 7: (A) A larva drawing showing the oenocytes, the fat body, and the gut localization. (B) An adult fly drawing showing the abdominal dorsal ribbons and ventral clusters of oenocytes (green). Adapted from Makki *et al.* 2014 [36].

As for these enzymes are concerned, the principal biosynthetic CHC pathway is conserved (Figure 8A) , in spite of the diversity of CHC mixtures among insects[13]. The CHC biosynthesis initiates from acetyl-CoA, which is elongated so as to form a long-chain fatty acyl-CoA[27,45]. This elongation reaction is catalyzed by two different forms of fatty acid synthases (FAS); a cytosolic and a microsomal FAS, which are involved in the elongation of non-branched fatty acyl-CoA (precursors of linear alkanes and alkenes) and of branched fatty acyl-CoA (precursors of methyl-branched CHCs), respectively[27,45]. Elongases then elongate these fatty acyl-CoAs to specific lengths, and desaturases add double bonds to the fatty acyl-CoA chain during this chain elongation process[27,45]. Then, the fatty acyl-CoAs are reduced to aldehydes by fatty acid reductases[9,27,45]. Lastly, the aldehydes serve as substrates for the last oxidative decarbonylation step (CO_2 is released) for the production of CHCs, which is performed by CYP4Gs[9,27,36,46].

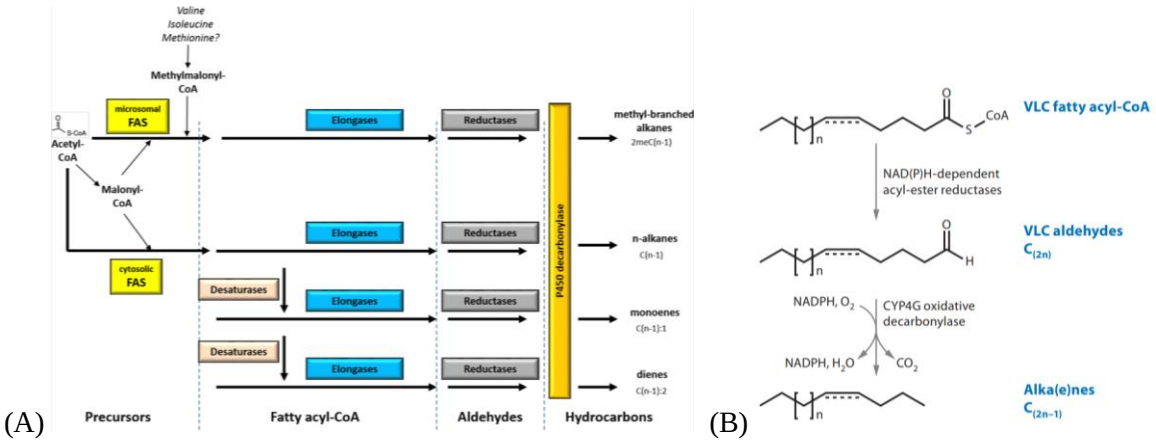


Figure 8: (A) CHCs are synthesized in the oenocytes from acetyl-CoA, which undergoes an elongation reaction to form a long-chained fatty acyl-CoA. This initial elongation reaction is catalyzed by either a microsomal or a cytosolic fatty acid synthase. Elongases then further lengthen these fatty acyl-CoAs to specific lengths, and desaturases add double bonds during the chain elongation process[13]. The fatty acyl-CoAs are then reduced to aldehydes by fatty acid reductases, before a decarbonylation reaction mediated by

a cytochrome P450, converts the aldehydes to hydrocarbons[9]. Adapted from Chung *et al.* 2015 [27]. (B) The last two steps of CHC biosynthesis. Adapted from Makki *et al.* 2014 [36].

1.3.3 P450 indispensable partners

As shown in (Figure 8B), CYP4G oxidative decarbonylases require NADPH as reducing equivalents[36,46]. These are provided by the CYP4Gs' interactors or redox partners, the NADPH cytochrome P450 reductase (CPR)[1,28,36,37,46,47]. CPRs are flavoproteins that employ both FAD and FMN as cofactors in order to receive the two electrons from the NADPH and then provide them to P450 enzymes[1,48]. In other words, CPRs are dual-substrate (NADPH and the electron acceptor, P450) and dual-product enzymes (NADP+ and the reduced electron acceptor, FAD and FMN)[1]. CPRs are also expressed into the oenocytes[47].

1.3.4 Cuticular thickening and insecticide resistance

Many publications investigate the possible implication of CYP4Gs in insecticide resistance phenotype. There are three main types of cuticular resistance phenotypes: thickening of the epicuticle (enriched lipids and CHCs), thickening of the procuticle (enriched cuticular proteins and chitin filaments) and altered cuticle composition (differences in cuticular proteins and chitin composition)[49]. RNAi of the endogenous *Cyp4gs* enhances insecticide penetration[3,19,50,51] and the finding of constitutive overexpressed *Cyp4g* genes[3,52–56] in insecticide-resistant strains point out the contribution of CHCs to insecticide resistance[3,21,25,57]. Lastly, a thicker epicuticular layer and a significant elevation in CHC levels was revealed by cuticular analysis using electron microscopy of resistant versus susceptible mosquitoes[57].

1.4 Tools to study CYP4Gs

1.4.1 Recombinant production of CYP4Gs in microbial cell factories

In order to investigate the function of P450s (substrate specificities, enzyme kinetics, catalysis), the recombinant production of these enzymes is explored in different expression systems, as well as in the heterologous system of *E. coli*[58]. *E. coli*, as an expression host, possesses many advantages; high yield, rapid growth, low cost cultivation parameters and easy genetic manipulation are some of them[1,58]. However, there are many challenges regarding the interaction with partners (CPR is necessary to catalyze the decarbonylase reaction), incorporation of the heme group, codon usage (different codon bias between bacteria and insects), post-translational modifications (absent in bacteria) and the fact that most eukaryotic P450s are membrane bound[58]. The latter one is also the most challenging because bacteria lack the internal membrane compartmentalization in which eukaryotic membrane proteins are anchored and hence there are issues with proper membrane insertion, protein folding and possible formation of toxic protein aggregates (inclusion bodies)[58]. These major experimental bottlenecks are circumvent with N-terminal

modifications; N-terminal deletion, N-terminal substitutions (MALLLAVF(L) sequence is the commonest) or fusion with bacterial signal peptides[58]. In addition, the P450-partner interactions are secured via the expression of chimeric fused holoenzymes (P450 C-terminus is fused via a linker with the N-terminus of CPR), expression of both proteins from a bicistronic plasmid, expression of both proteins from one plasmid with two promoters or expression from two different plasmids[58]. Moreover, co-expression of chaperones could assist the correct folding of the protein, which in turn would increase its activity[58]. Lastly, the optimization of culture conditions (temperature, medium etc.), the integration of the heme group (δ-ala-leuvenic acid, a heme-precursor, is usually supplied to the media), the fine-tuning of P450 expression (different IPTG[(*2)] levels can cause various outcomes) and the selection of the proper bacterial strain (see Material and Methods) are very important[58]. Depending each time of the gene of interest, some or all of these parameters have to be adjusted. Finally, there are also other heterologous systems for the recombinant production of insect P450s *in vitro* (yeast[58] and insect cell lines[25,59,60]).

1.4.2 *Drosophila melanogaster*

Drosophila melanogaster, the common fruit-fly, has been used for several years as model organism, because it is easily reared in the lab and they produce a large number of progeny in a short period of time[61]. More importantly, the whole genome is available since 2000, making it accessible for the uncharacterized genes and proteins and easy for genetic manipulations (forward and reverse genetics) in the interest of unraveling the mode of action of any gene of interest[61].

1.4.3 CRISPR-Cas9

One of the commonest genome editing techniques nowadays is CRISPR (Clustered, Regularly Interspaced Short Palindromic Repeats)-Cas9 technology. This system was originally discovered in bacteria representing an adaptive immune system against viruses. Cas9 is an RNA–guided protein with DNA cutting capability[62,63]. More specifically, a single-guide RNA sequence (sgRNA) of ~20 nt length leads the Cas9 to the target sequence in a sequence specific manner, generating a double stranded break (DSBs)[64,65], which in turn will provoke the recruitment of either of two DNA repair mechanisms, the Non-Homologous End Joining pathway (NHEJ) and the Homologous-Directed Repair (HDR) mechanism[66]. In order to recruit the HDR mechanism, a homologous template should be available and in the case of *Drosophila*, a donor plasmid embodying two homology arms, ~1 Kb each on both sides of the desired region is more than suitable[67,68]. Lastly, two plasmids containing the sgRNAs, targeting before and after the cleavage site, offer the position specificity[67,68].

1.5 Aims of this study

This study aims to:

- Understand the role of CYP4G1 insertion by creating transgenic flies that express a deleted form of the insertion of the *Cyp4g1* gene,
- Elucidate the function of CYP4G15 by creating transgenic flies that ectopically express CYP4G15 in the oenocytes and
- Successfully express recombinant *Anopheles gambiae* CYP4Gs in bacteria by overcoming the commonly experienced limitations of standard expression systems.

2. Materials and methods

2.1 Insect strains

2.1.1 Fly strains

All *Drosophila* strains described in Table 2 and discussed in the next sections were reared on a standard fly diet at 25ºC temperature, 60-70% humidity and 12:12 hour photoperiod.

Table 2: All the *Drosophila* strains utilized in this work.

Strain Name	Characteristics	Genotype
Nanos	Nanos-Cas9 system	$\frac{+}{+ \text{ or } \neg}; \frac{+}{+}; \frac{\text{Cas9}}{+}$
VK13	Based on Φc31 integrase vector system; the attP site is located in 3rd chr	$\frac{+}{+ \text{ or } \neg}; \frac{+}{+}; \frac{\Phi\text{−integrace, w+}}{+}$
yw	Different eye and body color background	$\frac{\text{yw}}{\text{yw}}; \frac{+}{+}; \frac{+}{+}$ or $\frac{\text{yw}}{\neg}; \frac{+}{+}; \frac{+}{+}$
If/CyO	Double balancer for 2nd chr	$\frac{\text{w}}{\text{w}}; \frac{\text{If}}{\text{CyO / LacZ}}; \frac{\text{TM3}}{\text{TM6B}}$
TM3/TM6B	Double balancer for 3rd chr	$\frac{\text{yw}}{\text{yw}}; \frac{+}{+}; \frac{\text{TM3}}{\text{TM6B}}$
R46/R53	Co-inherited 2nd and 3rd chr	$\frac{\text{yw}}{\text{yw}}; \frac{\text{R46}}{\text{T(2,3)CyO}} \frac{\text{R53}}{\text{TM6B}}$
REGAL4	RE-Gal4 driver (2nd chr)	$\frac{\text{yw}}{\neg}; \frac{\text{REGAL4}}{\text{REGAL4}}; \frac{+}{+}$
Cyp4g1 KD	RNAi mediated knock-down of CYP4G1 (2nd chr)	$\frac{\text{yw}}{\neg}; \frac{\text{UAS −dsCyp4g1, w3, y+}}{+}; \frac{+}{+}$
FM7	Balancer for X chr	$\frac{FM7c}{\neg}; \frac{+}{+}; \frac{+}{+}$
Δ4	Cyp4g1$^{\Delta 4}$ lethal	$\frac{Cyp4g1\Delta 4}{FM7c}; \frac{+}{+}; \frac{+}{+}$

2.1.2 Mosquito strains

The *An. gambiae* strain N'Gousso (Cameroon) was reared under standard insectary conditions at 27 °C temperature, 70–80% humidity, with a 12-h:12-h photoperiod.

2.2 Protocols

2.2.1 DNA extraction

DNA from *D. melanogaster* was extracted with DNAzol (MRC), following the instructions of the manufacturer.

2.2.2 PCR

2.2.2.1 *Cyp4g15*

GoTaq® DNA Polymerase (Promega) was used for the amplification of a 1740 bp fragment containing *Cyp4g15* ORF using primer pair BssHII 4g15 F/XhoI 4g15 R (Table S1) that introduce a 5′ BssHII site and a 3′ XhoI site. The template for the amplification of *Cyp4g15* was cDNA of adult Drosophila RNAs. PCR conditions were 95°C for 2 min, followed by 39 cycles of 95°C for 30 s, 62°C for 30 s, 72°C for 2 min and a final extension step of 72°C for 10 min.

2.2.2.2 *Cyp4g1Δinsertion*

GoTaq® DNA Polymerase (Promega) was used for the amplification of two fragments, 875 bp and 1327 bp respectively. On one hand, the 875 bp fragment contains the anterior *Cyp4g1* ORF; the primer pair of Start-F/Mid-R (Table S1), that introduce a 5′ BssHII site and a 3′ XhoI site, were used. On the other hand, the 1327 bp fragment contains the posterior *Cyp4g1* ORF; the primer pair of Mid-F/End-R (Table S1), that introduce a 5′ BssHII site and a 3′ XhoI site, were used. The template for the amplification of *Cyp4g1Δ* was cDNA of adult Drosophila RNAs. PCR conditions were 95°C for 2 min, followed by 39 cycles of 95°C for 30 s, 58.5°C for 30 s, 72°C for 2 min and a final extension step of 72°C for 10 min.

2.2.2.3 *Cyp4g16* and *Cyp4g17*

Kapa Long Range Taq DNA Polymerase (Kapa Biosystems) was used for the amplification of three different *Cyp4g16* and *Cyp4g17* ORFs. Primer pair 4G16 F/4G16 R (Table S1) was used to amplify a 1707 bp fragment containing the *Cyp4g16* ORF and introducing a 5′ NdeI and a 3′ SalI. Primer pair 4G17 F/4G17 R (Table S1) was used to amplify a 1702 bp fragment containing the *Cyp4g17* ORF and introducing a 5′ NdeI and a 3′ SalI. Primer pair ΔN4G16 F/4G16 R (Table S1) was used to amplify a 1596 bp fragment containing the N-terminally ablated *Cyp4g16* ORF and introducing a 5′ NdeI and a 3′ SalI. Primer pair ΔN4G17 F/4G17 R (Table S1) was used to amplify a 1582 bp fragment containing the N-terminally ablated *Cyp4g17* ORF and introducing a 5′ NdeI and a 3′ SalI. Primer pair ΔN4G16 F/ his4G16 R (Table S1) was used to amplify a 1614 bp fragment containing the N-terminally ablated *Cyp4g16* ORF with the addition of a C-terminal poly-His tail, and introducing a 5′ NdeI and a 3′ SalI. Primer pair ΔN4G17 F/ his4G17 R (Table S1) was used to amplify a 1596 bp fragment containing the N-terminally ablated *Cyp4g17* ORF with the addition of a C-terminal poly-His tail, and introducing a 5′ NdeI and a 3′ SalI. The templates for the amplification of *Cyp4g16* and *Cyp4g17* ORFs were cDNA of adult mosquitos RNAs and a pGem:CYP4G17 plasmid (previously constructed), respectively. PCR conditions

were 95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 56°C for 30 s, 72 for 1 min and 50 sec, and a final extension step of 72°C for 2 min.

2.2.3 Sub-cloning (pGem-T easy vector) and sequencing

All PCR products described above were purified by using the PCR clean-up gel-extraction kit (Mancherey Nagel), according to the manufacturer's protocol. Then, they were cloned into pGEM-T easy vector, according to the manufacturer's instructions (Promega Corporations). DH5a or Dam^-/Dmc^- DH5a (for methylated targets of enzymatic digestion) competent cells were transformed with the ligation reaction and then cultured overnight on Luria Bertani agar plates (100 μg/ml Ampicillin, 50 mg/ml Xgal and 0.5 mM IPTG) at 37°C. Blue-white screening disassociates the negative blue and the positive non-blue (white) colonies; the latter were picked for the creation of liquid cultures and subsequently the plasmid DNA isolation through Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit (Fisher Scientific UK), according to the manufacturer's protocol, took place. Several preps were screened with EcoRI (Minotech), since the enzyme cuts twice in between the insert of a poly-linker site of the pGEM-T easy vector. More than one positive clones for each gene were sequenced (CeMIA S.A.); T7 and SP6 promoter universal primers and the appropriate internal primers for each gene (Tables S1) were used. The sequencing results are available at Supplemental Information.

2.2.4 Cloning (pPelican vector)

2.2.4.1 *Cyp4g15* and *Cyp4g1Δinsertion*

After the validation of the PCR products sequence, *Cyp4g15* and *Cyp4g1Δinsertion* were cloned into the recipient plasmid, using the restriction sites added. More specifically, the plasmids were digested with the appropriate enzyme combination (BssHII/XhoI) and sub-cloned into the unique MluI/XhoI sites of the vector dPelican.attB.UAS-CYP6A51[59]; modified based on pPelican vector[69] that incorporates gypsy insulator sequences flanking the expression cassette (plasmid #30)[70]. BssHII and MluI are isoschizomers and after the digestion of pPelican, rSAP (0,5 ul, NEB) was added for 30 min in order to dephosphorylate the vector's hanging ends (inactivated at 80°C for 30 min) and make the consequent ligation more precise. More specifically, the *Cyp4g15*/*Cyp4g1Δinsertion* ORFs were placed in-between the 5xUAS-promoter sequence and the SV40 polyadenylation sequence and both *de novo* UAS expression recombinant plasmids (pPelican.attB.UAS-CYP4G15 and pPelican.attB.UAS-CYP4G1Δinsertion) embodied a mini-white Drosophila marker. Again, several preps were screened and the positive clones (correct length for both *Cyp4g15* and *Cyp4g1Δinsertion* and correct arrangement for the two parts of *Cyp4g1Δinsertion*) were used to transform DH5a cells, in order to end up with high-concentration purified plasmids (like 2.2.3, but no blue-white screening took place, because the pPelican plasmid does not incorporate β-galactosidase).

2.2.4.2 *Cyp4g16* and *Cyp4g17*

In a similar manner, cloning of *Cyp4g16/17* (wt, ΔN and ΔNhis) from sequenced pGEM-T easy constructs using NdeI/NotI restriction sites into the corresponding sites of pCW or pCDF.Duet-1:MdCPR was accomplished. Furthermore, the already prepared, from other lab members, fused constructs (pCW:OmpA.CYP4G16/17-linker-ΔNAgCPR) were also tested. The pCW:P450 plasmid bears an IPTG-inducible tac-tac promoter. The pCDF.Duet-1:MdCPR bears two IPTG-inducible T7 promoters which control the expression of both *Cyp4g* (MCS1) and *Cpr* from *Musca domestica* (MCS2). The pCW:OmpA.CYP4G16/17_ΔNAgCPR construct bears an OmpA leader sequence fused to the N-terminus of each one *Cyp4g*. Outer membrane protein A (OmpA) is a conserved protein that binds non-covalently peptidoglycan, expressed in the outer bacterial membrane[71]. In addition, the CYP4G is fused via a dipeptide linker (Ser-Ser) to the N-terminally-truncated CPR from *An. gambiae*. The resulted P450 derivatives (wt, ΔN, fusion with OmpA) are shown in Figure 9.

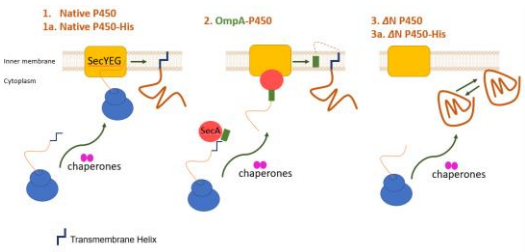


Figure 9: Schematic representation of the work flow towards P450s functional expression in bacteria. (1) Native P450s will be expressed and targeted to the plasma membrane via the intrinsic properties of their N-terminal transmembrane helix. (2) The OmpA signal peptide fused to P450s will guide them to the bacterial plasma membrane through the SecA ATPase and (3) N-terminal truncations of the predicted transmembrane helix will produce cytoplasmic P450s. Created by Vasileia Balabanidou.

2.3 Drosophila crosses

2.3.1 Crosses of CYP4G15 and CYP4G1Δinsertion

All *Drosophila* crosses are conducted according to Kefi *et al.* 2019[26] and are described in detail in the interest of this work. The *de novo* UAS expression recombinant plasmids (pPelican.attB.UAS-CYP4G15 and pPelican.attB.UAS-CYP4G1Δinsertion) were injected to preblastoderm embryos of the *D. melanogaster* strain VK13 (y[1] M{vas-int.Dm}ZH-2A w[*]; PBac{y[+]-attP-9A}VK00013, #24864 in Bloomington Drosophila Stock Center); φC31 integrase expression is accomplished under vasa promoter in X chromosome and an attP landing site is located at the 3rd chromosome[72]. The survivors of the G0 injected VK13 flies were crossed with yw flies (different eye and body color background). The G1 offspring was screened for red eyes; the w+ phenotype is proof of the recombinant plasmid integration (mini-white). For the distinct positive lines, a cross with a balancer strain for the 3rd chromosome (yw; TM3 Sb/TM6B Tb Hu) took place. The G2 progeny with red eyes and respective marker phenotype were picked out and crossed with each other in order to produce homozygous flies and establish the transgenic responder lines (Figure 10A & 11A).

2.3.2 Generation of UAS responder flies for rescue experiments

Then, two different lines were employed as a means to generate flies with oenocyte-specific *Cyp4g1* RNAi KD and/or oenocyte-specific CYP4G15/CYP4G1Δinsertion expression. On one hand, the RE-Gal4 driver line[73] carries the RE fragment of the *Desat1* gene promoter, whose expression is primarily limited to oenocytes in *Drosophila* adults; however, low levels of expression can also be observed in male accessory glands[73]. On the other hand, the UAS-*Cyp4g1*-KD responder strain (#102864KK from Vienna Drosophila Resource Center) expresses a oenocyte-specific hairpin that is responsible for the RNAi mediated KD of *Cyp4g1*.

Both strains were used so as to generate homozygous lines bearing both UAS-*dsCyp4g1* (2nd chromosome) and UAS-CYP4G15 (3rd chromosome); the series of the genetic crosses are presented in Figure 10B. Afterwards, different combinations of genetic crosses provided all the genotypes used for downstream experiments (Table 3).

Only the RE-Gal4 driver line was employed to generate homozygous lines bearing UAS-CYP4G1Δinsertion (3rd chromosome); the series of the genetic crosses are displayed in Figure 11B. Afterwards, one more cross took place with a CYP4G1-null mutant strain (Cyp4g1$^{\Delta 4}$)[28], in order to check the deleted form of CYP4G1 (CYP4G1Δinsertion) under zero background conditions (Figure 11C).

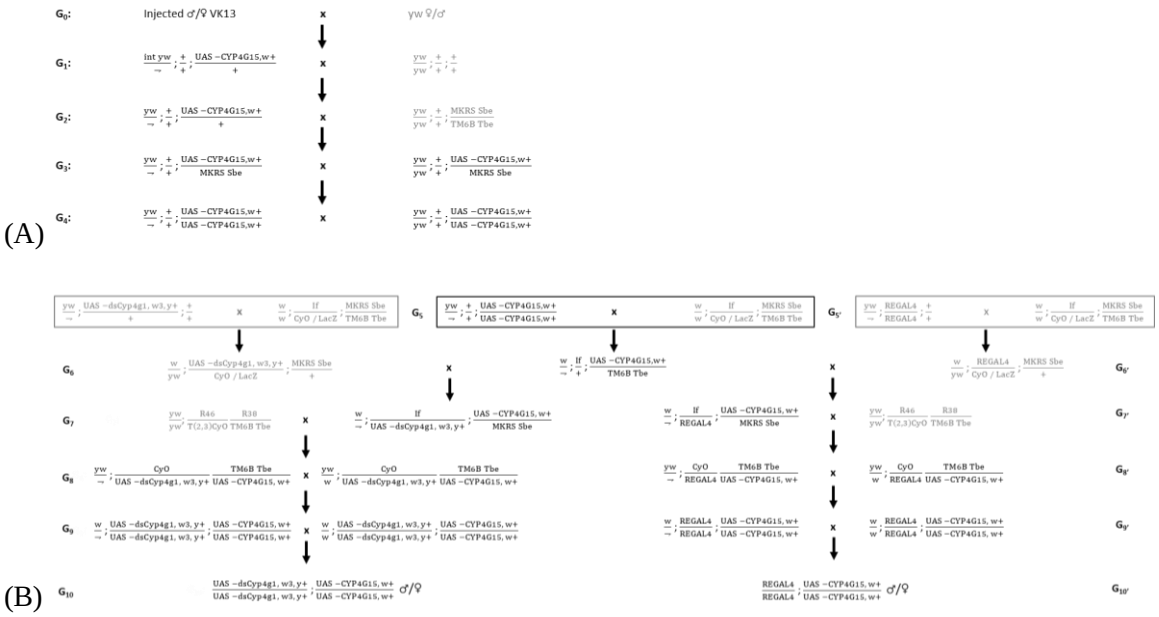


Figure 10: Series of genetic crosses for the generation of flies with both oenocyte-specific *Drosophila Cyp4g1* RNAi knock-down and CYP4G15 expression. (A) Injections in *Drosophila* embryos with plasmids containing CYP4G15 were followed by a series of standard genetic crosses to end up with 3rd chromosome homozygotes for CYP4G15. (B) Series of crosses for construction of double homozygote flies containing CYP4G15 (3rd chromosome) with *dsCyp4g1* (2nd chromosome) (left). Series of crosses for construction of double homozygote flies containing CYP4G15 (3rd chromosome) with REGal4 (2nd chromosome) (right). Adapted from Kefi *et al.* 2019 [26].

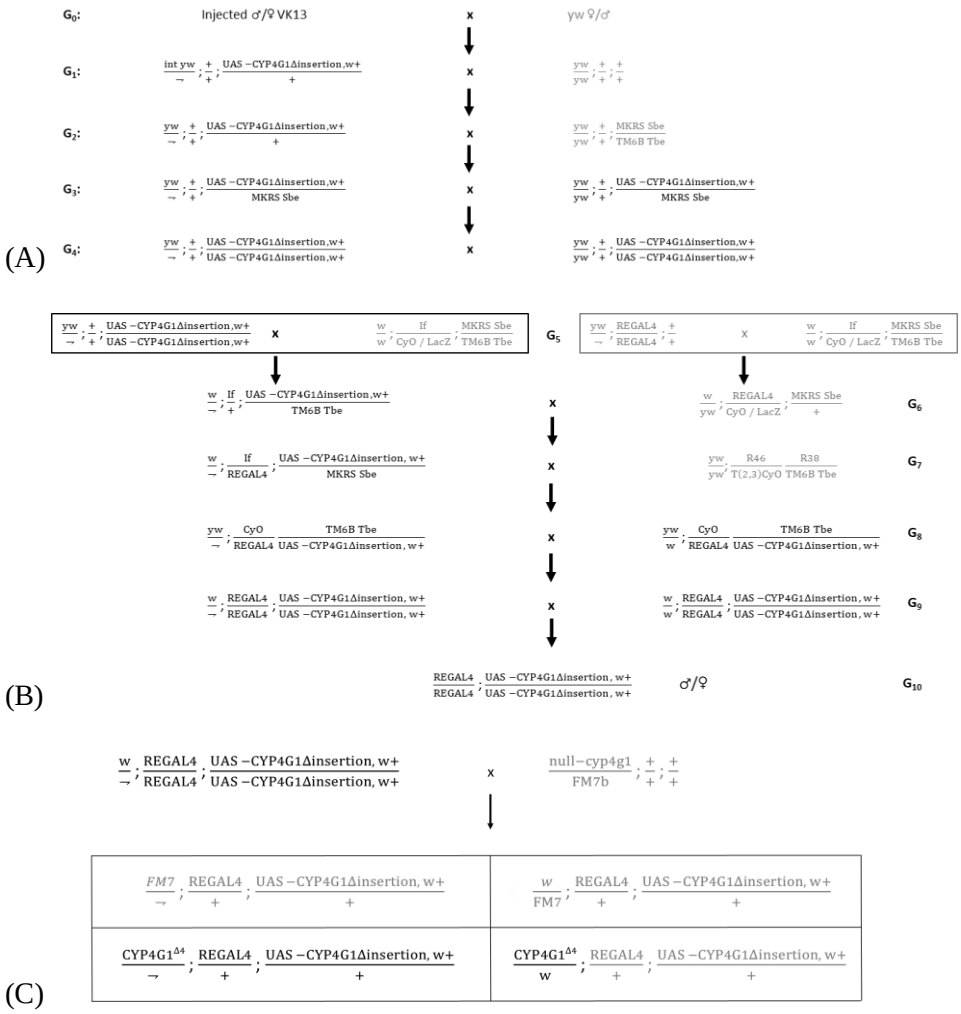


Figure 11: Series of genetic crosses for the generation of flies with oenocyte-specific CYP4G1Δinsertion expression under Cyp4g1$^{\Delta 4}$ background. (A) Injections in *Drosophila* embryos with plasmids containing CYP4G1Δinsertion were followed by a series of standard genetic crosses to end up with 3rd chromosome homozygotes for CYP4G1Δinsertion. (B) Series of crosses for construction of double homozygote flies containing CYP4G1Δinsertion (3rd chromosome) with REGal4 (2nd chromosome). Adapted from Kefi *et al.* 2019 [26]. (C) The genetic cross that generates the males with oenocyte-specific CYP4G1Δinsertion expression under Cyp4g1$^{\Delta 4}$ background.

Table 3: Combinations of crosses for the production of all genotypes used for downstream experiments (eclosion estimation and phenotypic observation of flies). A color code is illustrated: males (black), females (white), negative (red) and positive (green) control, one (purple) and two (dark blue) copies of CYP4G15.

♀ \ ♂	UAS −dsCyp4g1 / UAS −dsCyp4g1 ; + / +	REGAL4 / REGAL4 ; UAS −CYP4G15 / UAS −CYP4G15	UAS −dsCyp4g1 / UAS −dsCyp4g1 ; UAS −CYP4G15 / UAS −CYP4G15
REGAL4 / REGAL4 ; + / +	REGAL4 / UAS −dsCyp4g1 ; + / +	-	REGAL4 / UAS −dsCyp4g1 ; UAS −CYP4G15 / +
REGAL4 / REGAL4 ; UAS −CYP4G15 / UAS −CYP4G15	REGAL4 / UAS −dsCyp4g1 ; UAS −CYP4G15 / +	REGAL4 / REGAL4 ; UAS −CYP4G15 / UAS −CYP4G15	REGAL4 / UAS −dsCyp4g1 ; UAS −CYP4G15 / UAS −CYP4G15
UAS −dsCyp4g1 / UAS −dsCyp4g1 ; UAS −CYP4G15 / UAS −CYP4G15	-	-	UAS −dsCyp4g1 / UAS −dsCyp4g1 ; UAS −CYP4G15 / UAS −CYP4G15

2.3.3 CRISPR of CYP4G1

2.3.3.1 Strategy of genome editing

The CRISPR-Cas9 strategy (Figures 12A) was incorporated with the purpose of generating a deleted form of the insertion of *Drosophila* CYP4G1[74,75]. First of all, *Cyp4g1* sequence was obtained from FlyBase (http://flybase.org/), the database of Drosophila genes and genomes. Based on this sequence, multiple CRISPR targets in the examined region were suggested by the online tool Optimal Target Finder (http://targetfinder.flycrispr.neuro.brown.edu/)[74]; two targets located upstream (Start) and downstream (End) of the insertion, with no predicted off-target effects, were chosen (Figures 12B). Based on those targets, single-stranded DNA oligos were designed (Figures 10C & Supplemental Table 1). Each ssDNA pair was heated at 100ºC for 5 min and left to cool slowly until the mix reached RT in order for them to anneal and create a double-stranded DNA oligo (Figures 12C). The dsDNA oligos have 5' and 3' single stranded overhangs, facilitating the ligation into digested with BbsI and dephosphorylated gRNA vector pU6-BbsIchiRNA; a RNA expressing plasmid[74].

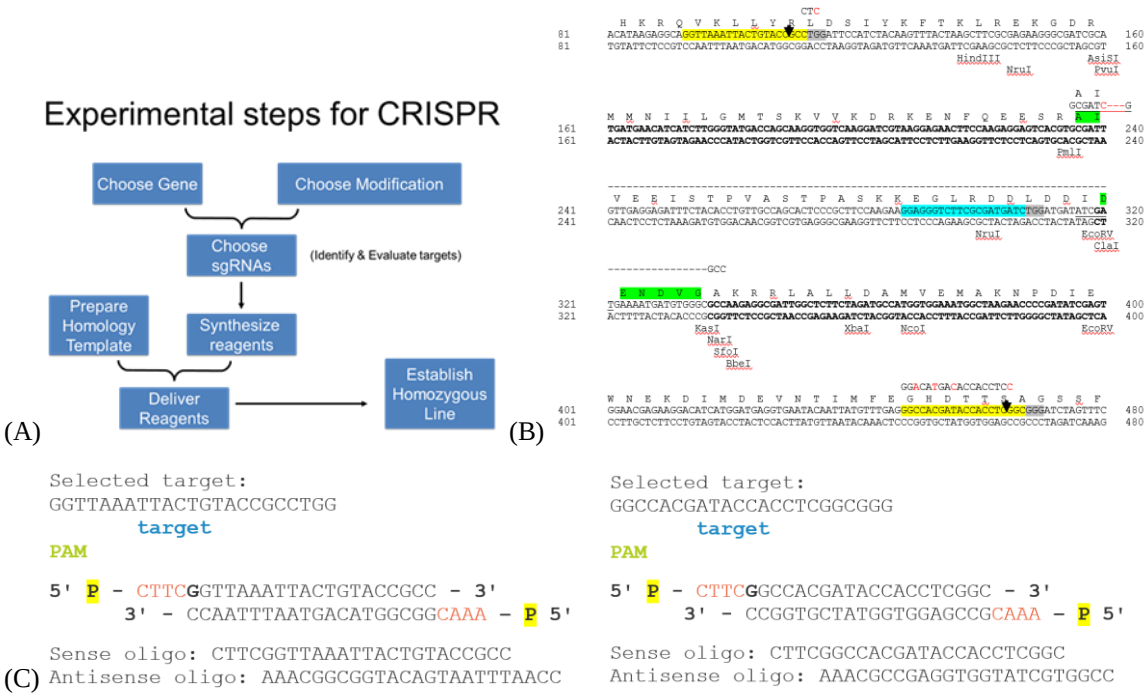


Figure 12: CRISPR design strategy. (A) The experimental steps for CRISPR application. (B) The two targets located upstream (1st arrow) and downstream (2nd arrow) of the insertion (green boundaries). (C) The dsDNAs produced by the ssDNAs of the two targets.

After the ligation, the standard protocol of transformation into DH5a competent cells and plasmid purification took place (section 2.2.3). Five different colonies from each plasmid were checked accordingly, initially with PCR (T7 universal primer, Supplemental Table 1) and then by sequencing (CeMIA S.A.). For the generation of a deleted form of

the insertion of CYP4G1, a donor plasmid was designed and ordered by Genscript (pUC57). It comprised two ~1000 bp homology arms (to facilitate HDR) at both sides of the deleted region (96 bp) and had various synonymous mutations in the gRNAs and PAM sequences so as to avoid possible DSBs in the donor plasmid and/or the HDR-modified flies by the endogenous CRISPR mechanism (Figure 13). The endogenous CRISPR mechanism is present due to the expression of the endonuclease Cas9 under the control of the promoting element nanos of the lab strain y1 M{nos-Cas9.P}ZH-2A w* (Bloomington Drosophila stock center)[75].

TCACTCAAAGGCGGTAGTTGCCAGGATGGGCTTAGAAGCGGGCAAATCCTTTGATGTTCATGACTATATGTCGCAGACCACGGTTGACATCC
TGTTGTCTACCGCCATGGGTGTGAAGAAGCTTCCGGAGGGTAACAAGAGTTTCGAATACGCCCAAGCCGTCGTCGACATGTGTGATATCATA
CATAAGAGGCAGGTTAAATTACTGTACCGCCTGGATTCCATCTACAAGTTTACTAAGCTTCGCGAGAAGGGCGATCGCATGATGAACATCAT
CTTGGGTATGACCAGCAAGGTGGTCAAGGATCGTAAGGAGAACTT|CCAAGAGGAGTCACGTGCGATTGTTGAGGAGATTTCTACACCTGTT
GCCAGCACTCCCGCTTCCAAGAAGGAGGGTCTTCGCGATGATCTGGATGATATCGATGAAAATGATGTGGGCGCCAAGAGGCGATTGGCTC
TTCTAGATGCCATGGTGGAAATGGCTAAGAACCCCGATATCGAGTGGAACGAGAAGGACATCATGGATGAGGTGAATACAATTATGTTTGA
GGGCCACGATACCACCTCGGCGGGATCTAGTTTCGCCCTCTGCATGATGGGAATCCACAAGGACATCCAGGCTAAAGTCTTCGCCGAACAGA
AGGCCATCTTCGGGGATAATATGCTGAGGGATTGCACCTTTGCCGATACCATGGAGATGAAATATTTGGAGCGCGTAATTTTAGAGACTTTG
AGGTTGTACCCACCAGTACCACTTATCGCCAGGCGTCTGGACTACGACCTGAAGTTGGCCAGTGGTCCGTACACGGTTCCCAAGGGCACTAC

Figure 13: The donor plasmid appropriate for the CRISPR strategy. The synonymous mutations in the sgRNAs (yellow) and PAM (grey) sequences are depicted in red. The start of the two homology arms is indicated with a dashed green line. The 96 bp deleted region is illustrated with the gray strikethrough line.

2.3.3.2 genetic crosses for the generation of modified flies

All the plasmids described in the previous section (2x pU6-BbsI chiRNA and the donor plasmid) were injected (final concentration of 100 ng/ul for pCU57 and 400 ng/ul for the two pU6-BbsIchiRNA) to approximately 500 *Drosophila* eggs of the nanos-Cas9 strain. The surviving instar larvae 24 hours after the injection procedure were collected and transferred into standard fly-food. The emerged adults, each one considered as a different line, however are not modified, because nanos is an embryonic marker, which is expressed in the posterior pole of the egg during oogenesis, and thus HDR occurs only in the pole cells of the egg[76]. That is why these adults (G0) were backcrossed with nanos-Cas9 flies. The progeny (G1) was pooled in batches of 30 pupae and genomic DNA was extracted (section 2.2.1) in order to perform screening PCR (section 2.3.3.3). G1 flies originated from positive G0 lines were again backcrossed with nanos-Cas9 flies and then screened for the identification of positive heterozygotes. According to mendelian inheritance, the 50% of the G2 generation is expected to be positive for the modification. G2 females originated from positive G1 lines, thus potential heterozygotes, were crossed with male flies from strain w+oc/Fm7yBHw that contain an X chromosome balancer and can maintain the mutation at a heterozygous state. Similarly, the 50% of the G3 generation is expected to be modified. Subsequently, these G2 females were screened and the progeny females (G3) from positive G2 flies were crossed again with balancer male flies. Now, non-Bar G4 males (selection against the balancer) are hemizygous, meaning that they have solely the modified allele to survive. If the CRISPR event is not lethal, G4 individuals are crossed among themselves so as to produce homozygous flies (selection against the balancer) and establish the transgenic responder lines. Figure 14 shows all the crosses described in this paragraph.

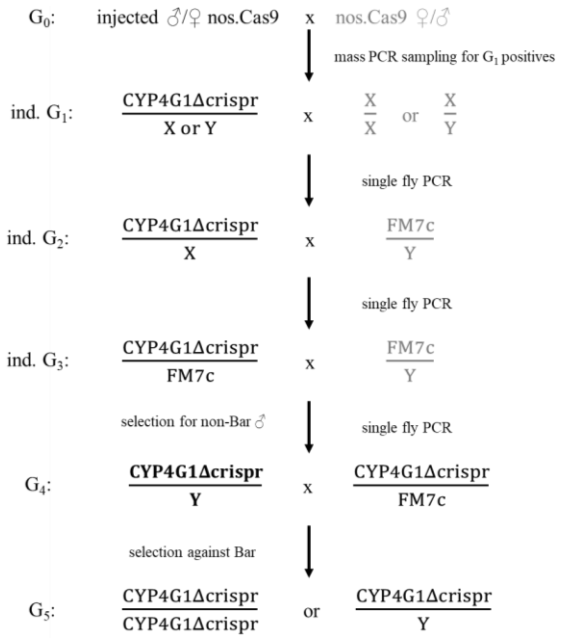


Figure 14: Nanos-Cas9 embryos are injected and surviving adults (G0) are back-crossed to nanos-Cas9. The G1 progeny is sampled (n≈ 30) and if positive, individual G1 flies are crossed to nanos-Cas9 and then screened with single fly PCR for homologous directed repair (HDR), i.e. CYP4G1Δcrispr. Individual G2 females are crossed to males of a strain bearing X chromosome balancer FM7c marked with Bar and then screened for CYP4G1Δcrispr. Individual G3 females with heterozygous Bar phenotype are crossed to the balancer strain males and then screened for CYP4G1Δcrispr. G4 females with Bar phenotype (bearing the desired mutation opposite to FM7c) are crossed with male siblings selected against Bar (i.e. hemizygous for the genome modified chromosome bearing the HDR-derived allele) and their progeny (G5) is selected against Bar to generate homozygous lines bearing the desired mutation. Adapted from Samantsidis *et al.* 2019[91].

2.3.3.3 *Screening* PCR amplification with specific primers.

PCR was performed with Kapa Taq polymerase (Kapa Biosystems) for the amplification of the possible modified *Cyp4g1* sequence. Two sets of primer pairs were used: primers Start F and END R result in a fragment of 2180 bp for the wild-type and 2079 bp for the CRISPR form and primers Delta Cyp4g1 seq and 4G1Rq result in a fragment of 843 bp for the wild-type and 742 bp for the CRISPR form (Supplemental Table 1). The templates for the amplification of *Cyp4g1* were pupae pools, single flies and cDNA of adult Drosophila RNAs. PCR conditions were 94°C for 2 min, followed by 34 cycles of 94°C for 45 s, 57°C for 30 s, 72°C for 1 or 2 min and a final extension step of 72°C for 10 min. All molecular screening experiments of the modified flies were conducted along with two positive controls, pGEM:CYP4G1Δinsertion plasmid already available containing the deleted form (section 2.2.3) for monitoring the size difference between the deleted and the wild type *Cyp4g1* and positive G1 gDNA from pupae pools, negative control from nanos-Cas9 gDNA and blank (no template).

2.4. Quantification of eclosion (adult survival and adult mortality)

So as to quantify the percentage of successful eclosion of the CYP4G15-transgenic flies, the steps described in Kefi *et al.* 2019 [26] were followed (described in detail again here). Five virgin females and 5 males of the appropriate genotypes were crossed. After approximately 5 days, 20 to 25, 3rd instar larvae were transferred into fly-food (approximately 100 larvae per biological replicate). Pupae were then measured to evaluate the pupation efficiency and successfully eclosed adults were counted. To address eclosion, we calculated the adult survivors (females and males) after 3 days, while newly emerged

adults that deceased right after eclosion were measured distinctly so as to address the adult mortality, in three biological replicates.

2.5 RNA extraction

RNA from *D. melanogaster* was extracted with PicoPure™ RNA Isolation Kit (Thermo Fisher), following the instructions of the manufacturer.

2.6 cDNA synthesis

cDNA from *D. melanogaster* RNA was synthesized with MINOTECH RT polymerase (MINOTECH), following the instructions of the manufacturer.

2.7 Reverse-transcriptase (RT) PCR

Reverse-transcriptase (RT) PCR was performed with GoTaq® Flexi DNA Polymerase (Promega) for the amplification of *Cyp4g15* and the potentially modified *Cyp4g1* sequences. The templates for the amplification of *Cyp4g15* were cDNAs of the different transgenic flies and the primers BssHII 4G15F and CYP4G15R resulting in a fragment of 497 bp were employed (Supplemental Table 1). The templates for the amplification of *Cyp4g1* were cDNAs of the different transgenic flies and the primers ΔCYP4G1seq and 4G1Rq resulting in a fragment of 843 bp for the wild type and 742 bp for the CYP4G1Δinsertion form were employed (Supplemental Table 1). For both cases, RPL11_Fq and RPL11_Rq for the reference gene *RPL11* were used in order to be able to monitor the comparable cDNA presence among the different samples tested. PCR conditions were 95°C for 2 min, followed by 25 cycles of 95°C for 30 s, 57°C (4G1) or 58ºC (4G15) for 30 s, 72°C for 50 sec and a final extension step of 72°C for 5 min. All molecular screening experiments of the modified flies were conducted along with two negative controls, no-RT and no-template control.

2.8 Antibodies

The rabbit polyclonal antibodies that were used for the specific targeting of CYP4G16 and CYP4G17 have previously been described (Balabanidou *et al*., 2016 [24] and Ingham *et al*., 2014 [77], respectively). The antibodies against penta-His were purchased from Qiagen.

2.9 Western blots

Bacterial cultures before and after (4h and 24h) IPTG induction were homogenized into 5x protein loading buffer (100 mM Tris-HCL pH 6.8, 10% SDS, 40% glycerol and 0.4% bromophenol blue). More specifically, 1 ml of cultures before IPTG induction were lysed and homogenized into 50 ul 5x protein loading buffer. 1 ml of cultures 4h after IPTG induction were homogenized into 100 ul 5x protein loading buffer. 200 ul of cultures 24h

after IPTG induction were homogenized into 100 ul 5x protein loading buffer. Abdominal walls from adult mosquitos or flies were homogenized into homogenization buffer in RIPA buffer, supplemented with protease inhibitors (1% EDTA, 1% EGTA, 1% PMSF and 1% phosphate inhibitors). Approximately, 10 ul of homogenization buffer were used for every mosquito/fly. Next, the extracted polypeptides were separated according to their size on a 10% acrylamide SDS-PAGE and the separated polypeptides were electro-transferred on nitrocellulose membrane (GE Healthcare, Whatman) and subsequently probed with anti-CYP4G16 or anti-CYP4G17 or anti-His. The dilution of the antibodies used were 1:250 in 1x TBS-Tween (TBST) 5% skimmed milk for the first three and 1:2000 dilution in 1xTBST supplemented with 3%BSA for the last one, respectively. A secondary antibody coupled to horseradish peroxidase (Cell Signaling) was used (goat anti-rabbit for CYP4G16/17 and goat anti-mouse for penta-his, diluted 1:5000 in 1% or 3% skimmed milk in TBS-Tween buffer, respectively) and visualized using a horseradish peroxidase sensitive ECL Western blotting detection kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The result was recorded using ChemiDoc Imaging System (Bio-Rad).

2.10 Expression system

2.10.1 Bacterial strains

The bacterial strains used for the CYP4G16/17 expression are listed in Table 4. All of them are chemically competent *E. coli* cells suitable for high efficiency transformation and are appropriate for expressing proteins. Furthermore, Lemo21 (DE3) and C43 cells are effective in overexpressing toxic membrane proteins.

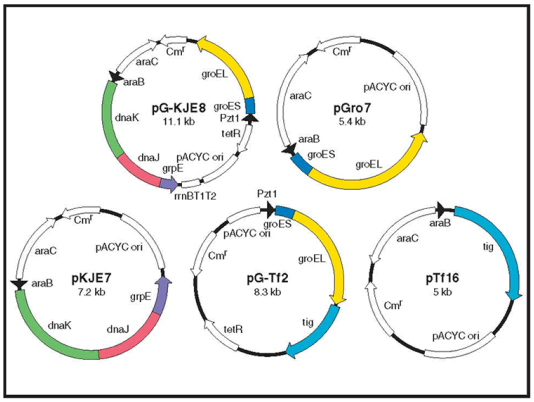
Table 4: The bacterial strains tested in this work.

BACTERIAL STRAIN	SOURCE
DH5a	https://international.neb.com/products/c2987-neb-5-alpha-competent-e-coli-high-efficiency#Product%20Information
JM109	https://www.chem-agilent.com/pdf/strata/200235.pdf
Lemo 21 (DE3)	https://international.neb.com/products/c2528-lemo21de3-competent-e-coli#Product%20Information
C43	http://wolfson.huji.ac.il/expression/procedures/bacterial/lucigen'sC43_41strains.pdf
BL21 (DE3) Star	https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Foneshotbl21star_man.pdf&title=T25lIFNob3QgQkwyMSBTdGFyIChERTMpIGFuZCBPbmUgU2hvdCBCTDIxIFN0YXIgKERFMykgcEx5c1MgQ29tcGV0ZW50IENlbGxz

2.10.2 Plasmids expressing members of the chaperone team

Additionally, plasmids overexpressing members of the chaperone team (DnaK, DnaJ, GrpE, GroE, GroES and trigger factor) in various combinations and controlled conditions (dose-dependent L-arabinose and tetracycline addition) were incorporated with the

intention of facilitating the folding, aggregation, and stability of the heterologous proteins in the *E. coli* strains (Takara Bio, Figure 15)[78,79]. Plasmids pG-KJE8 (referred in this work as plasmid 5 chaperones) and Plasmid pGTf2 (referred in this work as plasmid 3 chaperones) were used.

(A) 

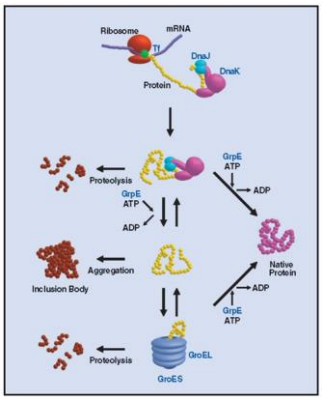
(B) 

Figure 15: (A) The Chaperone Plasmid Set consists of five different plasmids, each of which is designed to express multiple molecular chaperones that function together as a "chaperone team" to enable optimal protein expression and folding and to reduce protein misfolding. Each plasmid carries an origin of replication (ori) derived from pACYC and a chloramphenicol-resistance gene (Cm^r) gene, which allows the use of *E. coli* expression systems containing ColE1-type plasmids that confer ampicillin resistance. The chaperone genes are situated downstream of the *araB* or *Pzt-1* (tet) promoters; as a result, expression of target proteins and chaperones can be individually induced if the target gene is placed under the control of different promoters (e.g., *lac*). These plasmids also contain the necessary regulator (*araC* or tet^r) for each promoter. (B) Possible model for chaperone-assisted protein folding in *E. coli*. Adapted from Takara Bio.

2.10.3 Bacterial expression protocol

The induction of P450s was accomplished in accordance with previous works of lab members[59,60].

<u>Culture conditions (1st attempt)</u>

1. Transform *E. coli* cells with ~10 ngr plasmid DNA (pCW:P450wt, pCW:P450ΔN, pCDF.Duet-1:P450_CPR or pCW:OmpA.P450_ΔNAgCPR).
2. Pick up a single colony from the transformation plate (agar plate/ampicillin or chloramphenicol or streptomycin) and inoculate in 5 mL LB supplemented with ampicillin (100 μg/ml) or chloramphenicol (34 μg/ml) or streptomycin (50 μg/ml). Leave the pre-culture to grow for 16 h at 37^0 C, while shaking.
3. Next day, inoculate 10 ml of Terrific Broth with 100 ul of the preculture (in a 50 ml falcon tube) (1/100 dilution) also supplemented with ampicillin or chloramphenicol or streptomycin.
4. Grow the cell culture at 37^0C with 150 rpm shaking until OD_{600} = 0.9.

5. Decrease temperature at 25^0C (150 rpm shaking) and continue growing for another 30 minutes. Collect a cell culture sample (–IPTG sample).
6. Then induce cells by adding 0,5 mM IPTG, supplement with 1 mM ALA and grow for 24 hours (25^0 C, 150 rpm shaking). Collect a cell culture sample after 4h and 24h (+ IPTG samples).
7. All the –/ + IPTG samples were normalized according to their OD value at 600 nm and the corresponding proteins were extracted via homogenization in 5x protein loading buffer and centrifuging at 7000 rpm, 4^oC for 2 min (see also section 2.9).
8. Protein samples were loaded on 10% SDS gel (prepare a 10- or 15-well gel and load 5 ul per lane) and upon separation, the polypeptides were analyzed by Western blots.

<u>Culture conditions (2nd attempt)</u>

A different temperature was tested (steps four to six altered).

- Grow the cell culture at 30^0C with 150 rpm shaking until OD_{600} = 0.5-0.6.
- Collect a cell culture sample (–IPTG sample).
- Then induce cells by adding 0,5 mM IPTG, supplement with 1 mM ALA and grow for 24 hours (30^0 C, 150 rpm shaking). Collect a cell culture sample after 4h and 24h (+ IPTG samples).

2.10.4 Negative immune-absorbance experiment

Bacterial cell compartments (cytosol and membranes) expressing a control P450 were used for negative immune-absorption of anti-CYP4G17, in order to eliminate the non-specific background in Western blot analysis of bacterial cells expressing CYP4G17. The following steps were carried out:

- Pick up a single colony from the transformation plate of the reference plasmid and cell combination (agar plate/ampicillin) and inoculate in 5 mL LB supplemented with ampicillin (100 μg/ml) to grow for 16 h at 37^0 C, while shaking.
- Next day, inoculate 400 ml of Terrific Broth with 4 ml of the pre-culture (in a 2 L flask) (1/100 dilution) also supplemented with ampicillin.
- Grow the cell culture at 30^0C with 150 rpm shaking until OD_{600} = 0.5-0.6.
- Centrifuge the cell culture for 5 min, 5000 g, 4^oC.
- Re-suspend the cell pellets in ice-cold 1x PBS.

- Break the cell walls via repeated cycles of sonication with ultra-sounds (pulsed, amplitude 50) and resting on ice.
- Centrifuge for 5 min, 4000 g, 4ºC.
- Centrifuge again the resulting supernatant for 30 min, 18000 g, 4ºC. After the end, the new supernatant corresponds to the cytosol and the pellet to the bacterial membranes.
- Use membranes corresponding to 100 ml initial culture and carefully re-suspend them in 40 ul of affinity-purified anti-CYP4G17.
- Incubate over-day (8-10 h) at 4ºC on a rotator wheel.
- Centrifuge for 20 min, 18000 g, 4ºC.
- Take the supernatant and repeat the step of resuspension with another 100 ml membrane compartments.
- Incubate over-night (~16 h) at 4ºC, while rotating on a rotator wheel.
- Next day, centrifuge for 20 min, 18000 g, 4ºC.
- Mix the supernatant with 1 ml from the cytoplasmic compartments solution and 9 ml 1% skimmed milk in TBST.
- Use the resulting absorbed-antibody for the Western blot analysis of bacterial crude protein extracts.

3. Results

3.1 UAS/Gal4 mediated expression of CYP4G15 rescued the *Cyp4g1*-KD flies suggesting that CYP4G15 acts partially as decarbonylase

As described in sections 2.2.3, 2.2.4.1 and 2.3.1, pGem:CYP4G15 was sequenced (Supplemental Figure 1) and pPelican.attB.UAS-CYP4G15 was injected, respectively. Both constructs are depicted in Figure 16.

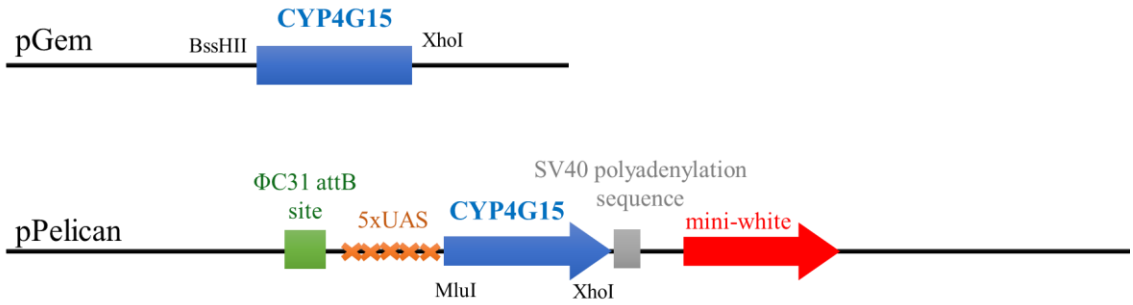


Figure 16: The pGem:CYP4G15 and pPelican.attB.UAS-CYP4G15 plasmids (see also sections 2.2.3, 2.2.4.1 and 2.3.1).

Then, as described in section 2.3, series of crosses for the production of all genotypes used for CYP4G15 ectopic expression took place. In Figure 17, these six different combinations and hence the progeny genotypes are presented; one positive control expressing both REGal4 and *dsCyp4g1*, two negative controls expressing either REGal4 or *dsCyp4g1* in combination with CYP4G15, two similar backgrounds expressing both REGal4 and *dsCyp4g1* in combination with one copy of CYP4G15 (different derivation crosses) and one background expressing both REGal4 and *dsCyp4g1* in combination with two copies of CYP4G15.

More specifically, the vast majority of the negative control flies successfully eclosed (Figure 17). On the contrary, almost all oenocyte-specific *Cyp4g1*-KD flies (positive control) are incapable of eclosing from the pupal case and die at emergence (Figure 17) as previously shown with two different oenocyte-specific Gal4 drivers[9,26]. This lethal phenotype could be rescued in the presence of two copies of CYP4G15, since transgenic survivors were detected (Figure 17-19). Then, based on the phenotypic evaluation, the eclosion efficiency, adult survival and adult mortality were estimated (section 2.4 and Figure 17). In fact, ~17% of the pupae expressing CYP4G15 in two copies successfully emerged into adults, meaning that CYP4G15 could complement the function of CYP4G1 at a certain degree, indicating that CYP4G15 is a functional oxidative decarbonylase, at least in our system. In addition, ~91,6% of the efficiently eclosed survivors are males (109 out of 119). As for the one copy background is concerned, both reverse crosses provoke the same phenotype; a significant amount of newly-emerged

individuals endure the eclosion barrier but decease nearly instantaneously and are spotted lying dead on the food. This indicates that the transgene rescues the lethal phenotype in a dose-dependent manner. Photos of flies arrested during eclosion and survivors are shown in Figure 18. Moreover, the reverse-transcriptase (RT) PCR for the different CYP4G15 backgrounds verifies the oenocyte-specific expression of CYP4G15 (Figure 19).

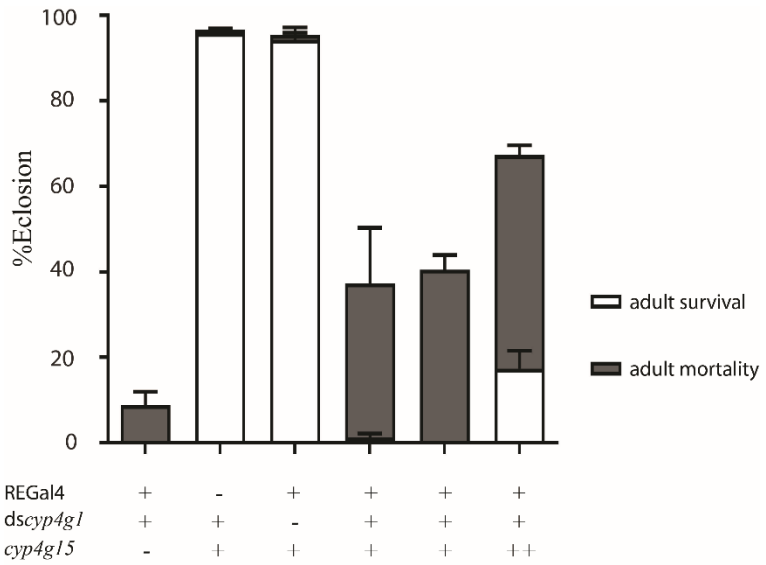


Figure 17: Percent eclosion of *D. melanogaster* flies in different CYP4G backgrounds. Quantification of adult flies that successfully eclosed corresponding to a known number of pupae. White bars represent successfully eclosed adults that survived (%), while flies that died as newly-emerged adults lying on the food were calculated to address mortality post successful eclosion (%) and are depicted with grey bars. Different CYP4G backgrounds are described at the bottom of the graph with "+" representing the presence and "–" the absence of a P450 gene (*Cyp4g1* and *Cyp4g15*) or the oenocyte-specific GAL4 driver (REGal4). Mean of 3 biological experiments +SEM. Adapted by Kefi *et al.* 2019[26].

(A)

(B)

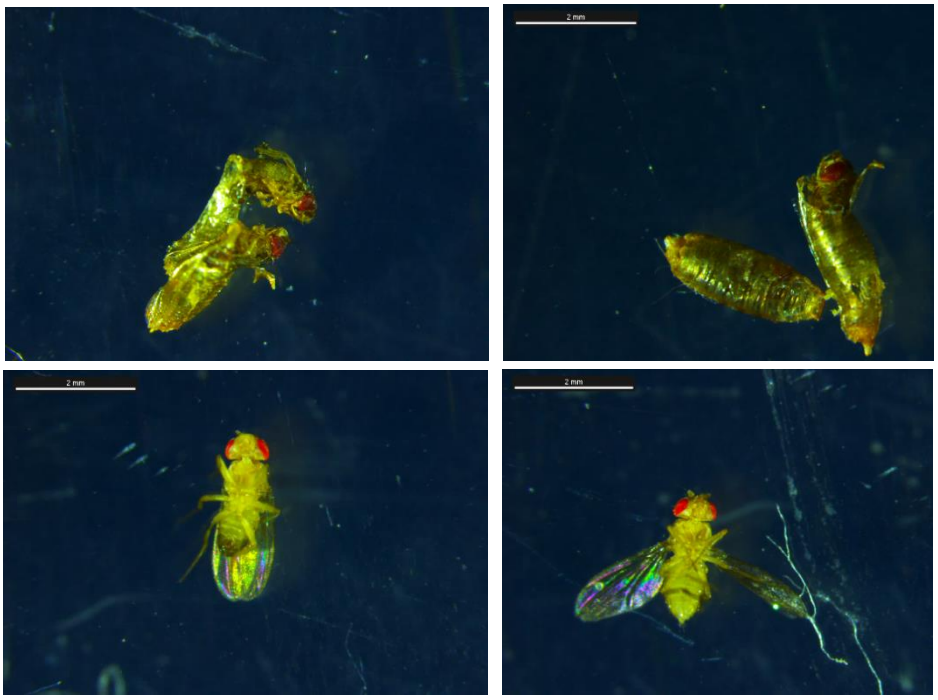


Figure 18: The different phenotypes of adult flies expressing two copies of CYP4G15. (A) Flies arrested during eclosion and (B) successfully eclosed survivors (left: male, right: female).

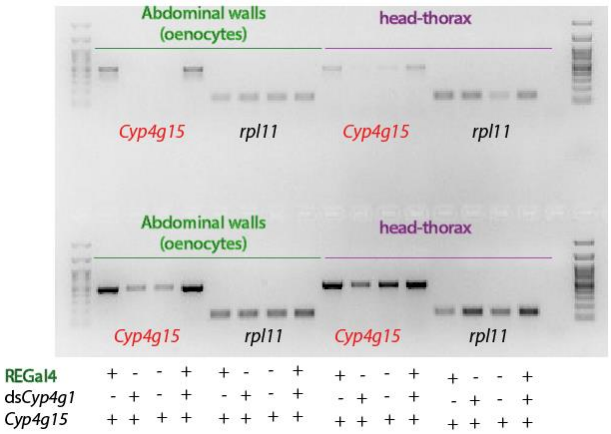


Figure 19: Reverse-transcriptase (RT) PCR for the abdominal walls and rest body of $\frac{\text{REGAL4}}{\text{REGAL4}}$; $\frac{\text{UAS}-\text{CYP4G15, w+}}{\text{UAS}-\text{CYP4G15, w+}}$, $\frac{\text{UAS}-\text{dsCyp4g1, w3, y+}}{\text{UAS}-\text{dsCyp4g1, w3, y+}}$; $\frac{\text{UAS}-\text{CYP4G15, w+}}{\text{UAS}-\text{CYP4G15, w+}}$, $\frac{+}{+}$; $\frac{\text{UAS}-\text{CYP4G15, w+}}{\text{UAS}-\text{CYP4G15, w+}}$ and $\frac{\text{REGAL4}}{\text{UAS}-\text{dsCyp4g1, w3, y+}}$; $\frac{\text{UAS}-\text{CYP4G15, w+}}{\text{UAS}-\text{CYP4G15, w+}}$ flies. *Cyp4g1* (tested) and *Rpl11* (used as quantity control) genes are targeted; 25 and 30 cycles are displayed top and bottom, respectively, in 1% agarose gel. Different CYP4G backgrounds are described at the bottom of the graph with "+" representing the presence and "–" the absence of a P450 gene (*dsCyp4g1* and *Cyp4g15*) or the oenocyte-specific GAL4 driver (REGal4).

3.2 CRISPR-mediated genome editing of CYP4G1 bearing the deletion for the CYP4Gs specific insertion results in flies surviving up to pharate adult developmental stage

As we discussed above (section 2.3.3), a series of crosses and screenings took place in order to possibly end up with the hemizygous males and homozygous females bearing the deletion for the specific 4G insertion. More specifically, 17 G0 adult survivors post injection were backcrossed with nanos cas9 flies. The G1 pupae pool screening showed that 3 out of 17 lines (4, 6 and 14) were positive for the modification (Figure 20).

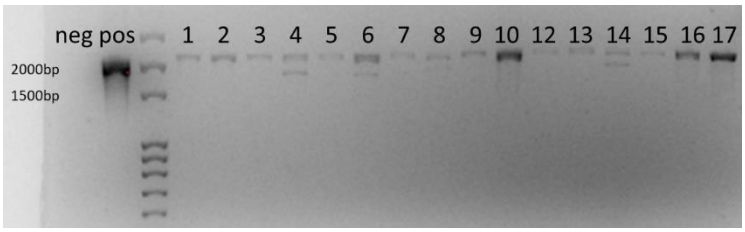
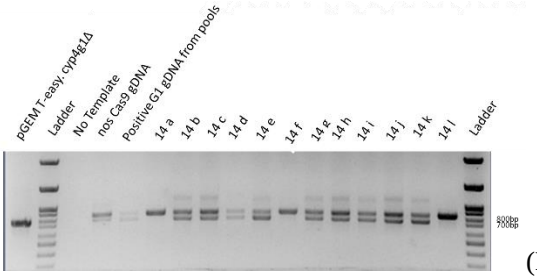
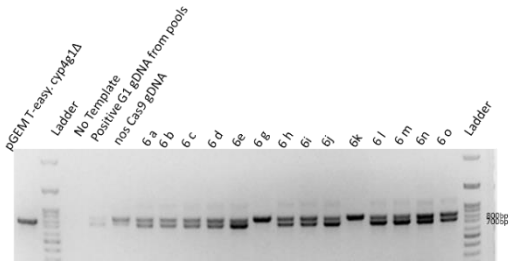
(A) 

Figure 20: Diagnostic screening of G1 pupae pools with specific primers yielding 2180 bp (wt) and 2079 bp (CRISPR) *Cyp4g1* products in 2% agarose gel. Neg indicates the blank (distilled water instead of the DNA template) and pos indicates the positive control (pGEM T-easy. cyp4g1Δinsertion plasmid). The numbers indicate the different lines.

G1 single-fly PCR screening from these 3 lines identified many positive heterozygotes, as shown in Figure 21. More specifically, line 14 revealed 9 (14b-e, g-k) out of 12 (14a-l) positive offspring, line 6 revealed 12 (6a-e, h-j, l-o) out of 14 (6a-o) positive offspring and line 4 revealed 11 (4b-f, h-j, l-o) out of 14 (4a-o) positive offspring.

(A) 

(B) 

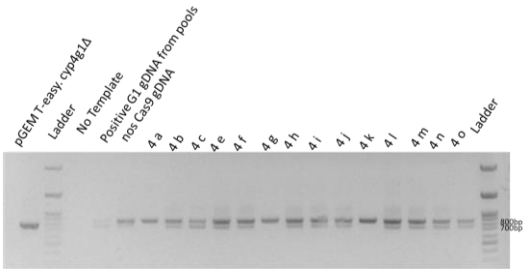
(C) 

Figure 21: Diagnostic screening of G1 single flies with specific primers yielding 843 bp (wt) and 742 bp (CRISPR) *Cyp4g1* products. pGEM T-easy. cyp4g1Δinsertion indicates the positive control of the deleted

form, No Template indicates the blank (distilled water instead of the DNA template), positive G1 gDNA from pools indicates the positive control and nos Cas9 gDNA indicates the negative control. (A) G1 single-fly screening from individuals of line 14. (B) G1 single-fly screening from individuals of line 6. (C) G1 single-fly screening from individuals of line 4.

G2 single-fly screening from some of the above positive flies also resulted to positive heterozygotes; the majority of the G1 flies were either sterile or died before leaving offspring. Here, G2 single-fly screening from line 4e revealed 5 (4e_2 - 4e_6) out of 7 (4e_1 - 4e_7) positive progeny and line 4f revealed 1 (4f_5) out of 7 (4f_1-4f_7) positive progeny. G2 single-fly screening from line 6e revealed 2 (6e_2 and 6e_3) out of 7 (6e_1 - 6e_7) positive progeny and line 6j revealed 4 (6j_1, 6j_4, 6j_6, 6j_7) out of 8 (6j_1 – 6j_8) positive progeny. G2 single-fly screening from line 14e revealed 3 (14e_1, 14e_2, 14e_8) out of 6 positive progeny and line 14i revealed 5 out of 7 (14i_1, 14i_3 - 14i_5, 14i_8) positive progeny. The PCR diagnostic screening based on the size difference of the deleted and the wt allele is shown in Figure 22.

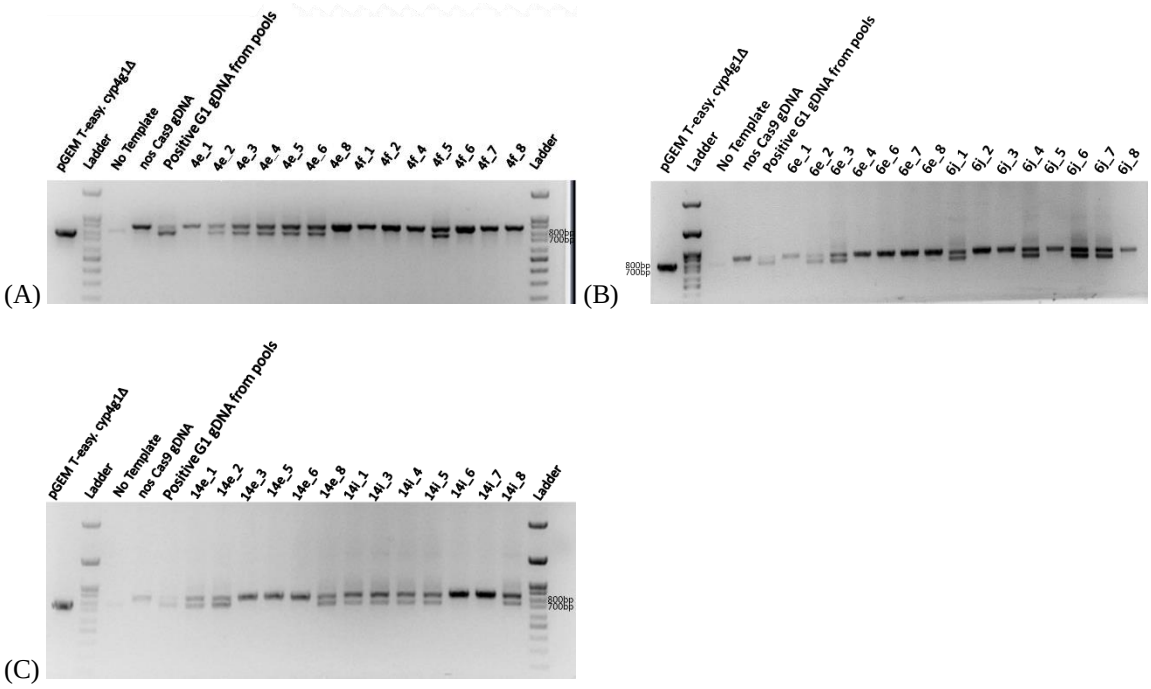


Figure 22: Diagnostic screening of G2 single flies with specific primers yielding 843 bp (wt) and 742 bp (CRISPR) *Cyp4g1* products. pGEM T-easy. cyp4g1Δinsertion indicates the positive control of the deleted form, No Template indicates the blank (distilled water instead of the DNA template), positive G1 gDNA from pools indicates the positive control and nos Cas9 gDNA indicates the negative control. (A) G2 single-fly screening from individuals of lines 4e and 4f. (B) G2 single-fly screening from individuals of lines 6e and 6j. (C) G2 single-fly screening from individuals of lines 14e and 14i.

The G3 screening (5-8 individuals per line) was held to a random of lines as shown in Figure 23. Lines 6e_3, 14_e, 4e_4, 6j_7, 4e_2, 14j_4, 6j_4 were tested for positive

CRISPR event.

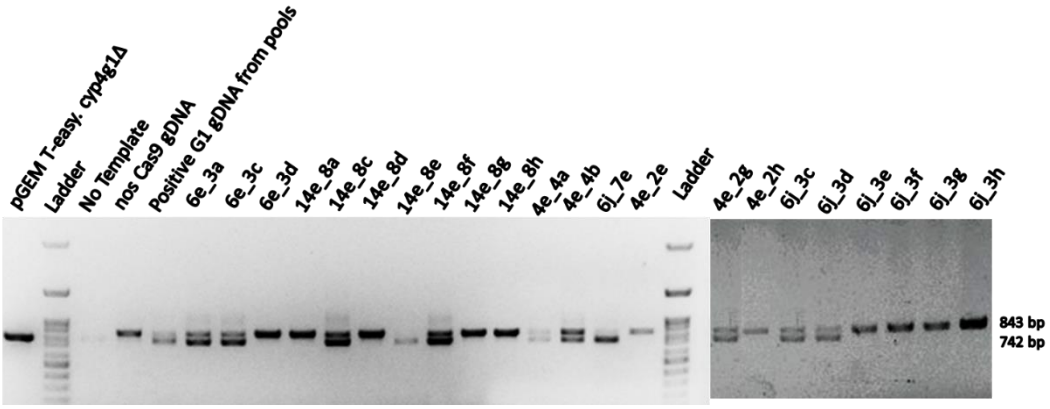
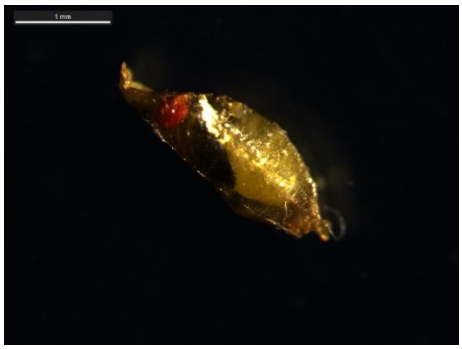
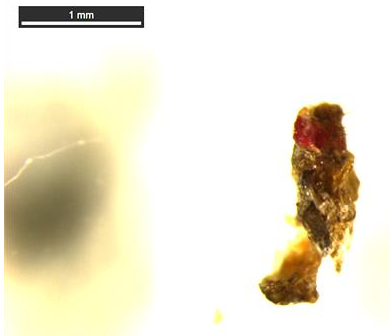


Figure 23: Diagnostic screening of different G3 fly pools with specific primers yielding 843 bp (wt) and 742 bp (CRISPR) *Cyp4g1* products. pGEM T-easy. cyp4g1Δinsertion indicates the positive control of the deleted form, No Template indicates the blank (distilled water instead of the DNA template), positive G1 gDNA from pools indicates the positive control and nos Cas9 gDNA indicates the negative control.

However, G4 showed no adult modified (non-bar) males, which means that our strategy leads to a lethal phenotype; only pharate individuals were observed. More specifically, flies arrested during adult eclosion were dissected from the pupal case; they were identified as the CYP4G1Δcrispr hemizygotes, based on the gender (male genitals) and the eye marker (red and non-bar). The phenotype of all flies described are shown in Figure 24. In other words, the modification is balanced only at a heterozygous female and homozygous flies are unable to occur. These lines are maintained in G3 generation and at heterozygous state. Figure 25 illustrates the alignment of the amino-acid sequences of CYP4G1wt, CYP4G1Δcrispr and CYP4G1Δinsertion.

(A) 



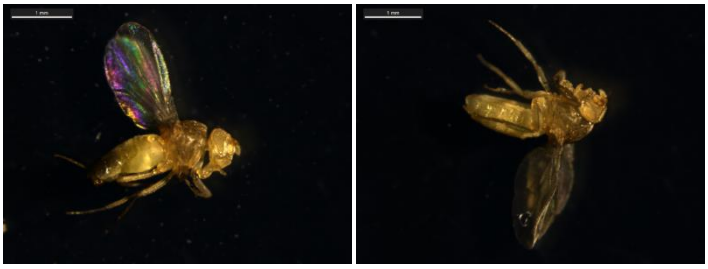
(B)  (C) 

Figure 24: The different phenotypes and genotypes (only for the X chromosome) of the CRISPR-flies. (A) CRISPR males with genotype: $\frac{\text{X\_HDR}}{\text{Y}}$. These CYP4G1Δcrispr hemizygotes arrested during adult eclosion (pupae left) are also shown dissected from the puparial case (right). (B) CRISPR females with genotype: $\frac{\text{X\_HDR}}{\text{FM7c}}$. (C) The FM7c balancer strain (left: male, $\frac{\text{FM7c}}{\text{Y}}$ and right: female, $\frac{\text{FM7c}}{\text{FM7c}}$).

```
cyp4g1      QHLTKAEEYRYFKPWFGDGLLISNGHHWRHHRKMIAPTFHQSILKSFVPTFVDHSKAVVA      180
uas         QHLTKAEEYRYFKPWFGDGLLISNGHHWRHHRKMIAPTFHQSILKSFVPTFVDHSKAVVA      180
crispr      QHLTKAEEYRYFKPWFGDGLLISNGHHWRHHRKMIAPTFHQSILKSFVPTFVDHSKAVVA      180
            ************************************************************

cyp4g1      RMGLEAGKSFDVHDYMSQTTVDILLSTAMGVKKLPEGNKSFEYAQAVVDMCDIIHKRQVK      240
uas         RMGLEAGKSFDVHDYMSQTTVDILLSTAMGVKKLPEGNKSFEYAQAVVDMCDIIHKRQVK      240
crispr      RMGLEAGKSFDVHDYMSQTTVDILLSTAMGVKKLPEGNKSFEYAQAVVDMCDIIHKRQVK      240
            ************************************************************

cyp4g1      LLYRLDSIYKFTKLREKGDRMMNIILGMTSKVVKDRKENFQEESRAIVEEISTPVASTPA      300
uas         LLYRLDSIYKFTKLREKGDRMMNIILGMTSKVVKDRKENFQEESRA--------------      286
crispr      LLYRLDSIYKFTKLREKGDRMMNIILGMTSKVVKDRKENFQEESRAI-------------      287
            *********************************************

cyp4g1      SKKEGLRDDLDDIDENDVGAKRRLALLDAMVEMAKNPDIEWNEKDIMDEVNTIMFEGHDT      360
uas         ------------IDENDVGAKRRLALLDAMVEMAKNPDIEWNEKDIMDEVNTIMFEGHDT      334
crispr      -------------------AKRRLALLDAMVEMAKNPDIEWNEKDIMDEVNTIMFEGHDT      328
                              **************************************

cyp4g1      TSAGSSFALCMMGIHKDIQAKVFAEQKAIFGDNMLRDCTFADTMEMKYLERVILETLRLY      420
uas         TSAGSSFALCMMGIHKDIQAKVFAEQKAIFGDNMLRDCTFADTMEMKYLERVILETLRLY      394
crispr      TSAGSSFALCMMGIHKDIQAKVFAEQKAIFGDNMLRDCTFADTMEMKYLERVILETLRLY      388
            ************************************************************

cyp4g1      PPVPLIARRLDYDLKLASGPYTVPKGTTVIVLQYCVHRRPDIYPNPTKFDPDNFLPERMA      480
uas         PPVPLIARRLDYDLKLASGPYTVPKGTTVIVLQYCVHRRPDIYPNPTKFDPDNFLPERMA      454
crispr      PPVPLIARRLDYDLKLASGPYTVPKGTTVIVLQYCVHRRPDIYPNPTKFDPDNFLPERMA      448
            ************************************************************
```

Figure 25: The alignment of the protein sequences of CYP4G1wt, CYP4G1Δinsertion and CYP4G1Δcrispr. CYP4G1Δinsertion and CYP4G1Δcrispr lack 26 and 32 amino-acids, respectively.

3.3 UAS/Gal4 mediated expression of CYP4G1 bearing a deletion for the CYP4Gs specific loop rescued CYP4G1 null mutants

As described in sections 2.2.3, 2.2.4.1 and 2.3.1, pGem:CYP4G1Δinsertion was sequenced (Supplemental Figure 2) and pPelican:attB.UAS-CYP4G1Δ was injected, respectively. Both constructs are depicted in Figure 26.

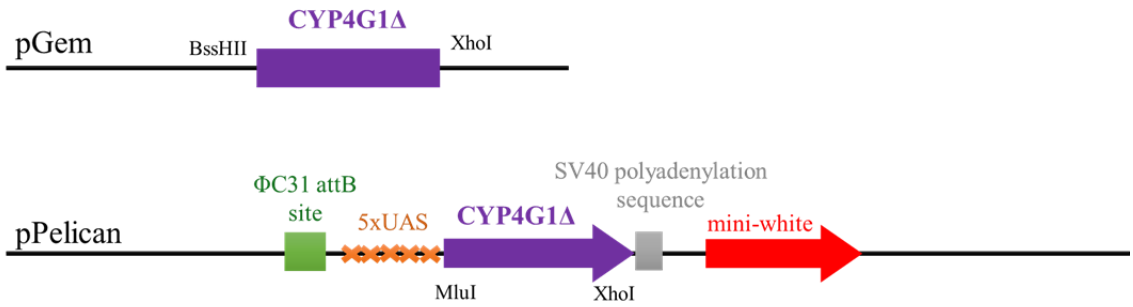


Figure 26: The pGem:CYP4G1 and pPelican.attB.UAS-CYP4G1 plasmids (see also sections 2.2.3, 2.2.4.1 and 2.3.1).

Afterwards, as described in section 2.3, combinations of crosses took place and the genetic cross that generates males with oenocyte-specific CYP4G1Δinsertion expression under CYP4G1-null background conditions was depicted in Figure 11C. These flies survive due to the induced expression of the transgene (REGal4 and UAS-CYP4G1Δinsertion), despite the fact that the native CYP4G1 is absent (CYP4G1$^{\Delta 4}$). In order to verify that, a reverse-transcriptase PCR (Figure 27) in the different *Cyp4g1* genetic backgrounds took place and revealed that the aforementioned males express only the transgene in their oenocytes (abdominal walls from each adult fly) as the transgenic allele is identified due to its ~100bp lower size comparing to the wild type one. All the successfully emerged adults from the last cross of this rescue experiment (Figure 11C) are shown in Figure 28.

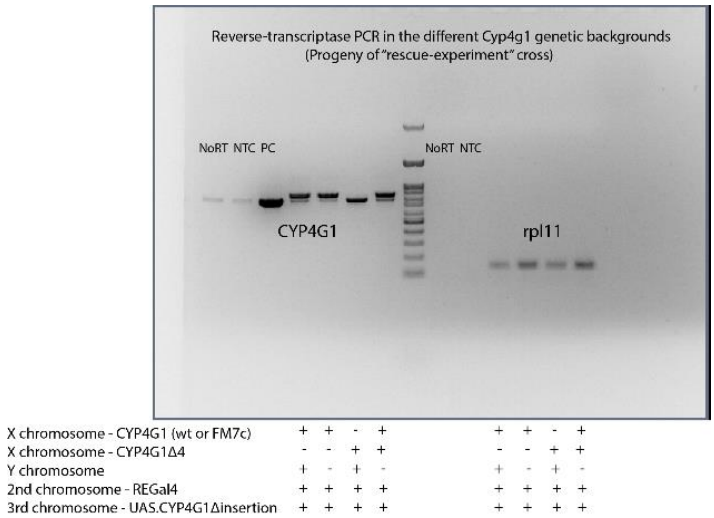


Figure 27: Reverse-transcriptase (RT) PCR of dissected abdominal walls for the four individuals of Figure 11C targeting *Cyp4g1* (tested) and *Rpl1* (used as quantity control) genes. The products of no-RT control, no-template control, positive control (pGem:CYP4G1Δinsertion), FM7 male ($\frac{FM7}{\rightarrow}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$), FM7/CYP4G1wt female ($\frac{w}{\text{FM7}}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$), CYP4G1$^{\Delta 4}$ male ($\frac{\text{CYP4G1}\Delta 4}{\rightarrow}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$) and CYP4G1$^{\Delta 4}$/CYP4G1wt female ($\frac{\text{CYP4G1}\Delta 4}{\text{w}}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$) are displayed in 1% agarose.

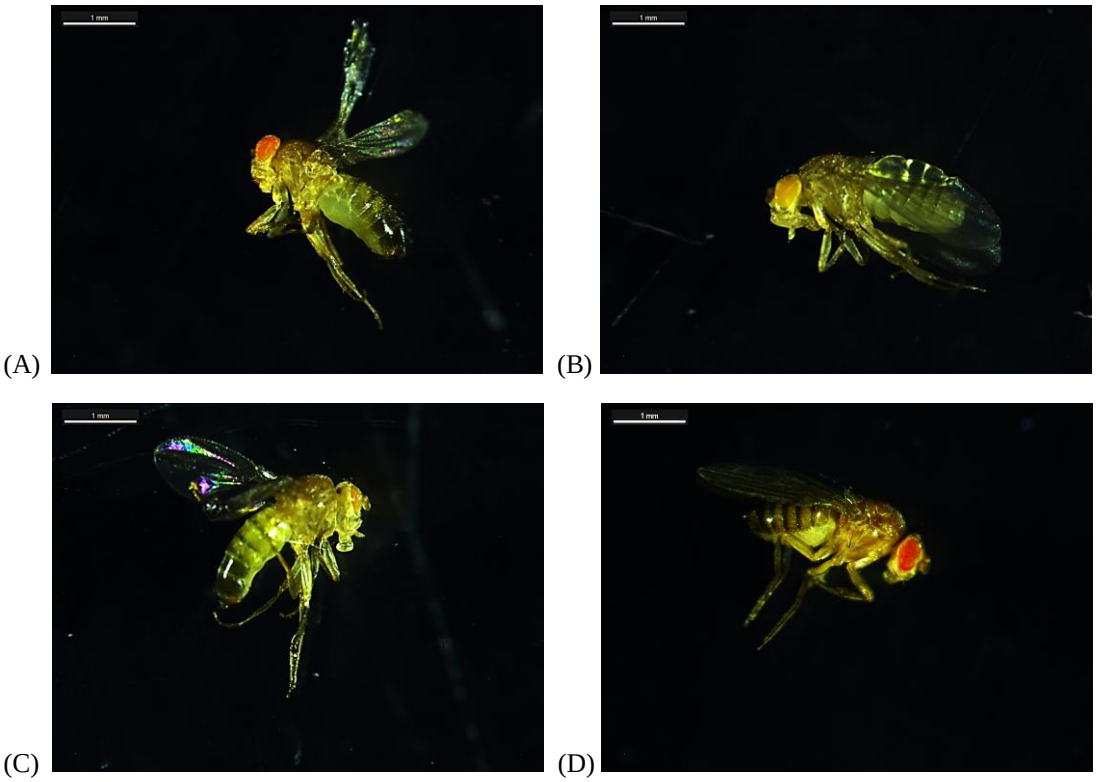


Figure 28: The different phenotypes and genotypes of flies expressing CYP4G1Δinsertion. (A) Alive male with non-Bar eyes and genotype: $\frac{CYP4G1\Delta 4}{\rightarrow}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$. (B) Alive female with non-Bar eyes and genotype: $\frac{CYP4G1\Delta 4}{\text{w}}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$. (C) Alive male with Bar eyes and genotype: $\frac{FM7}{\rightarrow}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$. (D) Alive female with heart-shaped eyes and genotype: $\frac{w}{\text{FM7}}$; $\frac{\text{REGAL4}}{+}$; $\frac{\text{UAS}-\text{CYP4G1}\Delta\text{insertion, w+}}{+}$.

3.4 Bacterial expression of CYP4G16 and CYP4G17

As described in sections 2.2.3 and 2.2.4.2, pGem:CYP4G16 and pGem:CYP4G17, both in wt or N-terminally truncated forms, were sequenced (Supplemental Figure 3) and the pCW:p450 and pCDF.Duet-1:p450_CPR were constructed. All constructs are depicted in Figure 29.

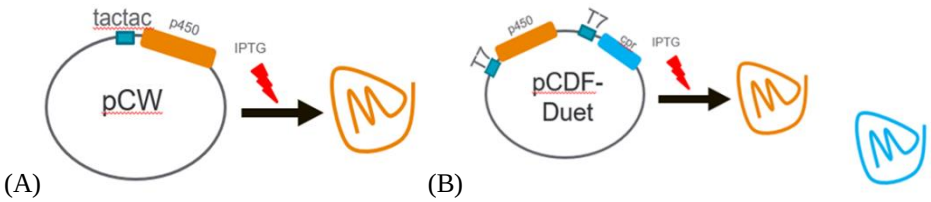


Figure 29: The pCW:P450 (A) and pCDF.Duet-1:P450_CPR (B) plasmids (where P450 declares either situation of CYP4G16 or CYP4G17; wt or N-terminally truncated) are presented in designed illustrations (see also sections 2.2.3 and 2.2.4.2). Created by Vasileia Balabanidou.

Five different bacterial strains were tested for their ability to express CYP4Gs under standard expression conditions, in the presence or not of chaperones (section 2.10, 1st attempt, Table 5). Under these experimental conditions (shown in Supplemental Figure 4), wt and N-terminally truncated CYP4G16 and CYP4G17 were not successfully expressed; at the same time another P450 protein, CYP6CM1 (constructed by Dimitra Tsakireli[59]), was successfully expressed (Figure 30). The calculated molecular weight of CYP4G16/17, CYP6CM1 and ΔNCYP4G16/17 is 63 kDa, 60 kDa and 58 kDa, respectively.

Table 5: The different combinations of plasmids and bacterial strains tested in this work are depicted with black (1st attempt) and red (2nd attempt) x, respectively.

E. coli /tested sequences	DH5a	JM09	Lemo21 (DE3)	BL21 (DE3) Star	C43
pCW:CYP4G16wt	x	x	x x	x x	x
pCW:CYP4G16wt + chaperone plasmid 3		x			
pCW:CYP4G16wt + chaperone plasmid 5		x			
pCW:CYP4G16ΔN	x	x	x	x x	x
pCW:CYP4G16ΔNhis				x	
pCW:CYP4G17wt	x	x x	x x	x x	x x
pCW:CYP4G17wt + chaperone plasmid 3				x	
pCW:CYP4G17wt + chaperone plasmid 5				x	
pCW:CYP4G17ΔN	x	x x	x	x x	x
pCW:CYP4G17ΔNhis		x		x	
pCDF.Duet-1:CYP4G16wt_MdCPR	x	x	x		x
pCDF.Duet-1:CYP4G17wt_MdCPR		x	x		x
pCW:OmpA.CYP4G16_ΔNAgCPR	x	x	x x	x	x
pCW:OmpA.CYP4G17_ΔNAgCPR			x x	x x	x x

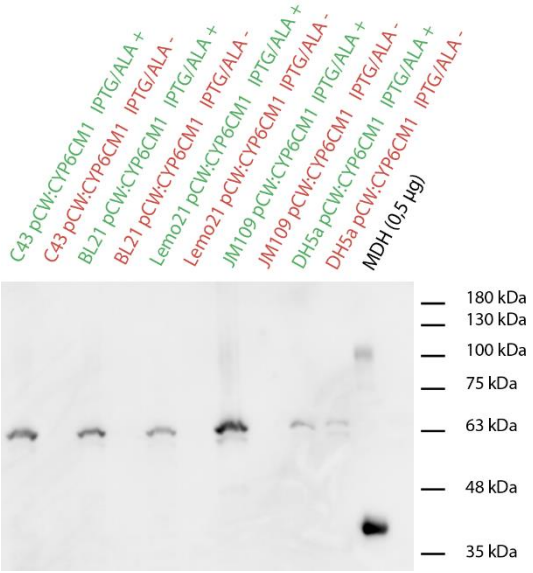


Figure 30: The expression patterns of CYP6CM1 among different bacterial strains were analyzed by Western blot using anti-his. Each lane corresponds to bacterial extracts before and after (4h) IPTG and ALA induction. As a positive control, 0.5 µg of purified His-MDH are loaded. All five bacterial strains (DH5a, JM09, Lemo21 (DE3), BL21 (DE3) Star and C43) were transformed with pCW:CYP6CM1.

To circumvent this, the culture conditions were changed (section 2.10.3, 2nd attempt and Table 5), again for all different combinations of constructs and bacterial cells. Western blot analysis of induced (+ IPTG) and non-induced (-IPTG) bacterial extracts using anti-CYP4G16 and anti-CYP4G17 revealed that both antibodies recognize a non-specific protein band that migrates at around 63 kDa (and a second band that migrates faster at around 35 kDa) since similar bands are recognized in the control cells non expressing CYP4Gs (Supplemental Figures 5-8). A representative example (JM109) is shown in Figure 31A, B. We performed negative immune-absorption of the anti-CYP4G17 against cytosolic and membrane control fractions of JM109 cells (section 2.10.4) and checked the resulting anti-CYP4G17 for immune-depletion. No band was detected, neither to the control nor to the CYP4G17 expressing samples (Figure 31C). Nevertheless, the expression of the his-tagged N-terminal deleted forms of both CYP4G16 and CYP4G17 revealed a specific band for the latter one, using the penta-his antibody (Figure 31D). In addition, most probably, the same band was detected with the absorbed anti-CYP4G17 antibody (Figure 31E), strongly indicating that the expression of the N-terminally truncated CYP4G17 was successful.

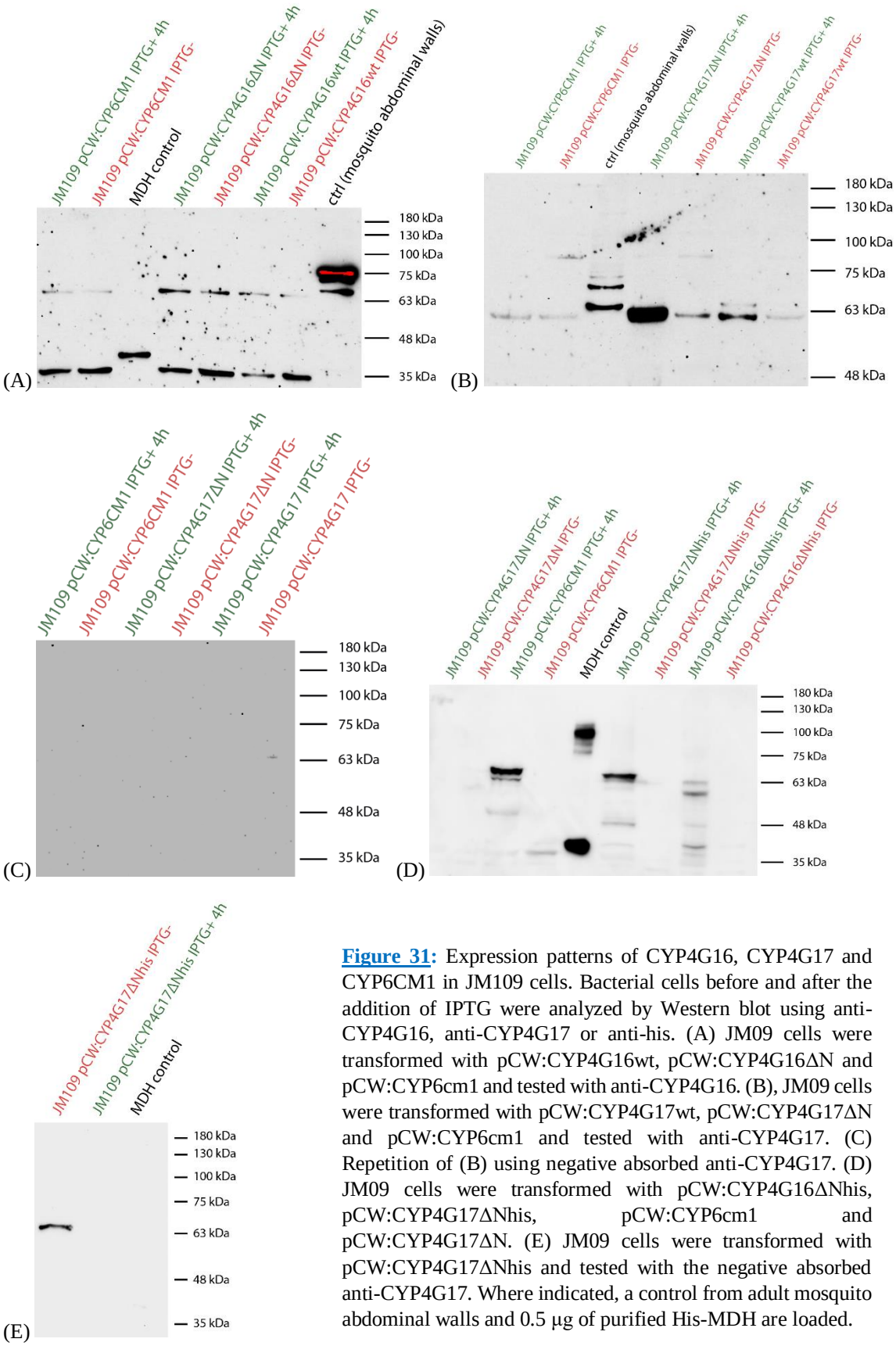


Figure 31: Expression patterns of CYP4G16, CYP4G17 and CYP6CM1 in JM109 cells. Bacterial cells before and after the addition of IPTG were analyzed by Western blot using anti-CYP4G16, anti-CYP4G17 or anti-his. (A) JM09 cells were transformed with pCW:CYP4G16wt, pCW:CYP4G16ΔN and pCW:CYP6cm1 and tested with anti-CYP4G16. (B), JM09 cells were transformed with pCW:CYP4G17wt, pCW:CYP4G17ΔN and pCW:CYP6cm1 and tested with anti-CYP4G17. (C) Repetition of (B) using negative absorbed anti-CYP4G17. (D) JM09 cells were transformed with pCW:CYP4G16ΔNhis, pCW:CYP4G17ΔNhis, pCW:CYP6cm1 and pCW:CYP4G17ΔN. (E) JM09 cells were transformed with pCW:CYP4G17ΔNhis and tested with the negative absorbed anti-CYP4G17. Where indicated, a control from adult mosquito abdominal walls and 0.5 µg of purified His-MDH are loaded.

4. Discussion

Insects are the most dominant species on earth, originating from 480 million years ago[10]. For this dominance, a number of structural, morphological and physiological factors are responsible: flight capacity, adaptability or universality, size, exoskeleton, resistance to desiccation, tracheal system of respiration, metamorphosis and reproductive potential[80]. In this work, focus is given on the exoskeleton, the insect cuticle, which coats all the tissues exposed to the external environment (body, fore-/hind-gut, luminal side of the tracheae) and serves multiple roles in protection against dehydration, predators, pathogen entry and locomotion[30]. More specifically, the insect-specific CYP4G enzymes, mainly responsible for hydrocarbon biosynthesis in insects, could be defined as a beneficial attribute[3,9], because their participation in cuticle formation counteracts insects significant surface-to-volume ratio and, hence, their inclination to lose water through evaporation[22].

CYP4Gs are members of the P450 subfamily that have orthologues distributed across the Insecta[9] and are characterized by a unique +44 residue insertion[3]. The majority of insects have approximately two CYP4Gs. These homolog pairs, however, seem to be not fully redundant[3]; CYP4G1 and CYP4G15 of *Drosophila melanogaster* (fruit-fly) have different localizations (oenocytes[9] vs brain[29]) and catalytic activities (examined in this work), CYP4G55 and CYP4G56 of *Dendroctonus ponderosae* (mountain pine beetle) marginally diverge in velocities[3,24], CYP4G76 and CYP4G115 of *Nilaparvata lugens* (brown planthopper) have small differences in their impact on insecticide penetration[3,21], CYP4G16 and CYP4G17 of *Anopheles gambiae* (mosquito) appear to differ in efficacy and slightly in substrate specificity (CYP4G17 displays preference for longer CHCs)[3,26] and CYP4G106 and CYP4G107 of *Rhodnius prolixus* (triatomine) have dissimilar specificities (molecular docking analysis estimates a preference for straight chain CHCs for CYP4G106 and methyl-branched chains for CYP4G107) and their differential contribution to insect desiccation (CYP4G107 has shown elevated relevance to desiccation stress tolerance)[22]. Therefore, there are many more to be discovered about the function of CYP4G enzymes[9].

Our goal was to create transgenic flies that express a deleted form of the insertion of the *Cyp4g1* gene and at first to check their viability and then elucidate the contribution of the insertion to the enzyme function. On one hand, the flies expressing the ablated CYP4G1 under the control of a UAS-REGal4 system at a null-CYP4G1 background had no differences in comparison with the wild-type flies, as far as successful pupation, eclosion or viability were concerned (section 3.3). On the other hand, the lines that were created by the CRISPR-Cas9 genome editing technique, had a systemic deleted form of the insertion that was a few amino-acids longer and could not reach adulthood at the hemizygous state, resulting in flies surviving up to the pharate adult state (section 3.2). This result can be attributed to two explanations: a CRISPR-Cas9-mediated deletion of CYP4G1 specific insertion has more severe effects than the “gentler” UAS/Gal4 system used to introduce the deletion in the CYP4G1 protein null mutants or/and the 6 amino-acids difference between the two cases is presumably crucial (Figure 25). Nonetheless, since there are survivors in the “milder” UAS-REGal4 system, the modified protein can be

considered correctly folded and functional, yet different potential defects of the flies bearing it could be present.

In order to verify if this case is true, a precise estimation of the survivors' lifetime will take place. In addition, CHC analysis of the survivors will shed light on the potential differences between the CHC profiles generated by the wild-type and the ablated form of CYP4G1. Moreover, as discussed in section 1.2.2, the insertion protrudes from the globular structure of the enzyme and is enriched in acidic amino acids, suggesting potential known and unknown interactions[3]. That is why an anti-CYP4G1 antibody has already been ordered and will be used for western blot analysis and cross-linking experiments with subsequent mass-spectrometry analyses, in order to initially detect the different forms of CYP4G1 in the aforementioned transgenic flies (CYP4G1wt and CYP4G1Δinsertion) and then identify their possible interactors, respectively. In this way, we could define if the loop mediates protein interactions and which they are, shedding light on different aspects of its function. Lastly, *in vitro* biochemical analysis of the different CYP4G1 forms, if successful, could further clarify this, by possibly distinguishing disparities in substrate preference, catalytic velocity, efficacy etc.

Secondly, apart from the CYP4G1 that has a key role in CHC biosynthesis and is expressed in oenocytes, CYP4G15 is also present in *Drosophila*; it is expressed in the brain[3,29], where its function is unknown, while *Cyp4g15*-RNAi flies are viable[3,16]. In order to further investigate CYP4G15, *D. melanogaster* individuals that ectopically express CYP4G15 in *Cyp4g1*-KD oenocytes were created. Our results revealed that two copies of CY4G15 could partially rescue the KD phenotype, suggesting that CYP4G15 can functionally substitute CYP4G1 at a degree (Figure 17). Apart from the survivors, a noteworthy amount of deceased, early-emerged adults were found lying on the food, distinct from the ones arrested during eclosion. This also occurred to flies expressing one copy of CYP4G15 (Figure 17); the null level of adult survival combined with the significant adult mortality entail that even a minute expression of CYP4G15 leads to ameliorated eclosion capability. In other words, CYP4G15 acts as a decarbonylase in the oenocytes.

The overall functionality of CYP4G15 seems to play a greater role than the gene copies number factor as far as the survival ability is concerned. However, this dose dependency is in proportion to the potentially reduced level of RE driver activation, the significant expression levels of native CYP4Gs[3,11,16,25], the, as of today, sluggish enzyme activity *in vitro* (in *Sf9* cells)[9,23,25] and has been observed in similar studies for one copy of CYP4G16/17 in transgenic flies[26]. It is important to point here that *Cyp4g1* is the most highly expressed P450 gene in *Drosophila melanogaster*[11], meaning that it is required in very high levels. The silencing of *Cyp4g1* and its replacement with *Cyp4g15* under the UAS-REGal4 system therefore may not be sufficient for a totally healthy phenotype as also dose dependency indicates. Maybe, the expression of *Cyp4g15* under the regulation of *Cyp4g1* promoter would be a more comparable way of studying CYP4G1 substitution in the oenocytes.

The probability of newly emerged adults to die shortly after eclosion was the reason why the adult survivors (females and males) were calculated three days post adulthood. Interestingly, the strong majority of rescued flies are males (more than 90%). This is in agreement with a preliminary study where increased desiccation resistance of females from temperate and tropical regions is demonstrated[81]. Laboratory flies have not be tested[81], and it is plausible that the females of our investigated strains may entail additional CHC load for desiccation resistance, making it harder to counterbalance a deficiency, especially in modified genetic background. In addition, the sexual dimorphism itself and the divergent sexual and natural selection between the sexes should also be taken into account[82–84]. Furthermore, potential deviations of the two genders in expression levels or spatiotemporal profile of the RE-Gal4 driver[73] might impact the efficiency and/or specificity of *Cyp4g1* KD and/or CYP4G15 expression.

Considering that CYP4G15 acts as a functional decarbonylase, the cuticular hydrocarbon profile of the survivors will be investigated. Nevertheless, a diminished amount of total hydrocarbons and differences in the relative abundance of the hydrocarbon blend in flies expressing CYP4G15 opposed to the wild-type is expected, similar to previous studies[26]. An anti-CYP4G15 antibody has already been ordered and will be used to identify the exact localization (cell type) of the enzyme in the CNS of *D. melanogaster* (sub-cellularly more precise than the RNA *in situ* hybridization[29]). Moreover, precise estimation of the survivors' lifetime and an *in vitro* biochemical analysis, if successful, could further assist the comprehension of the CYP4G15 function and elucidate the differences between the two *Drosophila* CYP4Gs.

Lastly, the substrate selectivity and the kinetic parameters of insect P450s are hard to quantitatively describe[1]. Biochemical analysis *in vitro*, as discussed in section 1.4.1, demands successful recombinant P450 expression. Towards this direction, we tried to develop a robust methodology by: balancing P450 expression (three vectors: pCW, pCDF.Duet-1 and pCW:OmpA.CYP4G16/17_ΔNAgCPR), exploring N-terminal modifications (native P450, ΔN-P450 and OmpA-P450), examining of the reductase partner (*Md*CPR cloned in pCDF.Duet-1), fine-tuning *E. coli* for membrane protein expression (test five different bacterial strains: BL21 (De3) star, JM109, DH5a, C43 (DE3) and Lemo21 DE3) and optimizing culture conditions (temperature and incubation time). In our experiments, almost all combinations of plasmids and bacterial strains showed unsuccessful expression patterns (Figure 31A-C and Supplemental Figures 4-8). Only the N-terminally truncated CYP4G17 was expressed in JM109 cells, as evidenced by the detection of a specific band by both antibodies (anti-His and negatively absorbed anti-CYP4G17) at the correct size (Figure 31D, E). New parameters should be tested again in small scale in all five bacterial strains. Furthermore, the expressed proteins will be examined for correct folding via CO differential spectrum recordings. It would be the first time that a CYP4G enzyme to be efficiently expressed in *E. coli* and could assist with the delineation of substrates specificity, kinetic parameters etc.

In other words, it is of great importance to elucidate CYP4Gs function and specificity, particularly in insects that express multiple *CYP4G* genes. Hopefully, this work will assist towards this direction.

Glossary

(∗1)

The constituent lipid molecules of the cuticle have a hydrophobic nature that requires a specialized system to transport them through the hemolymph[39], which surrounds all tissues and organs, having roles in the transfer of nutrient substrates and metabolic wastes to sites of usage and excretion, respectively[39]. This transport is carried through lipoprotein vehicles, called lipophorins[36,41]. Lipophorins are synthesized and secreted by the fat body[40,41], which serves analogous roles to the vertebrate adipose tissue and liver, like being the principal metabolic center and the major lipid storage organ[36,39]. They are released into the hemolymph providing and convey numerous lipids (hydrocarbons, cholesterol, diacylglycerol and carotenoids) from sites of storage, absorption, or synthesis to sites of exploitation as precursors for synthesis, metabolic fuel or structural components of cell membrane and cuticle[36,40,85–89]. These utilization locations are targeted through the lipophorin receptors[41–44].

(∗2)

Isopropyl β-d-1-thiogalactopyranoside (IPTG) is a molecular biology reagent that mimics allolactose, a lactose metabolite that triggers transcription of the lac operon. IPTG is used to induce protein expression where the gene is under the control of the lac operator[90].

References

1. Feyereisen, R. Insect CYP Genes and P450 Enzymes. in *Insect Molecular Biology and Biochemistry* 236–316 (Elsevier, 2012). doi:10.1016/B978-0-12-384747-8.10008-X.

2. Nelson, D. R. & Nebert, D. W. Cytochrome P450 (*CYP*) Gene Superfamily. in *eLS* (ed. John Wiley & Sons Ltd) 1–19 (John Wiley & Sons, Ltd, 2018). doi:10.1002/9780470015902.a0005667.pub3.

3. Feyereisen, R. Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. *Mol. Phylogenet. Evol.* **143**, 106695 (2020).

4. OMURA T, S. R. The Carbon Monoxide-binding Pigment of Liver Microsomes. *J Biol Chem.* **239**, 2370–2378 (1964).

5. Mansuy, D. The great diversity of reactions catalyzed by cytochromes P450. *Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol.* **121**, 5–14 (1998).

6. Guengerich, F. P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **14**, 611–650 (2001).

7. Willingham, A. T. & Keil, T. A tissue specific cytochrome P450 required for the structure and function of Drosophila sensory organs. *Mech. Dev.* **121**, 1289–1297 (2004).

8. Blomquist, G. J., Dillwith, J. W. & George Pomonis, J. Sex pheromone of the housefly. *Insect Biochem.* **14**, 279–284 (1984).

9. Qiu, Y. *et al.* An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc. Natl. Acad. Sci.* **109**, 14858–14863 (2012).

10. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).

11. Daborn, P. J. A Single P450 Allele Associated with Insecticide Resistance in Drosophila. *Science* **297**, 2253–2256 (2002).

12. Carugo, O. Amino acid composition and protein dimension. *Protein Sci.* **17**, 2187–2191 (2008).

13. Howard, R. W. & Blomquist, G. J. ECOLOGICAL, BEHAVIORAL, AND BIOCHEMICAL ASPECTS OF INSECT HYDROCARBONS. *Annu. Rev. Entomol.* **50**, 371–393 (2005).

14. Ferveur, J.-F. Cuticular Hydrocarbons: Their Evolution and Roles in Drosophila Pheromonal Communication. *Behav. Genet.* **35**, 279–295 (2005).

15. Qian, J.-L. *et al.* Identification of cytochrome P450, odorant-binding protein, and chemosensory protein genes involved in Type II sex pheromone biosynthesis and transportation in the tea pest, Scopula subpunctaria. *Pestic. Biochem. Physiol.* **169**, 104650 (2020).

16. Chung, H. *et al.* Characterization of Drosophila melanogaster cytochrome P450 genes. *Proc. Natl. Acad. Sci.* **106**, 5731–5736 (2009).

17. Yu, Z. *et al.* LmCYP4G102: An oenocyte-specific cytochrome P450 gene required for cuticular waterproofing in the migratory locust, Locusta migratoria. *Sci. Rep.* **6**, 29980 (2016).

18. Chen, N. *et al.* Cytochrome P450 gene, CYP4G51, modulates hydrocarbon production in the pea aphid, Acyrthosiphon pisum. *Insect Biochem. Mol. Biol.* **76**, 84–94 (2016).

19. Chen, N., Pei, X., Li, S., Fan, Y. & Liu, T. Involvement of integument-rich *CYP4G19* in hydrocarbon biosynthesis and cuticular penetration resistance in *Blattella germanica* (L.). *Pest Manag. Sci.* **76**, 215–226 (2020).

20. Wang, S. Y., Hackney Price, J. & Zhang, D. Hydrocarbons catalysed by *TmCYP4G122* and *TmCYP4G123* in *Tenebrio molitor* modulate the olfactory response of the parasitoid *Scleroderma guani*. *Insect Mol. Biol.* **28**, 637–648 (2019).

21. Wang, S., Li, B. & Zhang, D. NlCYP4G76 and NlCYP4G115 Modulate Susceptibility to Desiccation and Insecticide Penetration Through Affecting Cuticular Hydrocarbon Biosynthesis in Nilaparvata lugens (Hemiptera: Delphacidae). *Front. Physiol.* **10**, 913 (2019).

22. Dulbecco A. B. et al. Deciphering the role of Rhodnius prolixusCYP4G genes in straight and methyl-branched hydrocarbon formation and in desiccation tolerance. *Insect Mol. Biol.* (2020).

23. Calla, B. *et al.* Functional characterization of CYP4G11-a highly conserved enzyme in the western honey bee *Apis mellifera*. *Insect Mol. Biol.* **27**, 661–674 (2018).

24. MacLean, M., Nadeau, J., Gurnea, T., Tittiger, C. & Blomquist, G. J. Mountain pine beetle (Dendroctonus ponderosae) CYP4Gs convert long and short chain alcohols and aldehydes to hydrocarbons. *Insect Biochem. Mol. Biol.* **102**, 11–20 (2018).

25. Balabanidou, V. *et al.* Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in *Anopheles gambiae*. *Proc. Natl. Acad. Sci.* **113**, 9268–9273 (2016).

26. Kefi, M. *et al.* Two functionally distinct CYP4G genes of Anopheles gambiae contribute to cuticular hydrocarbon biosynthesis. *Insect Biochem. Mol. Biol.* **110**, 52–59 (2019).

27. Chung, H. & Carroll, S. B. Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. *BioEssays* **37**, 822–830 (2015).

28. Gutierrez, E., Wiggins, D., Fielding, B. & Gould, A. P. Specialized hepatocyte-like cells regulate Drosophila lipid metabolism. *Nature* **445**, 275–280 (2007).

29. Maïbèche-Coisne, M., Monti-Dedieu, L., Aragon, S. & Dauphin-Villemant, C. A New Cytochrome P450 from Drosophila melanogaster, CYP4G15, Expressed in the Nervous System. *Biochem. Biophys. Res. Commun.* **273**, 1132–1137 (2000).

30. Moussian, B. Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochem. Mol. Biol.* **40**, 363–375 (2010).

31. Gibbs, A. G. Water-Proofing Properties of Cuticular Lipids. *Am. Zool.* **38**, 471–482 (1998).

32. Gibbs, A. G. Lipid melting and cuticular permeability: new insights into an old problem. *J. Insect Physiol.* **48**, 391–400 (2002).

33. Wigglesworth, V. B. TRANSPIRATION THROUGH THE CUTICLE OF INSECTS. 19.

34. Gould AP et. al. Insect oenocytes: a model system for studying cell-fate specification by Hox genes. *J Anat.* **199(Pt 1–2)**, 25–33 (2001).

35. MARTINS, G F; RAMALHO-ORTIGÃO, J M. Oenocytes in insects. *Invertebrate Survival Journal* **9**, 139–152 (2012).

36. Makki, R., Cinnamon, E. & Gould, A. P. The Development and Functions of Oenocytes. *Annu. Rev. Entomol.* **59**, 405–425 (2014).

37. Lynd, A. *et al. Development of a functional genetic tool for Anopheles gambiae oenocyte characterisation: application to cuticular hydrocarbon synthesis*. http://biorxiv.org/lookup/doi/10.1101/742619 (2019) doi:10.1101/742619.

38. Jackson, A. & Locke, M. The formation of plasma membrane reticular systems in the oenocytes of an insect. *Tissue Cell* **21**, 463–473 (1989).

39. Gilbert LI, Chino H. Transport of lipids in insects. J Lipid Res. *J Lipid Res.* **15(5)**, 439–456 (1974).

40. Chino, H. Lipid Transport by Hemolymph Lipoprotein. in *Energy Metabolism in Insects* (ed. Downer, R. G. H.) 155–167 (Springer US, 1981). doi:10.1007/978-1-4615-9221-1_6.

41. Blacklock, B. J. & Ryan, R. O. Hemolymph lipid transport. *Insect Biochem. Mol. Biol.* **24**, 855–873 (1994).

42. Herz, J., Chen, Y., Masiulis, I. & Zhou, L. Expanding functions of lipoprotein receptors. *J. Lipid Res.* **50**, S287–S292 (2009).

43. Van der Horst, D. J., Roosendaal, S. D. & Rodenburg, K. W. Circulatory lipid transport: lipoprotein assembly and function from an evolutionary perspective. *Mol. Cell. Biochem.* **326**, 105–119 (2009).

44. Lu, K., Chen, X., Li, Y., Li, W. & Zhou, Q. Lipophorin receptor regulates Nilaparvata lugens fecundity by promoting lipid accumulation and vitellogenin biosynthesis. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **219–220**, 28–37 (2018).

45. Wicker-Thomas, C. *et al.* Flexible origin of hydrocarbon/pheromone precursors in *Drosophila melanogaster*. *J. Lipid Res.* **56**, 2094–2101 (2015).

46. Reed, J. R., Hernandez, P., Blomquist, G. J., Feyereisen, R. & Reitz, R. C. Hydrocarbon biosynthesis in the house fly, Musca domestica: Substrate specificity and cofactor requirement of P450hyd. *Insect Biochem. Mol. Biol.* **26**, 267–276 (1996).

47. Lycett, G. J. *et al.* Anopheles gambiae P450 reductase is highly expressed in oenocytes and in vivo knockdown increases permethrin susceptibility. *Insect Mol. Biol.* **15**, 321–327 (2006).

48. Murataliev, M. B., Feyereisen, R. & Walker, F. A. Electron transfer by diflavin reductases. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1698**, 1–26 (2004).

49. Balabanidou, V., Grigoraki, L. & Vontas, J. Insect cuticle: a critical determinant of insecticide resistance. *Curr. Opin. Insect Sci.* **27**, 68–74 (2018).

50. Guo, G., Geng, Y., Huang, D., Xue, C. & Zhang, R. Level of CYP4G19 Expression Is Associated with Pyrethroid Resistance in *Blattella germanica*. *J. Parasitol. Res.* **2010**, 1–7 (2010).

51. Kalsi, M. & Palli, S. R. Cap n collar transcription factor regulates multiple genes coding for proteins involved in insecticide detoxification in the red flour beetle, Tribolium castaneum. *Insect Biochem. Mol. Biol.* **90**, 43–52 (2017).

52. Pridgeon, J. W., Zhang, L. & Liu, N. Overexpression of CYP4G19 associated with a pyrethroid-resistant strain of the German cockroach, Blattella germanica (L.). *Gene* **314**, 157–163 (2003).

53. Jones, C. M. *et al.* The dynamics of pyrethroid resistance in Anopheles arabiensis from Zanzibar and an assessment of the underlying genetic basis. *Parasit. Vectors* **6**, 343 (2013).

54. Pittendrigh, B. Cytochrome P450 genes from Helicoverpa armigera: Expression in a pyrethroid-susceptible and -resistant strain. *Insect Biochem. Mol. Biol.* **27**, 507–512 (1997).

55. Højland, D. H. & Kristensen, M. Analysis of Differentially Expressed Genes Related to Resistance in Spinosad- and Neonicotinoid-Resistant Musca domestica L. (Diptera: Muscidae) Strains. *PLOS ONE* **12**, e0170935 (2017).

56. Martínez-Paz, P., Morales, M., Martínez-Guitarte, J. L. & Morcillo, G. Characterization of a cytochrome P450 gene (CYP4G) and modulation under different exposures to xenobiotics (tributyltin, nonylphenol, bisphenol A) in Chironomus riparius aquatic larvae. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **155**, 333–343 (2012).

57. Balabanidou, V. *et al.* 1 Mosquito build leg cloak to fight against insecticides. 41.

58. Hausjell, J., Halbwirth, H. & Spadiut, O. Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Biosci. Rep.* **38**, BSR20171290 (2018).

59. Tsakireli, D., Riga, M., Kounadi, S., Douris, V. & Vontas, J. Functional characterization of CYP6A51, a cytochrome P450 associated with pyrethroid resistance in the Mediterranean fruit fly Ceratitis capitata. *Pestic. Biochem. Physiol.* **157**, 196–203 (2019).

60. Riga, M. *et al.* Abamectin is metabolized by CYP392A16, a cytochrome P450 associated with high levels of acaricide resistance in Tetranychus urticae. *Insect Biochem. Mol. Biol.* **46**, 43–53 (2014).

61. Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the Fly: A Primer on the *Drosophila* Model System. *Genetics* **201**, 815–842 (2015).

62. Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell* **58**, 568–574 (2015).

63. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096–1258096 (2014).

64. Wiedenheft, B. *et al.* Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).

65. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

66. Bassett, A. & Liu, J.-L. CRISPR/Cas9 mediated genome engineering in Drosophila. *Methods* **69**, 128–136 (2014).

67. Beumer, K. J., Trautman, J. K., Mukherjee, K. & Carroll, D. Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. *G3amp58 GenesGenomesGenetics* **3**, 657–664 (2013).

68. Beumer, K. J. & Carroll, D. Targeted genome engineering techniques in Drosophila. *Methods* **68**, 29–37 (2014).

69. Barolo, S., Carver, L. A. & Posakony, J. W. GFP and β-Galactosidase Transformation Vectors for Promoter/Enhancer Analysis in *Drosophila*. *BioTechniques* **29**, 726–732 (2000).

70. Piwko, P., Vitsaki, I., Livadaras, I. & Delidakis, C. The Role of Insulators in Transgene Transvection in *Drosophila*. *Genetics* **212**, 489–508 (2019).

71. Ishida, H., Garcia-Herrero, A. & Vogel, H. J. The periplasmic domain of Escherichia coli outer membrane protein A can undergo a localized temperature dependent structural transition. *Biochim. Biophys. Acta BBA - Biomembr.* **1838**, 3014–3024 (2014).

72. Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage φC3. 8.

73. Bousquet, F. *et al.* Expression of a desaturase gene, desat1, in neural and nonneural tissues separately affects perception and emission of sex pheromones in Drosophila. *Proc. Natl. Acad. Sci.* **109**, 249–254 (2012).

74. Gratz, S. J. *et al.* Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* **194**, 1029–1035 (2013).

75. Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc. Natl. Acad. Sci.* **111**, E2967–E2976 (2014).

76. Dahanukar, A., Walker, J. A. & Wharton, R. P. Smaug, a Novel RNA-Binding Protein that Operates a Translational Switch in Drosophila. *Mol. Cell* **4**, 209–218 (1999).

77. Ingham, V. A. *et al.* Dissecting the organ specificity of insecticide resistance candidate genes in Anopheles gambiae: known and novel candidate genes. *BMC Genomics* **15**, 1018 (2014).

78. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, inEscherichia coli. *Appl. Environ. Microbiol.* **64**, 1694–1699 (1998).

79. Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in Escherichia coli. *Appl. Environ. Microbiol.* **66**, 884–889 (2000).

80. https://preparmy.com/life-science/biology/agriculture/contributory-factors-abundance-insects/, http://ecoursesonline.iasri.res.in/mod/page/view.php?id=101147, http://eagri.org/eagri50/ENTO231/lec02.pdf,https://biocontrol.entomology.cornell.edu/bio.php, https://projects.ncsu.edu/cals/course/ent425/text01/success.html

81. Parkash, R. & Ranga, P. Sex-specific divergence for adaptations to dehydration stress in Drosophila kikkawai. *J. Exp. Biol.* **216**, 3301–3313 (2013).

82. Foley, B., Chenoweth, S. F., Nuzhdin, S. V. & Blows, M. W. Natural Genetic Variation in Cuticular Hydrocarbon Expression in Male and Female *Drosophila melanogaster*. *Genetics* **175**, 1465–1477 (2007).

83. Sharma, M. D., Mitchell, C., Hunt, J., Tregenza, T. & Hosken, D. J. The Genetics of Cuticular Hydrocarbon Profiles in the Fruit Fly Drosophila simulans. *J. Hered.* **103**, 230–239 (2012).

84. Curtis, S. *et al.* Epicuticular Compounds of Drosophila subquinaria and D. recens: Identification, Quantification, and Their Role in Female Mate Choice. *J. Chem. Ecol.* **39**, 579–590 (2013).

85. Diehl, P. A. Synthesis and release of hydrocarbons by the oenocytes of the desert locust, Schistocerca gregaria. *J. Insect Physiol.* **21**, 1237–1246 (1975).

86. Haruhito, K. & Haruo, C. Transport of hydrocarbons by the lipophorin of insect hemolymph. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* **710**, 341–348 (1982).

87. Katase, H. & Chino, H. Transport of hydrocarbons by haemolymph lipophorin in Locusta migratoria. *Insect Biochem.* **14**, 1–6 (1984).

88. Pho DB, Pennanec'h M, Jallon JM. Purification of adult Drosophila melanogaster lipophorin and its role in hydrocarbon transport. 15.

89. Fan Y, Zurek L, Dykstra MJ, Schal C. Hydrocarbon synthesis by enzymatically dissociated oenocytes of the abdominal integument of the German Cockroach, Blattella germanica. **90**, 121–126 (2003).

90. https://en.wikipedia.org/wiki/Isopropyl_%CE%B2-D-1-thiogalactopyranoside

91. Samantsidis, G.-R., O'Reilly, A. O., Douris, V. & Vontas, J. Functional validation of target-site resistance mutations against sodium channel blocker insecticides (SCBIs) via molecular modeling and genome engineering in Drosophila. *Insect Biochem. Mol. Biol.* **104**, 73–81 (2019).

Supplementary information

Gene of interest	Primer name	Primer sequence
Universal	T7	TAATACGACTCACTATAGGG
	SP6	TATTTAGGTGACACTATAG
CYP4G15	BssHII 4g15 F	GCGCGCACCATGGAGGTGCTGAAGAAGGAC
	Xhoi 4g15 R	CTCGAGTCAACTGGTCCGCGGCTG
	CYP4G15F	GAGTTCGTGTTCACCCTGACG
	CYP4G15R	GAGTTCTCGTTGAACAACTCGA
	CYP4G15_qF1	CGAGTTCGTGTTCACCCTGA
	CYP4G15_qR1	TTCTTGCTGCGGATCACCTT
	CYP4G15_qF2	CTCGCGTAATGTGGTGAGGA
	CYP4G15_qR2	GCGTTTTCTTCGACACACCC
CYP4G1 (deletion)	4G1 START F	GCGCGCACCATGGCAGTGGAAGTAGTTCAGGA
	4G1 R seq	ACTACCGCCTTTGAGTGATCC
	4G1 F seq	TATCCACCATCGTGAGGAACTA
	4G1 END R	ATCGATCTCGAGGTTTTGCAGCTCTATCGGCTAC
	4G1 Fq	CGACACGGAGGCAGATTTCA
	4G1Rq	ACGCTTCTCCAACGAGACAT
	4G1DscrR	GGTATCGTGGCCCTCAAACA
	4g1Dscr2F	CATCCTGTTGTCTACCGCCA
	4G1Dscr2R	GAGGGCGAAACTAGATCCCG
	ΔCYP4G1seq	AGCAAGGTGGTCAAGGATCG
CYP4G1 (CRISPR)	Sense-Start	Phospho-CTTCGGTTAAATTACTGTACCGCC
	Antisense-Start	Phospho-AAACGGCGGTACAGTAATTTAACC
	Sense-End	Phospho-CTTCGGCCACGATACCACCTCGGC
	Antisense-End	Phospho-AAACGCCGAGGTGGTATCGTGGCC
	Start F	GCGCGCACCATGGCAGTGGAAGTAGTTCAGGA
	END R	ATCGATCTCGAGGTTTTGCAGCTCTATCGGCTA
	Delta Cyp4g1 seq	AGCAAGGTGGTCAAGGATCG
	4G1Rq	ACGCTTCTCCAACGAGACAT
CYP4G16	4G16 F	CATATGTCAGCAACAATTGCGCATAC
	4G16 R	GTCGACTCATAATGTCTTCGATTTGCGTTGAC
	4G16FQold	GGCAGCCTTTGATACCGGTACGCG
	4G16RQold	GGCCTTCCACTGTGTTTGTCTTGG
	ΔN4G16 F	CATATGGCGCGTTTATCACGACGCCACATG
	his4G16 R	GTCGACTCAATGATGATGATGATGATGTAATGTCTTCGATTT GCGTTGAC
CYP4G17	4G17 F	CATATGGGCATTGAAACGATCCCG
	4G17 R	TATAGTCGACTCATGCCCTCGGCTCCA
	4g17FQold	CACCTGCGCCTGGACCCAC

	4G17RQold	CAGCGACGGGGACGGCATC
	ΔN4G17 F	CATATGGCGCAAACGCGCCGGTACGT
	his4G17 R	GTCGACTCAATGATGATGATGATGATGTGCCCTCGGCTCCA
RPL11	RPL11_Fq	CGATCCCTCCATCGGTATCT
	RPL11_Rq	AACCACTTCATGGCATCCTC

Supplemental Table 1: Sequences of primers used for DNA amplification and/or sequencing.

Supplemental Figure 1: Sequencing results of pCW:CYP4G15.

```
CYP4G15wt        ATGGAGGTGCTGAAGAAGGACGCCGCCCTGGGCTCACCCAGCAGCGTCTTCTACTTCCTG  60
pGem:CYP4G15     ATGGAGGTGCTGAAGAAGGACGCCGCCCTGGGCTCACCCAGCAGCGTCTTCTACTTCCTG  60
                 ************************************************************

CYP4G15wt        CTGCTGCCCACGCTGGTGCTTTGGTACATCTACTGGCGATTATCGCGGGCTCACCTGTAC  120
pGem:CYP4G15     CTGCTGCCCACGCTGGTGCTTTGGTACATCTACTGGCGATTATCGCGGGCTCACCTGTAC  120
                 ************************************************************

CYP4G15wt        AGGTTGGCCGGACGACTGCCGGGACCACGGGGACTGCCCATCGTGGGTCATCTGTTCGAT  180
pGem:CYP4G15     AGGTTGGCCGGACGACTGCCGGGACCACGGGGACTGCCCATCGTGGGTCATCTGTTCGAT  180
                 ************************************************************

CYP4G15wt        GTGATTGGACCCGCTTCATCTGTTTTCCGAACAGTCATCCGCAAAAGCGCCCCTTTCGAG  240
pGem:CYP4G15     GTGATTGGACCCGCTTCATCTGTTTTCCGAACAGTCATCCGCAAAAGCGCCCCTTTCGAG  240
                 ************************************************************

CYP4G15wt        CACATAGCCAAGATGTGGATCGGGCCCAAGCTGGTGGTCTTCATCTACGATCCGCGGGAC  300
pGem:CYP4G15     CACATAGCCAAGATGTGGATCGGGCCCAAGCTGGTGGTCTTCATCTACGATCCGCGGGAC  300
                 ************************************************************

CYP4G15wt        GTGGAGTTGCTGCTGAGCAGCCATGTCTACATCGACAAGGCATCCGAGTACAAGTTCTTC  360
pGem:CYP4G15     GTGGAGTTGCTGCTGAGCAGCCATGTCTACATCGACAAGGCATCCGAGTACAAGTTCTTC  360
                 ************************************************************

CYP4G15wt        AAGCCCTGGCTGGGCGACGGACTGCTGATCAGCACAGGTCAAAAATGGCGATCGCACCGC  420
pGem:CYP4G15     AAGCCCTGGCTGGGCGACGGACTGCTGATCAGCACAGGTCAAAAATGGCGATCGCACCGC  420
                 ************************************************************

CYP4G15wt        AAACTGATTGCACCCACATTCCATCTGAATGTCTTGAAGAGCTTCATCGAGTTGTTCAAC  480
pGem:CYP4G15     AAACTGATTGCACCCACATTCCATCTGAATGTCTTGAAGAGCTTCATCGAGTTGTTCAAC  480
                 ************************************************************

CYP4G15wt        GAGAACTCGCGTAATGTGGTGAGGAAACTGCGTGCGGAGGATGGTCGCACTTTTGATTGC  540
pGem:CYP4G15     GAGAACTCGCGTAATGTGGTGAGGAAACTGCGTGCGGAGGATGGTCGCACTTTTGATTGC  540
                 ************************************************************

CYP4G15wt        CATGATTACATGAGCGAGGCAACCGTGGAGATTCTATTGGAGACTGCGATGGGTGTGTCG  600
pGem:CYP4G15     CATGATTACATGAGCGAGGCAACCGTGGAGATTCTATTGGAGACTGCGATGGGTGTGTCG  600
                 ************************************************************

CYP4G15wt        AAGAAAACGCAGGACAAATCGGGATTCGAGTATGCGATGGCCGTGATGCGGATGTGCGAC  660
pGem:CYP4G15     AAGAAAACGCAGGACAAATCGGGATTCGAGTATGCGATGGCCGTGATGCGGATGTGCGAC  660
                 ************************************************************

CYP4G15wt        ATCCTTCACGCCCGCCATCGGAGCATCTTTCTGCGCAACGAGTTCGTGTTCACCCTGACG  720
pGem:CYP4G15     ATCCTTCACGCCCGCCATCGGAGCATCTTTCTGCGCAACGAGTTCGTGTTCACCCTGACG  720
                 ************************************************************

CYP4G15wt        CGCTACTACAAGGAGCAGGGTCGCCTGCTGAACATCATCCACGGCCTGACCACCAAGGTG  780
pGem:CYP4G15     CGCTACTACAAGGAGCAGGGTCGCCTGCTGAACATCATCCACGGCCTGACCACCAAGGTG  780
                 ************************************************************

CYP4G15wt        ATCCGCAGCAAGAAGGCGGCCTTTGAGCAGGGCACCCGTGGATCCCTTGCCCAGTGCGAG  840
pGem:CYP4G15     ATCCGCAGCAAGAAGGCGGCCTTTGAGCAGGGCACCCGTGGATCCCTTGCCCAGTGCGAG  840
                 ************************************************************

CYP4G15wt        CTGAAGGCAGCCGCCTTAGAGCGGGAACGGGAACAGAATGGTGGTGTTGACCAGACCCCA  900
pGem:CYP4G15     CTGAAGGCAGCCGCCTTAGAGCGGGAACGGGAACAGAATGGTGGTGTTGACCAGACCCCA  900
                 ************************************************************

CYP4G15wt        TCAACAGCTGGCAGCGATGAGAAGGATCGGGAGAAGGATAAGGAGAAGGCCAGTCCGGTG  960
pGem:CYP4G15     TCAACAGCTGGCAGCGATGAGAAGGATCGGGAGAAGGATAAGGAGAAGGCCAGTCCGGTG  960
                 ************************************************************

CYP4G15wt        GCGGGACTGTCCTATGGACAGTCGGCTGGTCTCAAGGATGACCTGGATGTGGAGGACAAT  1020
pGem:CYP4G15     GCGGGACTGTCCTATGGACAGTCGGCTGGTCTCAAGGATGACCTGGATGTGGAGGACAAT  1020
                 ************************************************************

CYP4G15wt        GATATTGGTGAGAAGAAGCGGCTGGCCTTCCTTGATTTGATGCTCGAAAGTGCACAGAAT  1080
pGem:CYP4G15     GATATTGGTGAGAAGAAGCGGCTGGCCTTCCTTGATTTGATGCTCGAAAGTGCACAGAAT  1080
                 ************************************************************
```

```
CYP4G15wt        GGAGCACTCATTACGGACACGGAAATTAAGGAGCAGGTGGACACGATCATGTTCGAGGGA  1140
pGem:CYP4G15     GGAGCACTCATTACGGACACGGAAATTAAGGAGCAGGTGGACACGATCATGTTCGAGGGA  1140
                 ************************************************************

CYP4G15wt        CACGACACCACGGCAGCCGGCTCCTCGTTCTTTCTCTCGCTGATGGGCATCCACCAGGAC  1200
pGem:CYP4G15     CACGACACCACGGCAGCCGGCTCCTCGTTCTTTCTCTCGCTGATGGGCATCCACCAGGAC  1200
                 ************************************************************

CYP4G15wt        ATCCAGGATCGCGTACTGGCCGAACTCGACTCCATTTTTGGCGACTCCCAGCGACCGGCC  1260
pGem:CYP4G15     ATCCAGGATCGCGTACTGGCCGAACTCGACTCCATTTTTGGCGACTCCCAGCGACCGGCC  1260
                 ************************************************************

CYP4G15wt        ACATTTCAGGACACACTGGAGATGAAGTATCTGGAGCGGTGTCTGATGGAGACGCTGCGC  1320
pGem:CYP4G15     ACATTTCAGGACACACTGGAGATGAAGTATCTGGAGCGGTGTCTGATGGAGACGCTGCGC  1320
                 ************************************************************

CYP4G15wt        ATGTACCCACCCGTTCCATTGATCGCCCGCGAGCTGCAGGAGGATCTCAAGCTGAACTCG  1380
pGem:CYP4G15     ATGTACCCACCCGTTCCATTGATCGCCCGCGAGCTGCAGGAGGATCTCAAGCTGAACTCG  1380
                 ************************************************************

CYP4G15wt        GGCAACTACGTGATTCCCAGGGGCGCCACGGTGACGGTGGCCACCGTCCTGCTCCATCGC  1440
pGem:CYP4G15     GGCAACTACGTGATTCCCAGGGGCGCCACGGTGACGGTGGCCACCGTCCTGCTCCATCGC  1440
                 ************************************************************

CYP4G15wt        AATCCGAAGGTCTACGCCAATCCCAACGTCTTCGATCCGGACAACTTCCTGCCGGAGCGC  1500
pGem:CYP4G15     AATCCGAAGGTCTACGCCAATCCCAACGTCTTCGATCCGGACAACTTCCTGCCGGAGCGC  1500
                 ************************************************************

CYP4G15wt        CAGGCCAATCGCCACTACTACGCCTTTGTGCCGTTCTCTGCGGGACCACGCAGCTGTGTG  1560
pGem:CYP4G15     CAGGCCAATCGCCACTACTACGCCTTTGTGCCGTTCTCTGCGGGACCACGCAGCTGTGTG  1560
                 ************************************************************

CYP4G15wt        GGGCGCAAGTATGCGATGCTCAAGCTGAAGATCCTACTGTCGACCATTTTGCGGAACTAC  1620
pGem:CYP4G15     GGGCGCAAGTATGCGATGCTCAAGCTGAAGATCCTACTGTCGACCATTTTGCGGAACTAC  1620
                 ************************************************************

CYP4G15wt        CGTGTCTACTCGGACCTGACCGAATCGGACTTCAAACTGCAGGCGGACATCATCCTAAAG  1680
pGem:CYP4G15     CGTGTCTACTCGGACCTGACCGAATCGGACTTCAAACTGCAGGCGGACATCATCCTAAAG  1680
                 ************************************************************

CYP4G15wt        CGTGAGGAGGGATTCCGCGTGCGCCTCCAGCCGCGGACCAGTTGA 1725
pGem:CYP4G15     CGTGAGGAGGGATTCCGCGTGCGCCTCCAGCCGCGGACCAGTTGA 1725
                 *********************************************
```

Supplemental Figure 2: Sequencing results of pCW:CYP4G1Δinsertion.

```
CYP4G1wt            ATGGCAGTGGAAGTAGTTCAGGAGACGCTGCAACAAGCGGCGTCCAGTTCGTCGACGACG      60
CYP4G1Δinsertion    ATGGCAGTGGAAGTAGTTCAGGAGACGCTGCAACAAGCGGCGTCCAGTTCGTCGACGACG      60
                    ************************************************************

CYP4G1wt            GTCCTGGGATTCAGTCCTATGTTAACCACCTTAGTGGGCACCCTGGTGGCCATGGCATTG      120
CYP4G1Δinsertion    GTCCTGGGATTCAGTCCTATGTTAACCACCTTAGTGGGCACCCTGGTGGCCATGGCATTG      120
                    ************************************************************

CYP4G1wt            TACGAGTATTGGCGCAGGAATAGCCGGGAATACCGCATGGTTGCCAATATACCATCCCCA      180
CYP4G1Δinsertion    TACGAGTATTGGCGCAGGAATAGCCGGGAATACCGCATGGTTGCCAATATACCATCCCCA      180
                    ************************************************************

CYP4G1wt            CCGGAGTTGCCTATTTTGGGACAGGCTCATGTGGCCGCCGGCTTGAGCAATGCCGAGATC      240
CYP4G1Δinsertion    CCGGAGTTGCCTATTTTGGGACAGGCTCATGTGGCCGCCGGCTTGAGCAATGCCGAGATC      240
                    ************************************************************

CYP4G1wt            CTGGCCGTTGGCTTGGGTTACCTCAACAAGTACGGAGAAACCATGAAGGCCTGGTTGGGC      300
CYP4G1Δinsertion    CTGGCCGTTGGCTTGGGTTACCTCAACAAGTACGGAGAAACCATGAAGGCCTGGTTGGGC      300
                    ************************************************************

CYP4G1wt            AACGTCCTGTTGGTGTTTCTAACCAATCCCAGTGACATCGAGTTGATCCTGAGTGGGCAC      360
CYP4G1Δinsertion    AACGTCCTGTTGGTGTTTCTAACCAATCCCAGTGACATCGAGTTGATCCTGAGTGGGCAC      360
                    ************************************************************

CYP4G1wt            CAGCACTTGACCAAGGCGGAGGAGTATCGCTACTTCAAGCCCTGGTTCGGTGATGGTCTA      420
CYP4G1Δinsertion    CAGCACTTGACCAAGGCGGAGGAGTATCGCTACTTCAAGCCCTGGTTCGGTGATGGTCTA      420
                    ************************************************************

CYP4G1wt            CTGATCAGCAATGGACATCATTGGCGTCATCATCGTAAGATGATTGCCCCCACCTTCCAC      480
CYP4G1Δinsertion    CTGATCAGCAATGGACATCATTGGCGTCATCATCGTAAGATGATTGCCCCCACCTTCCAC      480
                    ************************************************************

CYP4G1wt            CAGAGCATCTTGAAGAGCTTCGTGCCTACATTTGTGGATCACTCAAAGGCGGTAGTTGCC      540
CYP4G1Δinsertion    CAGAGCATCTTGAAGAGCTTCGTGCCTACATTTGTGGATCACTCAAAGGCGGTAGTTGCC      540
                    ************************************************************

CYP4G1wt            AGGATGGGCTTAGAAGCGGGCAAATCCTTTGATGTTCATGACTATATGTCGCAGACCACG      600
CYP4G1Δinsertion    AGGATGGGCTTAGAAGCGGGCAAATCCTTTGATGTTCATGACTATATGTCGCAGACCACG      600
                    ************************************************************

CYP4G1wt            GTTGACATCCTGTTGTCTACCGCCATGGGTGTGAAGAAGCTTCCGGAGGGTAACAAGAGT      660
CYP4G1Δinsertion    GTTGACATCCTGTTGTCTACCGCCATGGGTGTGAAGAAGCTTCCGGAGGGTAACAAGAGT      660
                    ************************************************************

CYP4G1wt            TTCGAATACGCCCAAGCCGTCGTCGACATGTGTGATATCATACATAAGAGGCAGGTTAAA      720
CYP4G1Δinsertion    TTCGAATACGCCCAAGCCGTCGTCGACATGTGTGATATCATACATAAGAGGCAGGTTAAA      720
                    ************************************************************

CYP4G1wt            TTACTGTACCGCCTGGATTCCATCTACAAGTTTACTAAGCTTCGCGAGAAGGGCGATCGC      780
CYP4G1Δinsertion    TTACTGTACCGCCTGGATTCCATCTACAAGTTTACTAAGCTTCGCGAGAAGGGCGATCGC      780
                    ************************************************************

CYP4G1wt            ATGATGAACATCATCTTGGGTATGACCAGCAAGGTGGTCAAGGATCGTAAGGAGAACTTC      840
CYP4G1Δinsertion    ATGATGAACATCATCTTGGGTATGACCAGCAAGGTGGTCAAGGATCGTAAGGAGAACTTC      840
                    ************************************************************

CYP4G1wt            CAAGAGGAGTCACGTGCGATTGTTGAGGAGATTTCTACACCTGTTGCCAGCACTCCCGCT      900
CYP4G1Δinsertion    CAAGAGGAGTCACGTGCGA----------------------------------------      859
                    *******************

CYP4G1wt            TCCAAGAAGGAGGGTCTTCGCGATGATCTGGATGATATCGATGAAAATGATGTGGGCGCC      960
CYP4G1Δinsertion    -------------------------------------TCGATGAAAATGATGTGGGCGCC      882
                                                          **********************

CYP4G1wt            AAGAGGCGATTGGCTCTTCTAGATGCCATGGTGGAAATGGCTAAGAACCCCGATATCGAG      1020
CYP4G1Δinsertion    AAGAGGCGATTGGCTCTTCTAGATGCCATGGTGGAAATGGCTAAGAACCCCGATATCGAG      942
                    ************************************************************

CYP4G1wt            TGGAACGAGAAGGACATCATGGATGAGGTGAATACAATTATGTTTGAGGGCCACGATACC      1080
CYP4G1Δinsertion    TGGAACGAGAAGGACATCATGGATGAGGTGAATACAATTATGTTTGAGGGCCACGATACC      1002
                    ************************************************************

CYP4G1wt            ACCTCGGCGGGATCTAGTTTCGCCCTCTGCATGATGGGAATCCACAAGGACATCCAGGCT      1140
CYP4G1Δinsertion    ACCTCGGCGGGATCTAGTTTCGCCCTCTGCATGATGGGAATCCACAAGGACATCCAGGCT      1062
                    ************************************************************

CYP4G1wt            AAAGTCTTCGCCGAACAGAAGGCCATCTTCGGGGATAATATGCTGAGGGATTGCACCTTT      1200
CYP4G1Δinsertion    AAAGTCTTCGCCGAACAGAAGGCCATCTTCGGGGATAATATGCTGAGGGATTGCACCTTT      1122
                    ************************************************************

CYP4G1wt            GCCGATACCATGGAGATGAAATATTTGGAGCGCGTAATTTTAGAGACTTTGAGGTTGTAC      1260
CYP4G1Δinsertion    GCCGATACCATGGAGATGAAATATTTGGAGCGCGTAATTTTAGAGACTTTGAGGTTGTAC      1182
                    ************************************************************

CYP4G1wt            CCACCAGTACCACTTATCGCCAGGCGTCTGGACTACGACCTGAAGTTGGCCAGTGGTCCG      1320
CYP4G1Δinsertion    CCACCAGTACCACTTATCGCCAGGCGTCTGGACTACGACCTGAAGTTGGCCAGTGGTCCG      1242
                    ************************************************************

CYP4G1wt            TACACGGTTCCCAAGGGCACTACGGTCATCGTGCTGCAGTACTGCGTGCACAGACGTCCA      1380
CYP4G1Δinsertion    TACACGGTTCCCAAGGGCACTACGGTCATCGTGCTGCAGTACTGCGTGCACAGACGTCCA      1302
                    ************************************************************

CYP4G1wt            GACATCTACCCCAATCCCACCAAATTCGATCCGGACAACTTCCTACCCGAGAGGATGGCC      1440
CYP4G1Δinsertion    GACATCTACCCCAATCCCACCAAATTCGATCCGGACAACTTCCTACCCGAGAGGATGGCC      1362
                    ************************************************************
```

```
CYP4G1wt            AACAGGCATTACTACTCCTTCATTCCCTTTAGCGCTGGACCCAGAAGCTGTGTGGGCCGC      1500
CYP4G1Δinsertion    AACAGGCATTACTACTCCTTCATTCCCTTTAGCGCTGGACCCAGAAGCTGTGTGGGCCGC      1422
                    ************************************************************

CYP4G1wt            AAGTACGCCATGCTGAAGCTAAAGGTCCTGCTATCCACCATCGTGAGGAACTATATTGTC      1560
CYP4G1Δinsertion    AAGTACGCCATGCTGAAGCTAAAGGTCCTGCTATCCACCATCGTGAGGAACTATATTGTC      1482
                    ************************************************************

CYP4G1wt            CACTCCACCGACACGGAGGCAGATTTCAAGCTGCAGGCTGACATCATCCTAAAGCTTGAG      1620
CYP4G1Δinsertion    CACTCCACCGACACGGAGGCAGATTTCAAGCTGCAGGCTGACATCATCCTAAAGCTTGAG      1542
                    ************************************************************

CYP4G1wt            AATGGATTCAATGTCTCGTTGGAGAAGCGTCAGTACGCCACGGTGGCCTAGAATCCAGAA      1680
CYP4G1Δinsertion    AATGGATTCAATGTCTCGTTGGAGAAGCGTCAGTACGCCACGGTGGCCTAGAATCCAGAA      1602
                    ************************************************************

CYP4G1wt            ATCTAGGACCCCGACTACACACACGCAACCCCGAACCCGAAACCGGAATCCAGCCCTGTA      1740
CYP4G1Δinsertion    ATCTAGGACCCCGACTACACACACGCAACCCCGAACCCGAAACCGGAATCCAGCCCTGTA      1662
                    ************************************************************

CYP4G1wt            TATAGATGATGAATACCGATGAATATCCCAAACCGAAAACTTGATGACGAACTTATAAAT      1800
CYP4G1Δinsertion    TATAGATGATGAATACCGATGAATATCCCAAACCGAAAACTTGATGACGAACTTATAAAT      1722
                    ************************************************************

CYP4G1wt            CTAAAACACCGAATAAGAACCCAACGCACAAGCCAGCCAGAGAGTCAATTAATTTTTCTT      1860
CYP4G1Δinsertion    CTAAAACACCGAATAAGAACCCAACGCACAAGCCAGCCAGAGAGTCAATTAATTTTTCTT      1782
                    ************************************************************

CYP4G1wt            TCGTTTTTTAACTCGTTACTTTTATATTTGATTAATACCTTTTTGTTTGTTGGTCTTTAG      1920
CYP4G1Δinsertion    TCGTTTTTTAACTCGTTACTTTTATATTTGATTAATACCTTTTTGTTTGTTGGTCTTTAG      1842
                    ************************************************************

CYP4G1wt            CGAGTGGTGCCCCTATATAATGTATACGTATATACTATATATCCTTTTAACCAACTATTC      1980
CYP4G1Δinsertion    CGAGTGGTGCCCCTATATAATGTATACGTATATACTATATATCCTTTTAACCAACTATTC      1902
                    ************************************************************

CYP4G1wt            AACGCAACTGTTTGTGCTCTTCACCTTTTTAGTACTCCTACTTTTACCACTATCTATACT      2040
CYP4G1Δinsertion    AACGCAACTGTTTGTGCTCTTCACCTTTTTAGTACTCCTACTTTTACCACTATCTATACT      1962
                    ************************************************************

CYP4G1wt            TTTTTTTCGTAGCCATGTAGTGTGATTTTTTTTCTTTATTCTAGTATTTATTAAGTCAAA      2100
CYP4G1Δinsertion    TTTTTTTCGTAGCCATGTAGTGTGATTTTTTTTCTTTATTCTAGTATTTATTAAGTCAAA      2022
                    ************************************************************

CYP4G1wt            TGGTTTAAACGAAACCCAAAAAATATGAAAAATACACGTATGCGAGGCACGTAGCCGATA      2160
CYP4G1Δinsertion    TGGTTTAAACGAAACCCAAAAAATATGAAAAATACACGTATGCGAGGCACGTAGCCGATA      2082
                    ************************************************************

CYP4G1wt            GAGCTGCAAAAC      2172
CYP4G1Δinsertion    GAGCTGCAAAAC      2094
                    ************
```

Supplemental Figure 3: Sequencing results of (A) pGem:CYP4G16wt, (B) pGem:CYP4G16ΔN, (C) pGem:CYP4G17wt, (D) pGem:CYP4G17ΔN.

(A) pGem:CYP4G16wt

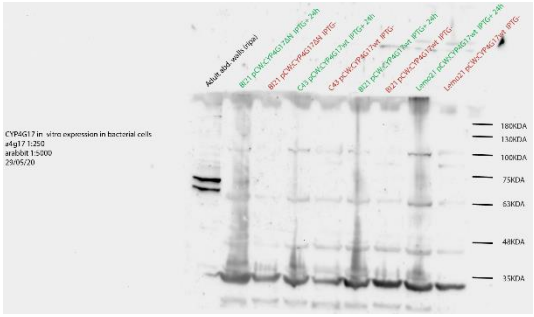
(B) pGem:CYP4G16ΔN

(C) pGem:CYP4G16ΔNhis

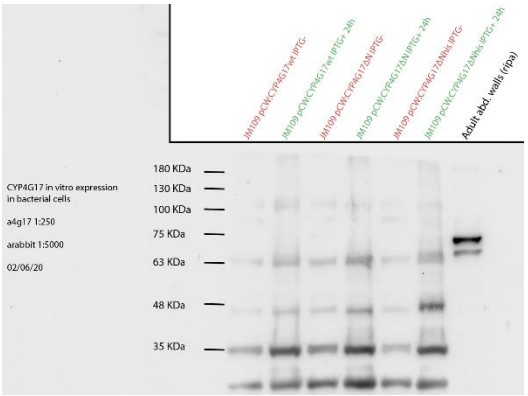
(D) pGem:CYP4G17wt

(E) pGem:CYP4G17ΔN

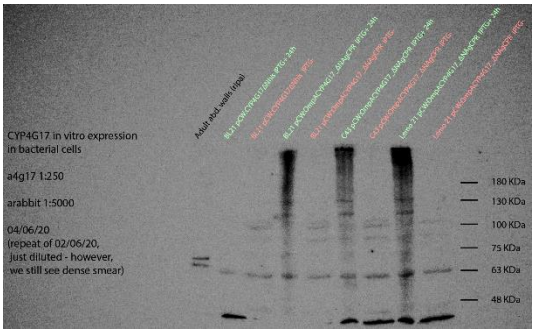
(F) pGem:CYP4G17ΔNhis



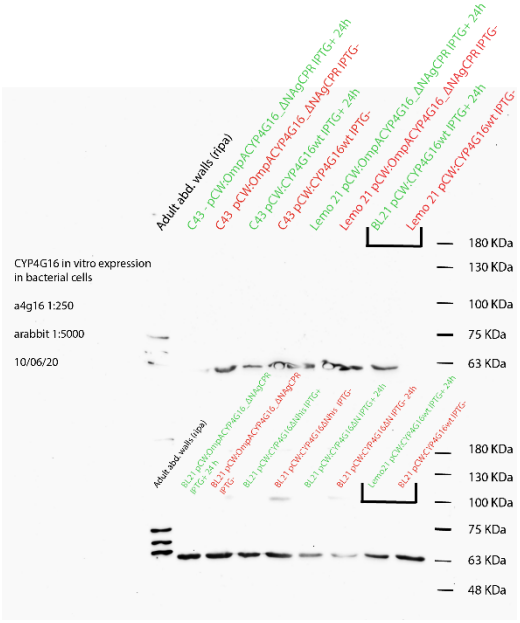
(A)



(B)

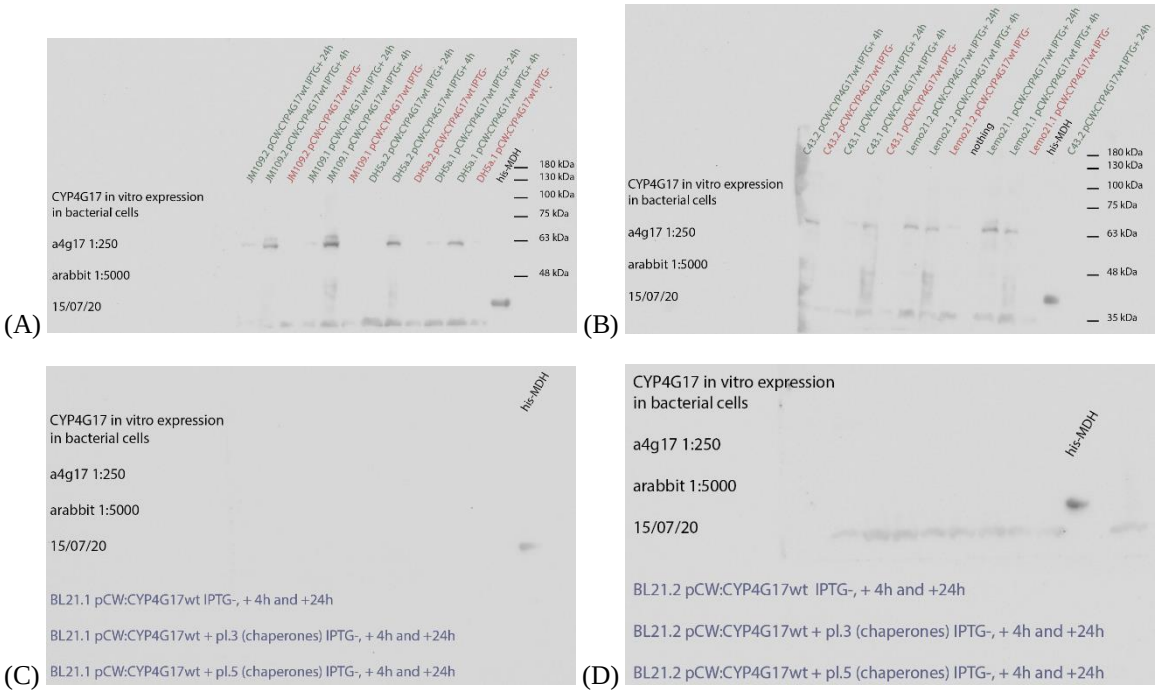


(C)

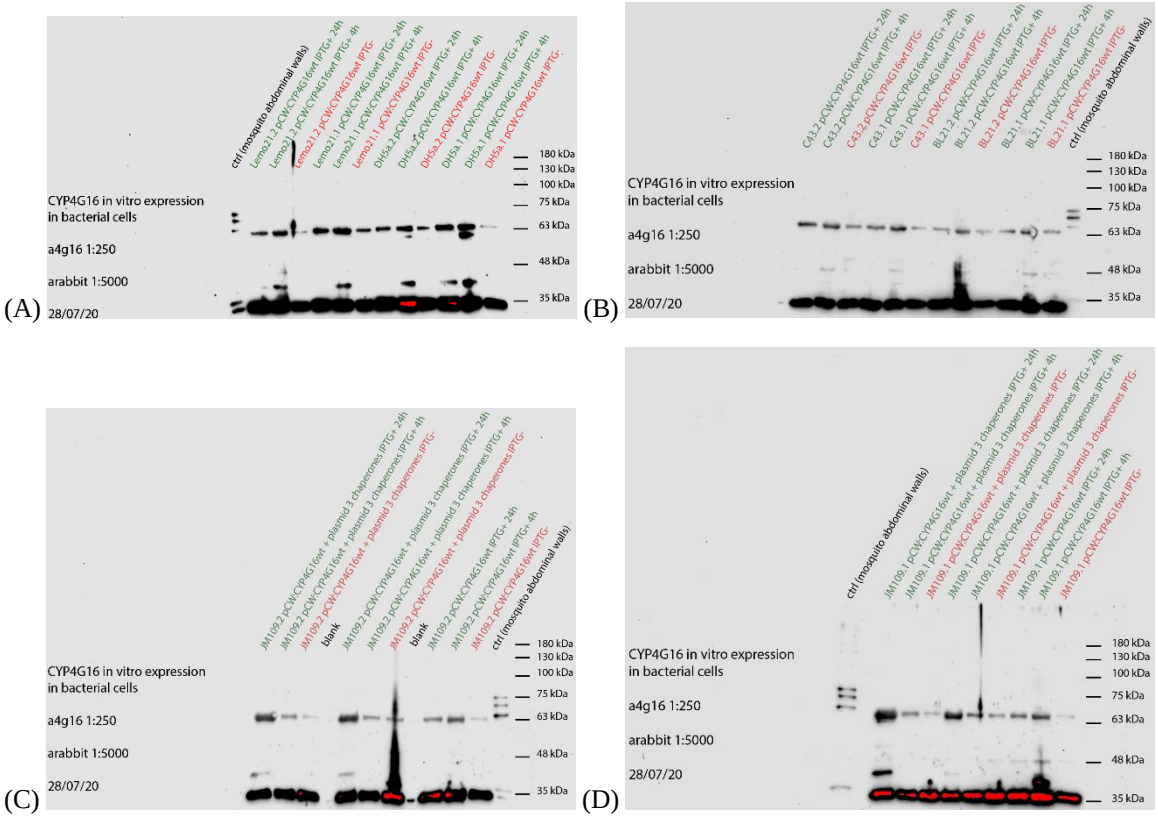


(D)

Supplemental Figure 4: Expression patterns of CYP4G16 or CYP4G17 among different combinations of plasmids and bacterial strains were analyzed by Western blot using anti-CYP4G16 or anti-CYP4G17. Each condition is examined before and after (4h) IPTG induction. In all gels, a control from adult mosquito abdominal walls and the ladder are depicted. (A) BL21 (DE3) Star, C43 and Lemo21 (DE3) cells were transformed with pCW:CYP4G17wt and BL21 (DE3) with pCW:CYP4G17ΔN. (B) JM109 cells were transformed with pCW:CYP4G17wt, pCW:CYP4G17ΔN and pCW:CYP4G17ΔNhis. (C) BL21 (DE3) Star, C43 and Lemo21 (DE3) cells were transformed with pCW:OmpA.CYP4G17_ΔNAgCPR and BL21 (DE3) cells with pCW:CYP4G17ΔNhis. (D) BL21 (DE3) Star cells were transformed with pCW:CYP4G16wt, pCW:CYP4G16ΔN, pCW:CYP4G16ΔNhis and pCW:OmpA.CYP4G16_ΔNAgCPR. C43 and Lemo21 (DE3) cells were transformed with pCW:CYP4G16wt and pCW:OmpA.CYP4G16_ΔNAgCPR.

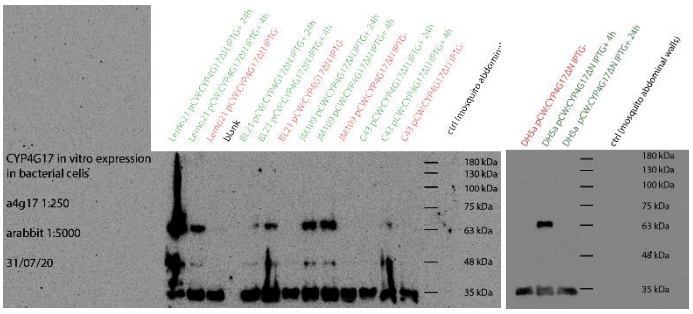


Supplemental Figure 5: Expression patterns of CYP4G17 among different combinations of plasmids and bacterial strains were analyzed by Western blot using anti-CYP4G17. In this experiment, 2 colonies for each combination were tested. Each condition is examined before and after (4h and 24h) IPTG induction. In all gels, a control from adult mosquito abdominal walls and the ladder are depicted. (A) DH5a and JM109 cells were transformed with pCW:CYP4G17wt. (B) Lemo21 (DE3) and C43 cells were transformed with pCW:CYP4G17wt. (C), (D) BL21 (DE3) Star cells were transformed with pCW:CYP4G17wt, pCW:CYP4G17wt in combination with chaperone plasmid 3 and pCW:CYP4G17wt in combination with chaperone plasmid 5.

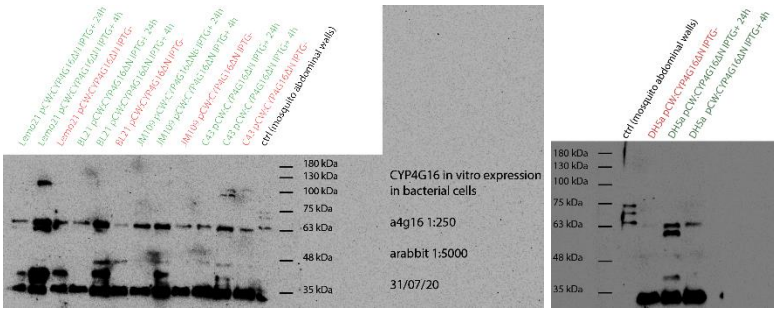


Supplemental Figure 6: Expression patterns of CYP4G16 among different combinations of plasmids and bacterial strains were analyzed by Western blot using anti-CYP4G16. In this experiment, 2 colonies for each combination were tested. Each condition is examined before and after (4h and 24h) IPTG induction. In all gels, a control from adult mosquito abdominal walls and the ladder are depicted. (A) DH5a and Lemo21 (DE3) cells were transformed with pCW:CYP4G16wt. (B) BL21 (DE3) Star and C43 cells were transformed with pCW:CYP4G16wt. (C), (D) JM109 cells were transformed with pCW:CYP4G16wt, pCW:CYP4G16wt in combination with chaperone plasmid 3 and pCW:CYP4G16wt in combination with chaperone plasmid 5.

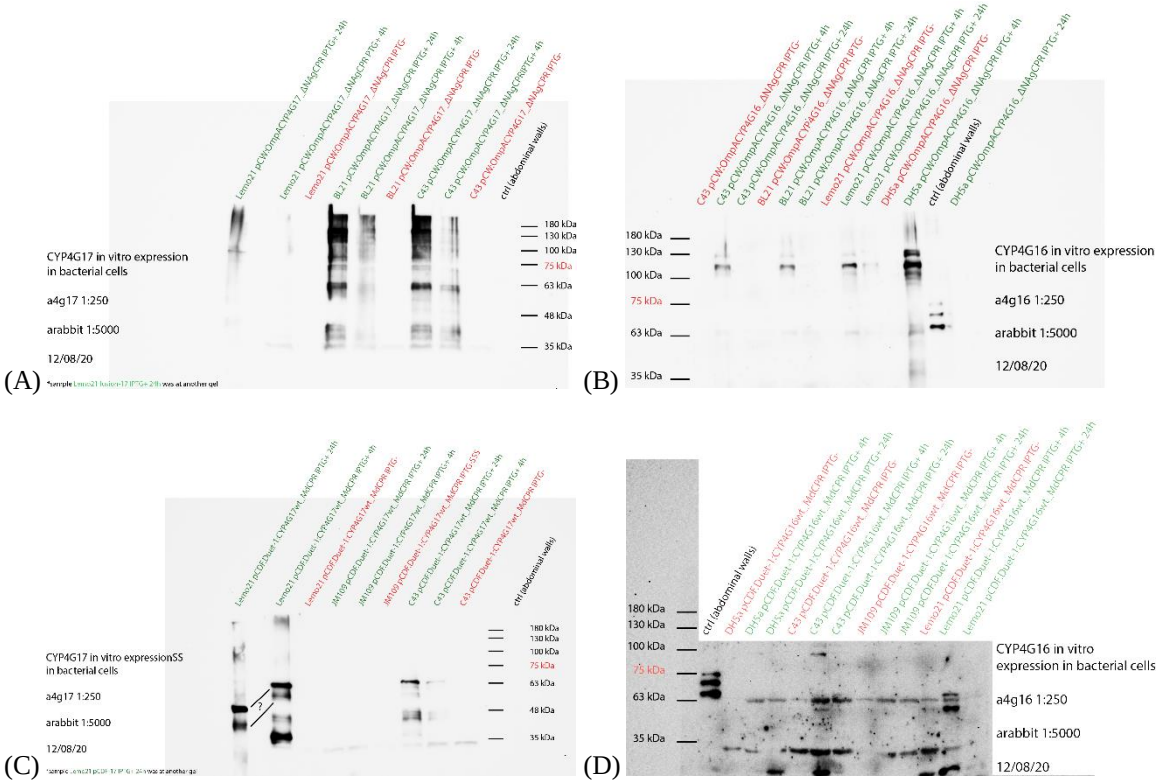
(A)



(B)



Supplemental Figure 7: Expression patterns of CYP4G16 or CYP4G17 among different combinations of plasmids and bacterial strains were analyzed by Western blot using anti-CYP4G16 or anti-CYP4G17. Each condition is examined before and after (4h and 24h) IPTG induction. In all gels, a control from adult mosquito abdominal walls and the ladder are depicted. (A) All five bacterial strains (DH5a, JM09, Lemo21 (DE3), BL21 (DE3) Star and C43) were transformed with pCW:CYP4G17ΔN. (B) All five bacterial strains (DH5a, JM09, Lemo21 (DE3), BL21 (DE3) Star and C43) were transformed with pCW:CYP4G16ΔN.



Supplemental Figure 8: Expression patterns of CYP4G16 or CYP4G17 among different combinations of plasmids and bacterial strains were analyzed by Western blot using anti-CYP4G16 or anti-CYP4G17. Each condition is examined before and after (4h and 24h) IPTG induction. In all gels, a control from adult mosquito abdominal walls and the ladder are depicted. (A) C43, BL21 (DE3) Star and Lemo21 (DE3) cells were transformed with pCW:OmpA.CYP4G17_ΔNAgCPR. (B) DH5a, C43, BL21 (DE3) Star and Lemo21 (DE3) cells were transformed with pCW:OmpA.CYP4G16_ΔNAgCPR. (C) C43, JM109 and Lemo21 (DE3) cells were transformed with pCDF.Duet-1:CYP4G17wt_MdCPR. (D) DH5a, C43, JM109 and Lemo21 (DE3) cells were transformed with pCDF.Duet-1:CYP4G16wt_MdCPR.