Genetic basis of octanoic acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped region

by

José Miguel Andrade López

A thesis submitted in partial fulfillment
of the requirement for the degree
Masters of Science
(Molecular, Cellular & Developmental Biology)
in the University of Michigan
2015

Master's Thesis Committee:

Professor Patricia Wittkopp, Chair
Assistant Research Scientist Joseph Coolon
Associate Professor Anuj Kumar
Associate Professor Orie Shafer

Table of Contents

I. List of Figures

II. Abstract

Drosophila sechellia is a species of fruit fly endemic to the Seychelles islands, which are located northeast of Madagascar off the coast of Africa. Unlike its generalist sister species *D. simulans* and *D. mauritiana,* and their closest relative *D. melanogaster*, *D. sechellia* evolved to specialize on a single plant species, *Morinda citrifolia*. Specialization on *M. citrifolia* is surprising because the fruit of the plant contains toxic compounds, primarily octanoic acid (OA), that are lethal to all *Drosophila* species except *D. sechellia*. Although the ecological and behavioral adaptations to this toxic fruit are known, the genetic basis for the evolutionary changes in OA resistance is not. Prior work showed that a genomic region on chromosome 3R, containing 18 genes, contributes to OA tolerance. To determine which gene(s) in this region might be involved in the evolution of OA resistance, I knocked-down expression of each gene in *D. melanogaster* with RNA interference (RNAi) (i) ubiquitously throughout development, (ii) during the adult stage, and (iii) within specific tissues in *D. melanogaster*. RNAi knockdown flies were tested for resistance to OA using the mixed effects Cox regression model. I found that knock-down of three neighboring genes, *Osiris 6*, *Osiris 7*, and *Osiris 8*, increased OA sensitivity. Tissue specific knockdowns, however, showed that decreasing expression of these genes in the fat body and salivary glands increases OA tolerance. I show that both *Osi6/7* are highly expressed during the first 24 hours of development and that exposure to different stressors induces expression in adults. Although *Osi6/7* have no coding change, RNA-seq data shows derived lower expression of these genes in *D. sechellia*; *Osi8* has two derived coding changes in *D. sechellia*. This study sheds light on the genetic basis of ecological adaptation to a toxic host within *Drosophila*, and insect-host specialization more broadly.

1. Introduction

Insects are among the most abundant and diverse group of organisms on the planet, with plant-feeding insects making up the majority of described species (Price 1980, Strong et al 1984, Jovilet 1992, Bernays and Chapman 1994). Most of these phytophagous insect species are considered specialists and feed on a small number of related plant species (Eastop 1973, Price 1980, Mitchell 1981, Ehrlich and Murphy 1988, Jovilet 1992, Bernays and Chapman 1994). Because these host-specific adaptations occur commonly, are key for ecological adaptation and are typically related to differences in plant chemistry, adaptations to novel host plants in phytophagous species are model traits for adaptive evolution in nature (Via et al 1999, Feder et al 2003). Typical adaptations associated with host-plant specialization include resistance to plant secondary defense compounds as well as preference behaviors associated with locating the new food source (Janike 1987, Via 1990, Futuyma 1991). However, little is known about the evolutionary mechanism by which this happens.

Fruit flies in the genus *Drosophila* are an excellent model for understanding the evolution of insect-host plant associations because of the incredible diversity of food sources used by these species and their frequent shifts between food sources (Matzkin et al 2006, Matzkin 2014, Linz et al 2013). The well-studied *Drosophila melanogaster* species group contains both generalist and specialist species, allowing dissection of the genetic basis of host transitions. The generalist species in this group include *D. melanogaster*, *D. simulans,* and *D. mauritiana,* which feed on the rotting fruit of several species of plants. Nested within this group of generalist species is a single derived specialist species, *D. sechellia*, which is endemic to the Seychelles islands and feeds almost exclusively on a single host plant: *Morinda citrifolia.*

Specialization on *M. citrifolia* is surprising because the fruit of the plant contains toxic defense compounds that are lethal to all other species of *Drosophila.* The primary toxin produced by *M. citrifolia* is octanoic acid (OA), a long chain fatty acid, which comprises 58% of the total volatile products in the fruit with hexanoic acid the second most abundant toxin making up 19% of the total volatile compounds (Legal 1999, Amlou 1998, Farnie et al. 1996 and Moreteau 1994). OA concentration varies during the ripening process with peak toxicity at full ripening (Legal 1994), and is detoxified over time by microorganisms, opening up the niche to the other *Drosophila* species (Figure S1) (R'Kha 1991). Because both adult and larval stages of *D. sechellia* are resistant to the OA levels present during the highest peak in toxicity (Jones 2001), *D. sechellia* has achieved a reproductive advantage through minimization of competition by being able to access the food source during an earlier time in the fruit's development.

Because the primary toxic compound in the fruit is OA, this toxin is used as a proxy for resistance studies in *D. melanogaster*. Sensitivity to OA varies within the *Drosophila* species group—*D. simulans* and *D. mauritiana* are both more sensitive to OA than *D. melanogaster,* and all three species are markedly more sensitive than *D. sechellia*, which shows tolerance to extremely high levels of OA (Jones 1998, Amlou et al 1997). In addition to resistance to OA, associated derived traits in *D. sechellia* differentiate the specialist from its sister species including increased egg production, attraction, and oviposition site preference for *M. citrifolia* (R'Kha 1991, Jones 2001). Recent studies have found overexpression of key genes involved in oogenesis and fatty acid metabolism, and suggest that the presence of l-DOPA in morinda fruit facilitated the specialization of *D. sechellia* on its toxic host (Dworkin and Jones 2009, Lavista-llanos et al 2014). The behavioral attraction to *M. citrifolia* is also observed with OA, and the opposite effect is seen with *D. simulans* (Jones 2001). Exposure of OA resembles that of

pyrethroid insecticides, causing flies to twitch and jerk in a nervous system-like response, suggesting that the volatile chemicals in *M. citrifolia* may act as neurotoxins (Beeman 1982, Legal 1992).

Genetic analyses of OA resistance in *D. sechellia* adults suggests that this trait is not highly polygenic. Five chromosomal regions have been mapped that contribute to variation in this trait, including a single region of large-effect on chromosome 3R (91A-93D) that explains ~15% of the difference between *D. simulans* and *D. sechellia* (Jones 1998). A recent study using introgression to move *D. sechellia* genomic regions conferring OA resistance into a *D. simulans* genetic background further narrowed this resistance locus to a single 170kb region containing 18 genes (Hungate et. al 2013). The genes in this region have a variety of predicted functions including three odor binding proteins (obp): *Obp83cd*, *Obp83ef*, and *Obp83g*; and nine *Osiris* genes which are biologically and molecularly uncharacterized but predicted to be transmembrane proteins and thought to be involved in the dosage-sensitive triple lethal locus (Dorer et al. 2003, Shah et al 2012).

To identify the gene(s) in this region most likely to contribute to the evolution of OA resistance, I used RNA interference (RNAi) in *D. melanogaster* using all available lines for these 18 genes (17/18) in the identified resistance locus to systematically test whether they play a role in OA resistance. Using two different screens of genes in this region, one knocking down each gene's function throughout development and the other knocking down each gene temporally in adults, I found that three genes, *Osiris 6* (*Osi6*), *Osiris 7* (*Osi7*), and *Osiris 8 (Osi8)*, increased sensitivity to the toxic effects of OA. Additionally, using a combination of different tissue-specific lines, I show knockdown of *Osi6* and *Osi7* in the fat body and salivary glands decreased sensitivity to OA, whereas *Osi8* does not. Knockdown of *Osi8* in the digestive system increased

sensitivity, suggesting other biological processes mediating sensitivity to toxins in *D. melanogaster*. Interestingly, *Osi6/7* are the only two genes in the resistance region that show derived, lower expression in *D. sechellia* based on whole-body RNA-seq data from *D. melanogaster*, *D. simulans*, and *D. sechellia*. *Osi6/7* also show induced expression in *D. melanogaster* when adult flies are exposed to different chemicals and environmental stressors, indicating possible functional roles, possibly even localized at specific tissues, of *Osi6/7* involved in toxin resistance. Because no coding change were found in *Osi6/7*, differences in their expression and/or protein sequence effects from the two coding changes in *Osi8* may have contributed to the evolution of OA resistance in *D. sechellia*.

2. Methods

2.1 Fly strains, rearing and husbandry

Strains of four species of *Drosophila* were used in this study: *D. melanogaster* (Canton S, Oregon R, *Zhr* (full genotype: XYS.YL.Df(1)Zhr), z30, 14021-0231.36, w1118, 60000), *D. simulans* (Tsimbazaza, 14021-0251.195), *D. mauritiana* (14021-0241.60) and *D. sechellia* (14021-0428.25, 14021-0428.08, 14021-0428.27, 14021-0428.07, 14021-0428.03). Additional *D. melanogaster* UAS-RNAi lines from the Vienna *Drosophila* UAS-RNAi Center, (VDRC# 10287, 42725, 18814, 40807, 33967, 7552, 5738, 33970, 9606, 43404, 26791, 42612, 5747, 102392, 44545, 8475, 5753, see Table S1) and a balanced ubiquitous GAL4 driver line (*actin*-GAL4/CyO). The GeneSwitch-UAS system was used to knockdown genes at the adult stage using the *Tubulin*-P[Switch] GAL4 driver. Tissue-specific GAL4 drivers were obtained from the Bloomington Stock Center, IN (Stock# 30843, 30844, 6357, 6890, 8527, 8765, 8180), and elav-GAL4 was obtained from the Bing Ye Lab (University of Michigan). Metabolism and Cytochrome P450(CYPs) lines were also obtained from Vienna *Drosophila* UAS-RNAi Center, (VDRC# 20183, 12016, 20183, 109463, 26603). All flies were reared on cornmeal medium using a 16:8 light:dark cycle at 25℃.

2.2 Octanoic acid mortality assay

Flies used in mortality assays were generated by crossing 3 virgin female with 3 males flies to control larval density. For ubiquitous RNAi experiments, *actin*-GAL4/CyO virgin females were crossed to UAS-RNAi males and all progeny were aged to 1-4 days post eclosion. Similarly, *Tubulin*-P[Switch] and tissue-specific GAL4 lines virgin females were crossed with UAS-RNAi males. Flies were then anesthetized with CO_2 and separated by sex and balancer

chromosome associated phenotypes (in *actin*-GAL4/CyO X UAS-RNAi, #8765 X UAS-RNAi cross progeny). Separated flies were then allowed to revive in empty fly vials (Genesee Scientific) with 10 flies per vial for 1.5 hours. Flies were then transferred into experimental vials containing 3.25g *Drosophila* instant media mix (0.75g *Drosophila* instant media flakes, 2.5g milli-Q H2O) (Carolina Biological) supplemented with 3.9μL of ≥99% octanoic acid (Sigma) to produce food with 1.2% OA. Gene-Switch crosses were reared at room temperature and F1 offspring were aged between 1-3 days. Aged flies were then transferred to fresh fly food mixed with mifepristone (RU486 Sigma, St. Louis) from a stock solution of 10 mg/ml in 100% EtOH to a final concentration of 10μg/ml overnight for 24 hours. Flies were then immediately used in the OA assay. The number of individuals "knocked down" (a fly was counted as "knocked down" when the individual was no longer able to walk or fly) was determined every five minutes for 60 minutes.

2.3 Mixed effect Cox regression analysis

A semi-parametric Cox proportional-hazard model was used to test the relative risk of OA exposure during gene knockdowns using the mixed effect Cox model , *coxme,* package in R (Terry 2012). The effects of the knocked-down genes on sensitivity to 1.2% OA were reported as regression coefficients, β, with 2SE as error bars. Because the Cox regression model is fitted using the *coxme* package in R, the β coefficient reads as a regression coefficient that when exponentiated gives the relative hazard in the treatment group, the RNA induced knockdown flies, compared against their sibling control. Vial number and day were included in the model as random effects, and sex was used as a multiplicative interaction variable:

coxme(Surv(Time,Status)~Gene*Sex+(1|Date)+(1|Vial),data=RNAi,ties=c("efron")).

A *coxph* survival curve estimate was jointly used from the fitted Cox model to visualize survival curves using OA concentrations ranging from 0.5 to 1.2 percent using 1-4 day old female *D. melanogaster* (*actin*-GAL4/CyO) individuals to determine the most appropriate OA concentration for subsequent assays (Cox 1992, Hertz-Picciotto and Rockhill 1997). The *coxph* package was also used to graphically represent proportional hazards within and between species as survival curves with 95% confidence intervals.

2.4 Gene expression analyses in Drosophila

Measures of gene expression were obtained from prior studies (Coolon et al 2014, Graveley et al 2011). The RNA gene expression measures were quantified using RNA-seq on whole adult (7-10 days post eclosion) female *D. melanogaster* (*zhr*), *D. simulans* (*tsimbazaza*) and *D. sechellia* (*droSec1*) were obtained from Coolon et al. (2014). Levels of gene expression quantified using RNA-seq on *D. melanogaster* (*y[1]; cn[1] bw[1] sp[1]*) across development (larvae, pupae, adult), *D. melanogaster* (*Oregon R*) in response to various perturbations (chemical exposure to Cadmium, Copper, Zinc, Caffeine, Paraquat, extended cold, cold shock, heat shock) and *D. melanogaster* (*Oregon R*) tissue-specific expression levels (whole fly males and females mated and unmated and aged 1,4 and 20 days, larval imaginal discs, larval and pupal central nervous system, larval and pupal salivary gland, larval and adult digestive system, larval and pupal fat body, ovary, testes, accessory gland and carcass) were obtained from the modENCODE project (Graveley et al 2011).

2.5 Sequence analyses: synonymous and nonsynonymous changes

Coding DNA sequences (CDS) for *Osi6*, *Osi7,* and *Osi8* were downloaded from FlyBase (Pierre et al 2014) for the *Drosophila* species with sequenced genomes publically available. Sequence was absent for the *D. simulans* ortholog of *Osi7* from the Flybase genome build, so I used recently published genomic sequence data from the *Tsimbazaza* isofemale line of *D. simulans* (Coolon et al 2014, McManus et al 2014). Sequence for the *D. mauritiana* orthologs of *Osi6*, *Osi7,* and *Osi8* was determined by Sanger sequencing the CDS. Protein sequences were aligned with GENEIOUS software (Biomatters Ltd.) and synonymous and nonsynonymous sequence changes were identified. Sequences of *Osi6/7/8* in *D. melanogaster, D. simulans,* and *D. sechellia* were re-confirmed with Sanger sequencing, and any observed coding changes in in *D. sechellia* were verified for fixation in the species by sanger sequencing of the additional *D. sechellia* laboratory strains. *Osiris* genes are known to be membrane proteins, and the system SOSUI was used to predict transmembrane regions for *D. sechellia*'s *Osi6/7/8* (Hirokawa et al 1998, Mitaku et al 1999, Mitaku et al 2002). SNPs that distinguish *D. sechellia* from its sister species were noted on the transmembrane SOSUI diagram for each of the gene.

3. Results

3.1 Quantifying octanoic acid knockdown in Drosophila

The resistance of *Drosophila sechellia* to the toxic effects of morinda fruit and its primary toxin OA is well documented; however, the assay by which toxicity was measured (e.g. exposure to OA vapor, OA in instant media food, and OA in *M. citrifolia* fruit) and the concentration of OA (0.1-0.5-100%) used in these studies varies considerably (Legal 1999, Amlou et al 1998, Morteau 1994, Farnie et al. 1996, Hungate et al 2013). To control the concentration of OA each fly experienced, I exposed flies to OA mixed into food (Amlou 1998). To determine the best concentration of OA to use for resistance experiments, I tested 1-3 day old adult female *D. melanogaster* (*actin*-GAL4/CyO) mortality associated with exposure to a series of six concentrations (0.5-1.2%, Figure 1). I found that mortality increased with increasing OA concentrations as expected. To ensure equal potential for identification of both increases and/or decreases from a baseline sensitivity, the data indicated that the best concentration to use in subsequent experiments was 1.2% OA, where approximately 50% mortality was observed within 60 minutes.

To quantify differences in OA tolerance among the members of the *melanogaster* species group (*D. melanogaster, D. simulans, D. sechellia,* and *D. mauritiana*), I performed mortality assays at a concentration of 1.2% OA (Figure 2A). Females are slightly more resistant than males, but not always significantly (Figure 2B), potentially due to differences in body size or other physiological difference between the sexes. Although some variation was observed between species strains (Figure 2C), the four species tested form distinct groups with varying levels of sensitivity to OA. The five *D. sechellia* lines were the most resistant having more than 80% survival at 60 minutes, both *D. simulans* lines were the least resistant and 100% death was

observed within 20 minutes of exposure, and the one *D. mauritiana* and the six *D. melanogaster* lines had intermediate resistance but at least some individuals survive past the 60 minutes of exposure to OA. Interestingly, resistance varied considerably within species among lines of both *D. sechellia* and *D. melanogaster* (Figure 2C).

I further analyzed these data using a Cox regression model (Cox 1992) to quantify the relative sensitivity of each line when exposed to OA using regression coefficients (β) with all lines treated with OA compared against one of the two *w1118* strains of *D. melanogaster* used in this study *(VDRC# 60000*, Figure 3). *Oregon-R* (β = -1.48, p = 0.015) and the other *w1118* strain (β = -1.2, p = 0.045) had the highest resistance to OA among the tested *D. melanogaster* strains. The five different lines of *D. sechellia* had extremely low, negative βs. Positive β values indicate increased sensitivity and negative values represent a decrease in OA sensitivity. β values for *D. sechellia* are lower than the *D. melanogaster w1118* baseline, therefore, indicating a decrease in OA sensitivity relative to the reference group (Figure 3). *D. mauritiana* showed high sensitivity to OA (β = 0.51, p = 0.036) similar to *D. simulans*; however, unlike *D. simulans*, some flies did survive exposure to OA after 60 minutes. The two *D. simulans* lines tested had the highest risk when exposed to OA (β = 1.48, p = 3.2E-9; β = 1.68, p = 7.7E-11), indicating an approximate 5-fold increased risk of knockdown upon exposure to OA for *D. simulans* relative to a *D. melanogaster* (*w1118)* across all the measured time intervals. Using the Cox regression model to quantify difference in OA sensitivity among the *Drosophila* species group, I was able to establish relative sensitivity differences between species that will allow us to compare changes in OA sensitivity in *D. melanogaster* when knocking-down individual genes in the resistance locus.

3.2 Ubiquitous RNAi knockdown of Osi6 *and* Osi7 *altered OA sensitivity*

To test the functional role of genes in the mapped chromosome 3R resistance region (Hungate et al 2013) on OA resistance, I used RNAi to knockdown expression of each gene in *D. melanogaster* and performed mortality assays. RNAi was performed in *D. melanogaster* lines containing transgenes that express hairpin RNAs under the control of the yeast upstream activating sequence (Dietzl et al 2007). When the hairpin-UAS lines are crossed to lines expressing the yeast GAL4 protein, the RNA hairpin is expressed and the gene targeted is knocked-down by the cells RNAi machinery (Figure 4A). Hairpin-UAS RNAi lines were available for 17 of the 18 genes in the resistance region. Only *Obp83g* was not available for this UAS/GAL4 system. Each UAS line was crossed to a ubiquitously expressed GAL4 line with expression driven by the *cis*-regulatory sequence from *Actin 5C*. To generate an internal control for each knockdown experiment, I used a line with *actin*-GAL4 on chromosome 2 that was heterozygous over a dominantly marked balancer chromosome (*CyO*) that when crossed to homozygous RNAi-UAS lines produced both RNAi-UAS/*actin*-GAL4 and RNAi/*CyO* progeny (Figure 4B). Using this approach I found that only two genes significantly altered OA sensitivity when knocked down, *Osi6* and *Osi7,* which both significantly increase sensitivity to OA (β = 2.65, p = 2E-5; β = 2.8, p = 7E-7, Figure 3B).

3.3 Stage-specific RNAi knockdown of Osi6 *and* Osi8 *altered OA sensitivity*

Because knockdown flies are subjected to OA as adults, and because the knocked-down genes might be important for other biological or developmental processes, I subsequently used the Gene-Switch system to induce gene knockdowns as adults and right before OA exposure. I assayed the mapped resistance locus with a hormone induced *Tubulin*-P[Switch] GAL4 driver. *Tubulin*-P[Switch] uses a modified chimeric GAL4 gene (Gene-Switch) that encodes the GAL4

DNA binding domain, the human progesterone receptor ligand-binding domain, and the activation domain form the human protein p65. The chimeric molecule only becomes active in the presence of the synthetic antiprogestin, mifepristone (RU486), and then binds to the UAS sequence to activate transcription of the RNA hairpin, knocking-down expression of that gene (Figure 5A). Using this inducible, stage-specific knockdown system allows us to test the function of key genes involved in OA resistance by removing expression of each gene only during the short window of OA exposure, bypassing other developmental stages in which the gene might be functionally important (Osterwalder et al 2001, Roman et al. 2001). Each UAS line was crossed to the hormone-induced *Tubulin*-P[Switch] GAL4 line and assayed in OA as was performed with the *actin*-GAL4 driver. *w1118* was used as control to exposure of RU486, and showed no effect of the treatment drug on OA sensitivity alone ($\beta = 0.47$, $p = 0.58$). All other knocked-down genes using this Gene-Switch system had their own control set as the baseline sensitivity, siblings unexposed to the drug. In this screen, knockdown of *Osi6* ($\beta = 2.02$, $p = 0.016$) and *Osi8* ($\beta = 2.13$, $p = 0.0015$) showed a significant increase in sensitivity compared against their respective siblings unexposed to RU486; whereas *Osi7* ($\beta = -0.19$, $p = 0.77$) had an indistinguishable change in sensitivity (Figure 5B).

3.4 Osi6/7/8 *expression varies by developmental stage, treatment, and tissue*

To determine the gene expression profiles for *Osi6*, *Osi7*, and *Osi8,* I obtained gene expression measures from *D. melanogaster* (Graveley et al 2011) including developmental stages, response to treatments, and across tissues to better understand a possible mechanism affecting OA resistance. I found that both *Osi6* and *Osi7* are highly expressed throughout development and show a cyclic expression profile, increasing and decreasing in expression

during each major developmental stage. Within the first 24 hours of development, *Osi6* and *Osi7* remain the most active with peak expression occurring at 16 hour (Figure 6A). They maintain this cyclic pattern throughout the larval and pupal stages, and show very low expression in adults. *Osi*8, however, is only expression during the pupal stage, peaking in expression at the two days post the larvae wandering pre-pupae stage (2d WPP), and no other expression throughout development was observed.

To determine the extent to which *Osi6*, *Osi7*, and *Osi8* are responsive to external stressors, I used modENCODE expression data and found that both *Osi6* and *Osi7* are sensitive to environmental toxicants, while *Osi8* does not show induced expression in response to environmental stress (Figure 6B). Both *Osi6* and *Osi7* are induced massively in response to heavy metal exposure (Cd, Cu, Zn), as well as organic compounds like caffeine and the pesticide paraquat. OA ($C_8H_{16}O_2$) and paraquat ($[(C_5H_4N)_2]Cl_2$), however, show no chemical similarities. Additionally, *Osi6* and *Osi7* are both induced in response to other types of environmental stress including heat shock, cold shock and extended cold (Figure 6B). *Osi6/7* expression increased almost 2-fold with increasing paraquat dose (5mM to 10mM). This is an interesting observation because among the 18 genes in the resistance locus, *Gasp* is the only other gene that shows some induction in adult flies caused from environmental toxicants, but no change in expression was shown between the 5mM and 10mM paraquat doses (no data was available for *CG31562*).

Tissue specific expression levels of *Osi6, Osi7*, and *Osi8* showed that all three genes are primarily expressed in the fat body and central nervous system (Figure 6C). Lower but detectable expression of both *Osi6* and *Osi7* was also observed in larval imaginal discs and salivary glands and in the adult male accessory gland. *Osi7* was generally expressed at a higher level than *Osi6* in each tissue, however, *Osi6* expression levels were higher in whole adults than *Osi7* suggesting

it may also be expressed in an untested tissue. Across development, both *Osi6* and *Osi7* are highly expressed at increasing levels with age, however, maximal observed expression is in pupal fat body. Because the adult fat body was not assessed in modENCODE, it remains unknown the degree to which these genes are expressed in this tissue.

3.5 Tissue-specific knockdown of Osi6/7/8

Because both the ubiquitous and stage-specific knockdown screens reduce expression of the genes throughout the entire organism, and because each gene's functional contribution to OA sensitivity might be localized in specific tissues, I used tissue-specific GAL4 lines to knockdown *Osi6/7/8* in specific tissues known to express the *Osiris* genes including the nervous system, salivary glands, and fat body. Interestingly, knockdown of both *Osi6* and *Osi7* in the salivary glands and fat body decreases sensitivity to OA in *D. melanogaster*; however, *Osi8* knockdown showed a different effect (Figure 7). Knockdown of *Osi8* in the salivary glands and hindgut show an increase in sensitivity to OA in *D.melanogaster* (Figure 7A). This is surprising because when *Osi6/7* are knocked-down in the whole body using an *actin*-GAL4 driver, sensitivity increases; however, when *Osi6/7* are exclusively knockdown in the fat body and salivary glands, the opposite effect is seen. *Osi8* knockdown follows a similar pattern as its ubiquitous knockdown.

Fatty acid metabolism genes have been shown to be differentially regulated between *D. sechellia* and *D. simulans* including *pdgy*, *Fad2*, *Arch42*, and *Cha* (Dworkin and Jones 2009). Because knockdown of *Osi6/7* in the fat body increases resistance to OA, I subsequently knockdown these metabolism genes in *D. melanogaster* using the Gene-Switch GAL4 driver (Figure 7B). However, no change in sensitivity was observed when these differentially regulated, metabolism genes were knocked-down (*pdgy*: $\beta = -0.27$, $p = 0.62$; *Fad2*: $\beta = 0.21$, $p = 0.78$;

Arc42: β = -0.69, p = 0.5; *Cha*: β = 0.98, p = 0.14).

3.6 Synonymous and nonsynonymous changes in Osiris *genes*

To identify possible sequence differences in *D. sechellia Osi6*, *Osi7*, and *Osi8* that could have functional consequences and therefore affect OA resistance, I aligned the coding sequence of *D. sechellia*, *D. simulans*, *D. mauritiana* and *D. melanogaster.* Orthologs of *Osi6* revealed 7 derived synonymous and no derived nonsynonymous changes on the *D. sechellia* lineage. *Osi7* showed 9 synonymous changes and a single derived nonsynonymous change in line 14021.0348.25 of *D. sechallia*; however, this is not a fixed difference because other *D. sechellia* strains do not share this sequence (Figure 8A). This rules out nonsynonymous changes in *Osi6/7* contributing to the resistance phenotype and suggests that non-coding sequences may be more likely to cause any functional differences that might exist between *D. sechellia* and other species at this locus. Different from that observed for *Osi6* and *Osi7*, the sequence of *D. sechellia's Osi8* has two derived nonsynonymous changes (*F95L and G129R*), suggesting that it may have functional consequences for the *Osi8* protein. Interestingly, most of the synonymous changes between *D. sechellia* and its three sister species occur at the domain of unknown function (dof1676), the segment of amino acids is right before the transmembrane region and inside of the cell, and includes the two derived coding changes in *Osi8* (Figure 8B).

While there has been much focus on the role that protein coding changes play in adaptive evolution for the last several years, many recent studies have shown that changes in gene regulation are equally important if not more common (Clark et al 2007, Orgogozo and Stern 2006, Orgogozo and Stern 2009, Andolfatto 2005). To identify possible species-specific differences in gene expression, orthologs of each of these genes in *D. melanogaster*, *D. simulans*

and *D. sechellia* whole adult females were obtained from published RNA-seq on whole female flies (Coolon et al. 2014). For both *Osi6* and *Osi7*, gene expression levels were identical in *D. melanogaster* and *D. simulans*, however, gene expression was significantly lower in *D. sechellia* (Figure 9). Levels of *Osi8* expression were extremely low and similar across all three species.

4. Discussion

Phytophagous insect-host specialization and host switching is a classical model of adaptive evolution and of great importance for agricultural crop pests. Understanding the genetic basis of such changes is therefore necessary, however, only a few case studies exist. Here I started with a resistance locus containing 18 genes, and used RNAi to functionally test the genes in the region, ultimately identifying three candidate genes, *Osi6*, *Osi7,* and *Osi8* that affect OA sensitivity in *D. melanogaster*. Using the UAS/GAL4 system, I have localized the genes involved in OA resistance in *D. melanogaster* and narrowing down the 170kb region to a 20kb region.

Studying the genetic basis of host-plant specialization can reveal useful information regarding the adaptation of species to novel habitats. *D. sechellia* is an excellent model system for such studies because of its recent divergence and adaptation to *M. citrifolia*, a plant that produces toxic fruit. Using the UAS/GAL4 system I ubiquitously knocked-down genes throughout development (*actin*-GAL4) and temporally (*Tubulin*-P[Switch]) in 3-day adult flies to screen all available *D. melanogaster* RNAi lines (17/18) from a genomic region on chromosome 3R shown to contribute to OA tolerance with introgression mapping. The mixed effect Cox regression model identified three neighboring genes, *Osi6, Osi7*, and *Osi8* where knock-down of the genes leads to an increase in OA sensitivity (Figure 4B and 5B). *Osi6* knockdown showed an effect in OA sensitivity using both screening methods, whereas *Osi7* and *Osi8* were only identified in one of the two screens. Although both GAL4 lines, in theory, ubiquitously knockdown gene expression in the whole organism, they use different promoters, either *actin* or *Tubulin*; however, difference in knockdown timing is more likely to contribute to the differences observed between the screens. To verify that the RNA inverted-repeat inserted in

each of the UAS lines for *Osi6*, *Osi7*, and *Osi8* were not having off-target effects among these similar genes, I aligned each inserted hairpin with each gene; however, no sequence similarities were observed. This suggests that the *Osiris* RNAi hairpin is not simultaneously knocking down the two other *Osiris* genes; however, it is still possible that the hairpin is having other off-target effects.

Worth noting is the number of progeny *actin*-GAL4/UAS-RNAi for *Osi6* and *Osi7* produce compared against the stage-specific GAL4/UAS-RNAi for *Osi6* and *Osi7* (Table S1). The former cross produces few F1 offspring, and mostly females, with an extended wing phenotype and limited in flight; however knocking down *Osi6/7* at specific tissues and using the *Tubulin*-P[Switch] GAL4 driver produces a normal number of offspring. Because *Osi6/7* show increased expression during the first 24 hours of embryonic development, it is possible that knocking down these genes may disrupt other important developmental properties not associated with OA sensitivity, producing sick individuals. The *actin*-GAL4/UAS-*Osi6/7* individuals do survive at least 4 days into adulthood, however, no longevity assay was performed on these individuals.

The *Osiris* gene family remains largely uncharacterized, but amino acid alignments of the 24 genes show conserved motifs including a signal peptide at the N-terminus and a single transmembrane region, denoted from the presence of a conserved hydrophobic region, suggesting that *Osi6*, *Osi7*, and *Osi8* are localized in the membrane (Shah et al. 2012). Considering that both *D. sechellia* larvae and pupae spend their developmental stages exposed to the toxic fruit whereas adult flies only feed and oviposit at the site, the developmental expression profile of these genes could function in resistance.

Among the 18 genes in the 170kb tolerance region only 3 genes (*Gasp*, *Osi6*, *Osi7*) show

induced expression to different chemicals and stressors (Graveley et al 2011). Among these is paraquat, a neurotoxin commonly found in pesticides; however, paraquat shares no chemical similarities to OA (Nistico et al 2011). The strong induction of these genes in response to other toxicants further suggests that gene expression may be involved in *D. sechellia'*s evolved resistance to OA. It would be interesting to comparatively look at changes in gene expression of *Osi6, Osi7,* and *Osi8* between species when exposed to OA. If *Osiris* genes change expression in response to environmental cues like OA, this might suggest a secondary regulatory mechanism mediating resistance in *D. sechellia*.

Temporal studies in *D. melanogaster* using RNA-seq data of larvae, pupae, and adult flies shows very high expression of *Osi6*, *Osi7*, and *Osi8* almost exclusively in the fat body and the central nervous system at least in pupae, and *Osi6* and *Osi7* are also primarily expressed in the digestive system and head for adult flies. This expression profile is interesting because the fat body is a common site of detoxification (Kilby 1963, Arrese and Soulages 2010). Additionally, because OA induces a twitching response in *Drosophila*, expression of these genes in the central nervous system could suggest that the CNS could be the tissue through which the *D. sechellia* alleles act. Because adult flies show expression in the digestive system and because assayed flies were quarantined for 1.5 hours before food and toxin exposure, the assay used in this study allows for multi-sensory response to occur at once, both peripheral (touch) and gustatory (feeding) effects of OA in *Drosophila*. This will ensure that whichever mode *D. sechellia* used to respond to OA exposure, it will be accounted for in the OA assay.

Interestingly, using tissue-specific GAL4 drivers to knockdown *Osi6/7/8* leads to a decrease in sensitivity when I knockdown both *Osi6* and *Osi7* in the fat body and salivary glands. Because studies have suggested that host specialization often contributes to the evolution

of novel morphological, behavioral, or physiological traits (Jones 2009), it is possible that *D. sechellia* might be using these *Osiris* genes in a new way, so future studies investigating the localization of these genes throughout development within and between species will yield useful information about *Osiris* genes' involvement in resistance to OA. Using the tissue-specific GAL4 driver, w^{*}; P{GawB}$c601^{c601}$ (Hgut), to knockdown *Osi8* simultaneously in the hindgut, ureter, malpighian tubules, and protocerebrum showed a large increase in sensitivity to OA. A similar increase in sensitivity was seen in the elav-GAL4 (NS+SG) driver, but not independently in the P{GAL4-elav.L}2 (NS) or P{Sgs3-GAL4.PD}TP1 (SG) GAL4 drivers suggesting that the latter two drivers use different tissue promoters than the NS+SG driver and might be localized in different spatial areas within those tissues.

Although RNAi knockdown using a UAS/GAL4 system approach can produce variability in successful disruption of a gene, this is the best means of screening the resistant locus for three reasons: one, mutant lines were not available for the entire 18 genes; two, the genes might be important during developmental and might produce sick offspring, making it difficult to quantify the sensitivity phenotype in adults; and three, because of the variability of OA sensitivity observed among different *D. melanogaster* lines, it is difficult to find an appropriate control to compare against the mutant lines. The UAS/GAL4 system allows us to use uninduced siblings (*actin*-GAL4/CyO and -RU486, *Tubulin*-P[Switch]/UAS-RNAi) as comparisons, controlling for environmental and genetic background effects on the OA sensitivity phenotype. Because alignment of the inserted hairpins designed to knockdown each of the *Osiris* genes do not appear to be having any off-target effects between *Osi6/7/8*, it is not likely that the candidate genes from *actin*-GAL4 and *Tubulin*-P[Switch] GAl4 drivers is due to the RNA inverted-repeat for *Osi6* knocking down *Osi6* and *Osi7* in the *actin*-Gal4 screen or vice versa. Instead, it might be a result

of difference in knockdown timing or other unknown biological properties. Existing lines containing mutant alleles of *Osi7* and *Osi8* were not available to verify UAS/GAL4 results and while there was a line carrying a mutant *Osi6* allele, sequence and expression analyses showed that it was actually not carrying a mutant allele (Figure S2).

The genetic screens using two different GAL4 drivers points to *Osi6* involved in OA sensitivity in *D. melanogaster* in both instances, and because there are no coding differences between species, the regulatory region upstream of *Osi6* might be involved in the resistance phenotype. It would be useful to look into *Osi6* controlling the regulation of *Osi7* and *Osi8* by using either qPCR or pyrosequencing to confirm any possible downstream interactions between the three genes.

Moving away from gene expression in *D. melanogaster* and looking at expression changes of the 18 genes in *D. sechellia* shows an interesting derived expression profile of *Osi6* and *Osi7*. RNA-seq expression data from whole female flies shows *Osi6* and *Osi7* having derived reduction in gene expression in *D. sechellia* relative to both *D. melanogaster* and *D. simulans*. No other gene in the resistance locus has this novel expression change. These findings suggest that if the mechanisms mediating resistance in *D. sechellia* is derived from changes in regulatory expression of genes, then both *Osi6* and *Osi7's* regulatory changes could contribute to OA resistance. However, a combination of regulatory and coding changes might also be a possible explanation, and *Osi8* is a candidate if this is the case.

From the three identified candidate genes, *Osi6*, *Osi7*, and *Osi8*, both *Osi6* and *Osi7* have no coding changes and are differentially expressed between species; therefore, the role they play in *D. sechellia* is likely through regulatory changes. Because there are no nonsynonymous changes in either *Osi6* or *Osi7*, sensitivity to OA might be due to a regulatory change in *D.*

sechellia or caused by either or both of the nonsynonymous changes in *Osi8*. These changes may contribute OA resistance in *D. sechellia*; however, regulatory changes cannot be ruled out. The two nonsynonymous changes in *D. sechellia's Osi8*, F95L and G129R, offer a possible candidate for a new genome editing approach that allows specific changes to the genome at the nucleotide level, the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system (Barrangou et al 2007). The system requires a 20bp chimeric guide RNA (gRNA) and the Cas9 endonuclease, which work in unison to target a DNA segment, create double stranded cut sites, and removing fragments of DNA. The cut DNA strand can then be reassembled through non-homologous end joining, which causes indel frameshifts and inactivates the gene, or by injecting a donor DNA into the organism to become incorporated between both cut sites of the genome through homology directed repair.

If regulatory regions are responsible for OA resistance between species, I can use this system to localize the regulatory regions responsible for OA resistance in *D. sechellia*. Whole regulatory regions can be swapped between organisms to identify regions important for OA sensitivity within the three *Osiris* genes. Similarly, if coding regions are responsible for OA resistance, I can make nonsynonymous changes in *D. melanogaster* and/or *D. simulan*s mirroring the derived coding change in *D. sechellia*. Making these nucleotide changes in the genome of other *Drosophila* species will allow us to investigate the single point mutations leading to amino acid substitutions in *D. sechellia* and their possible role in OA resistance. This system will also allow me to mutate *D. sechellia's Osi6/7/8* to look for decrease in OA resistance. This editing tool has already been used in several model organisms including *Drosophila* (Hwang et al 2013, Wang et al 2013, Gratz et al 2013, and Fang et al 2013), and might serve as the next step in studying OA resistance in *D. sechellia*.

Starting with a resistance locus containing 18 genes, I used RNAi to functionally test the genes in the region and identified three candidate genes, *Osi6*, *Osi7,* and *Osi8* that affect OA resistance in *D. melanogaster*. In the resistance locus, only *Osi6* and *Osi7* have divergent expression levels in *D. sechellia* and are expressed in tissues commonly associated with response to toxin exposure. Furthermore, *Osi6* and *Osi7* are strongly induced in response to other environmental toxicants making them good candidates for response to OA. Finally, the *D. sechellia Osi8* ortholog has two derived nonsynonymous mutation suggesting they may have functional consequences. While I cannot rule out that use of RNAi in a heterologous system (*D. melanogaster*) may have affected the results of our assay, these are the best candidates to date for a role in *D. sechellia* resistance to OA after more than 15 years of interrogation.

5. Acknowledgements

I would like to thank Patricia Wittkopp and Joseph Coolon for comments on this thesis, the Bloomington Drosophila Stock Center, the University of San Diego Stock Center and the Vienna Drosophila Resource Center, the Scott Pletcher and Bing Ye labs for fly lines. I also received invaluable experimental suggestions and advice from Anuj Kumar and Orie Shafer. I am grateful to Laura Olsen and Anuj Kumar for running the first MCDB Pathways Masters program at the University of Michigan, and for the amazing mentoring I received from them throughout my involvement in the program. Funding for this work was provided by the National Institutes of Health (5F32GM089009-02 to JDC), the National Science Foundation (MCB-1021398 to PJW, 1038099 for the Pathways Master's Program support for JMAL), and the Alfred P. Sloan Research Foundation (Fellowship to PJW). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Institutes of Health, National Science Foundation or Sloan Foundation.

6. References

Amlou, M, & Pla, E (1997). Genetic analysis by interspecific crosses. *Genetics Selection Evolution,* **29**, 511-522

Amlou, M, Moreteau, B, & David, JR (1998). Larval tolerance in the Drosophila melanogaster species complex toward the two toxic acids of the D. sechellia host plant. *Hereditas*, **129**, 7-14.

Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in Drosophila. *Nature*, ***437***(7062), 1149-1152.

Arrese, E. L., & Soulages, J. L. (2010). Insect fat body: energy, metabolism, and regulation. *Annual Review of Entomology*, ***55***, 207.

Barrangou, R, Fremaux, C, Deveau, H, Richards, M, Boyaval, P, Moineau, S, Romero, DA, Horvath, P (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.

Beeman, RW (1982). Recent advances in mode of action of insecticides. *Annual Review of Entomology*, **27**, 253-281.

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., ... & Civetta, A. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, ***450***(7167), 203-218.

Coolon, JD, McManus, CJ, Stevenson, KR, Graveley, BR, & Wittkopp, PJ (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome Research*, **24**, 797-808.

Cox, DR (1992). Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, **34**, 187-220.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S,

Scheiblauer S, Couto A, Marra V, Keleman K, Dickson BJ (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature*, **448**,151–156

Dorer, DR, Rudnick, JA, Moriyama, EN, & Christensen, AC (2003). A family of genes clustered at the Triplo-lethal locus of Drosophila melanogaster has an unusual evolutionary history and significant synteny with Anopheles gambiae. *Genetics*, **165**, 613-621.

Dudley, R. (2000). Evolutionary origins of human alcoholism in primate frugivory. *Quarterly Review of Biology.* **75**, 3–15.

Dudley, R. (2004). Ethanol, fruit ripening, and the historical origins of human alcoholism in primate frugivory. *Integrative and Comparative Biology*, ***44***(4), 315-323.

Dworkin I, Jones CD (2009). Genetic changes accompanying the evolution of host specialization in Drosophila sechellia. *Genetics* **181**, 721–736.

Ehrlich, PR and Murphy, DD (1988). Plant chemistry and host range in insect herbivores. *Ecology*, **69**, 908-909.

Farine, JP, Legal, L, Moreteau, B, & Le Quere, JL (1996). Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. *Phytochemistry*, **41**, 433-438.

Feng, Z, Zhang, B, Ding, W, Liu, X, Yang, DL, Wei, P, Cao, F, Zhu, S, Zhang, F, Mao, Y and Zhu, JK (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, *23*(10), 1229.

Gratz, SJ, Cummings, AM, Nguyen, JN, Hamm, DC, Donohue, LK, Harrison, MM, Wildonger, J, and O'Connor-Giles, KM (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, **194**, 1029-1035.

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., ... &

Celniker, S. E. (2011). The developmental transcriptome of Drosophila melanogaster. *Nature*, *471*(7339), 473-479.

Hertz-Picciotto, I and Rockhill, B (1997). Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics*, **53**, 1151-1156.

Hirokawa T., Boon-Chieng S., and Mitaku S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, **14**, 378-9.

Hungate, EA, Earley, EJ, Boussy, IA, Turissini, DA, Ting, CT, Moran, JR, ...& Jones, CD (2013). A locus in Drosophila sechellia affecting tolerance of a host plant toxin. *Genetics*, **195**, 1063-1075.

Hwang, WY, Fu, Y, Reyon, D, Maeder, ML, Tsai, SQ, Sander, JD, Peterson, RT, Yeh, J-R and Joung, JK (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, **31**, 227-229.

Kilby, B. A. (1963). The biochemistry of the insect fat body. *Advances in Insect Physiology*, **1**, 111-174.

Lavista-Llanos, S., Svatoš, A., Kai, M., Riemensperger, T., Birman, S., Stensmyr, M. C., & Hansson, B. S. (2014). Dopamine drives Drosophila sechellia adaptation to its toxic host. *eLife*, ***3***, e03785.

Legal, L, David, JR, & Jallon, JM (1992). Toxicity and attraction effects produced by Morinda citrifolia fruits on the Drosophila melanogaster complex of species. *Chemoecology*, **3**, 125-129.

Legal, L, Chappe, B, & Jallon, JM (1994). Molecular basis of Morinda citrifolia (L.): Toxicity on drosophila. *Journal of Chemical Ecology*, **20**, 1931-1943.

Legal, L, Moulin, B, & Jallon, JM (1999). The Relation between Structures and Toxicity of

Oxygenated Aliphatic Compounds Homologous to the Insecticide Octanoic Acid and the Chemotaxis of Two Species of Drosophila. *Pesticide Biochemistry and Physiology*, **65**, 90-101.

Linz, J, Baschwitz, A, Strutz, A, Dweck, HK, Sachse, S, Hansson, BS, & Stensmyr, MC (2013). Host plant-driven sensory specialization in Drosophila erecta. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20130626.

Jones, CD (1998). The genetic basis of Drosophila sechellia's resistance to a host plant toxin. *Genetics*, **149**, 1899-1908.

Jones, CD (2001). The genetic basis of larval resistance to a host plant toxin in Drosophila sechellia. *Genetical Research*, **78**, 225-233.

Matzkin, LM, Watts, TD, Bitler, BG, Machado, CA, & Markow, TA (2006). Functional genomics of cactus host shifts in Drosophila mojavensis. *Molecular Ecology*, **15**, 4635-4643.

Matzkin, LM (2014). Ecological genomics of host shifts in Drosophila mojavensis. *Advances in Experimental Medicine and Biology,* **781**, 233-247.

McManus, CJ, Coolon, JD, Duff, MO, Eipper-Mains, J, Graveley, BR, & Wittkopp, PJ (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, **20**, 816-825.

Mitaku, S., & Hirokawa, T. (1999). Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. *Protein Engineering*, ***12***(11), 953-957.

Mitaku S., Hirokawa T., and Tsuji T. (2002). Amphiphilicity index of polar amino acids as an aid

in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics*, **18**, 608-16.

Moreteau, B, R'Kha, S, & David, JR (1994). Genetics of a nonoptimal behavior: Oviposition preference of Drosophila mauritiana for a toxic resource. *Behavior Genetics*, **24**, 433-441.

Nistico, R., Mehdawy, B., Piccirilli, S., & Mercuri, N. (2011). Paraquat-and rotenone-induced models of Parkinson's disease. *International Journal of Immunopathol Pharmacol*, **24**, 313-22.

Orgogozo, V., Broman, K. W., & Stern, D. L. (2006). High-resolution quantitative trait locus mapping reveals sign epistasis controlling ovariole number between two Drosophila species. *Genetics*, ***173***(1), 197-205.

Orgogozo, V., & Stern, D. L. (2009). How different are recently diverged species? More than 150

phenotypic differences have been reported for the D. melanogaster species subgroup. *Fly*, ***3***(2), 117-117.

Osterwalder T, Yoon KS, White BH, Keshishian H (2001). A conditional tissue-specific transgene expression system inducible Gal4. *Proceedings of the National Academy of Sciences. USA* **98**, 12596-12601.

St. Pierre SE, Ponting L, Stefancsik R, McQuilton P, and the FlyBase Consortium (2014). FlyBase 102 - advanced approaches to interrogating FlyBase. *Nucleic Acids Research.* ***42****(D1), D780-D788.*

R'kha, S, Capy, P, & David, JR (1991). Host-plant specialization in the Drosophila melanogaster species complex: a physiological, behavioral, and genetical analysis. *Proceedings of the*

National Academy of Sciences, **88**, 1835-1839.

Roman G, Endo K, Zong L, Davis RL (2001). P{Switch}, a system for partial and temporal control of gene expression in Drosophila melanogaster. *Proceedings of the National Academy of Sciences. USA* **98**, 12602-12607.

Shah, N, Dorer, DR, Moriyama, EN, & Christensen, AC (2012). Evolution of a large, conserved, and syntenic gene family in insects. *G3*, **2**, 313-319.

Therneau, T. (2012). coxme: Mixed effects Cox models. *R package version*, *2*(3).

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910-918.

Figure 1

Using a combination of six different OA concentration mixed into fly food and exposing 3-day old adult female *D. melanogaster* (*actin*-GAL4/CyO) , I identified 1.2% OA as the optimum dosage, yielding about 50 percent death within 60 minutes. Sensitivity to OA increases with OA dose.

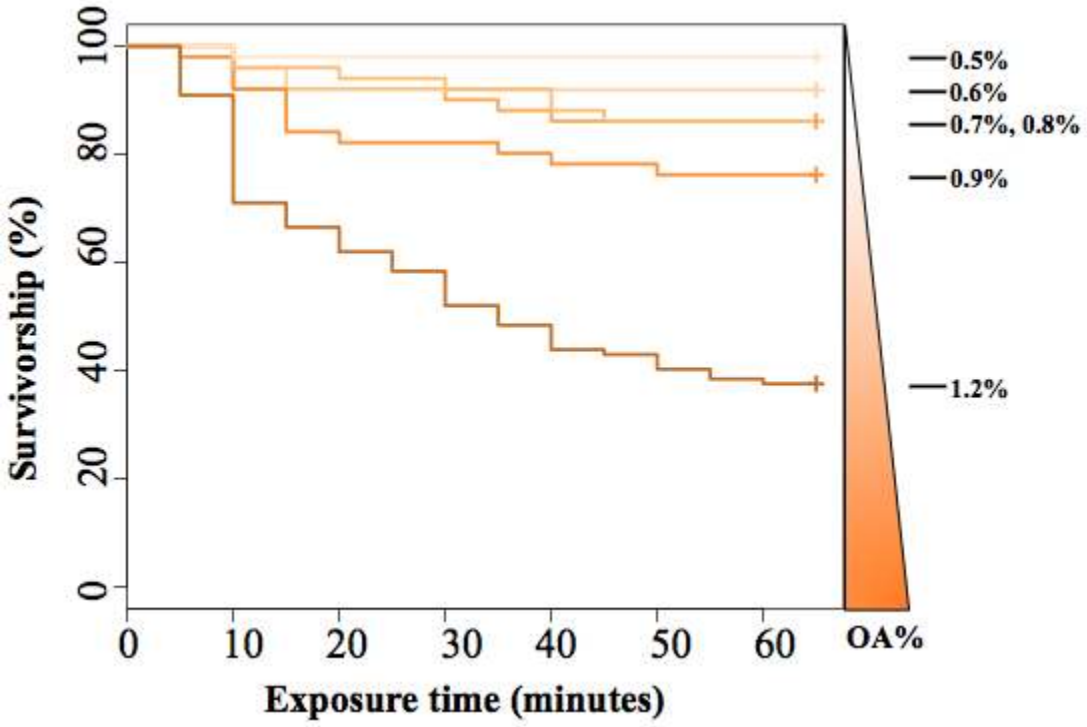


Figure 2A

Survival curves showing a representative lab strain from each of the *Drosophila melanogaster* species group. Dotted lines are represented as 95 percent confidence intervals. Distinct significant differences are observed between species, where *D. sechellia* is the most resistant to 1.2% OA, *D. melanogaster* has intermediate resistance, *D. mauritiana* and *D. simulans* are most sensitive to OA. Among the three species lines tested, only *D. simulans* individuals died within the first 20 minutes of exposure, all other species lines had some individuals survive past 60 minutes.

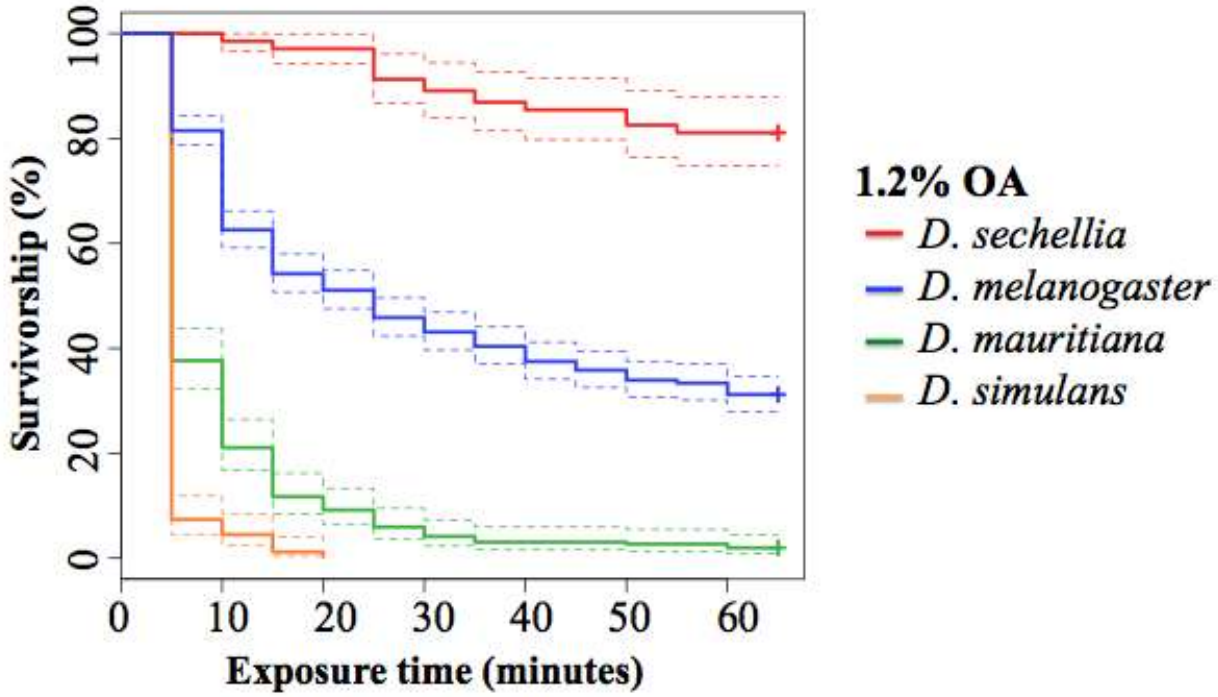


Figure 2B

Survival curves showing sex-specific sensitivity difference within *Drosophila*, dark curves represent females and light curves males, with 95 percent confidence intervals represented by dotted lines for each sex. The representative *D. sechellia* line shows significant difference between males and females, females being the most resistant of the sex, the other three species do not show significance. Not all *D. sechellia* lines show significant sex-specific difference in OA sensitivity.

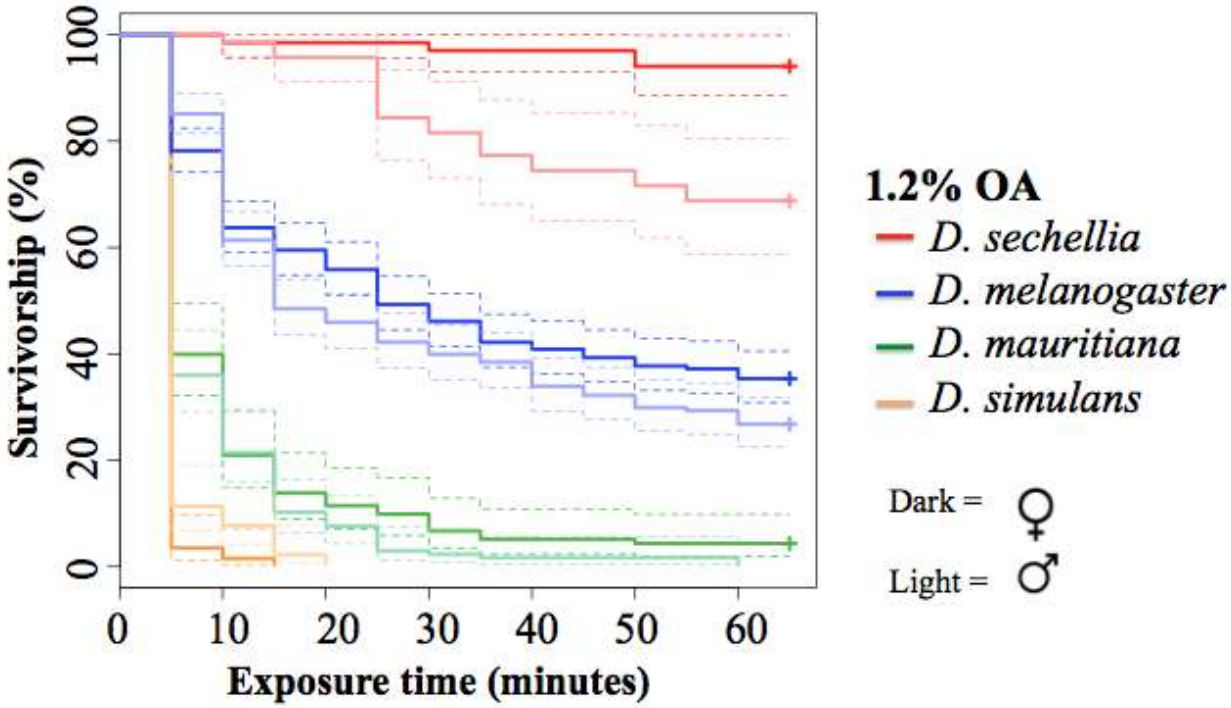


Figure 2C

Survival curves for all available species line showing variation, especially in *D. sechellia* and *D. melanogaster*. *D. sechellia* shows at least 80% survival to OA after 60 minutes, and in the other extreme, both *D. simulans* lines did not survive past 20 minutes of OA exposure. *D. mauritiana* also has low tolerance to OA, but some individuals did survive expose past 60 minutes.

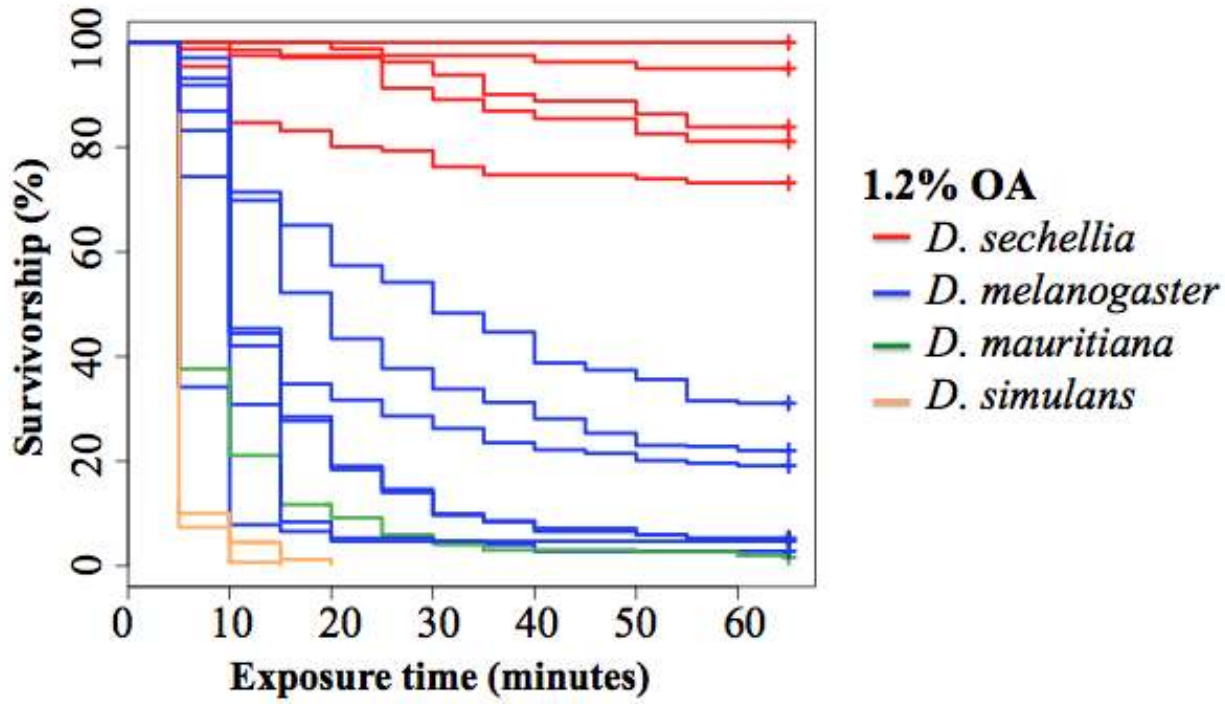


Figure 3

Using a mixed effect Cox regression model, *coxme* package in R, survival curves for all the lab species strains were converted into beta regression coefficients, β. Error bars are represented by 2SE. Exponentiating β gives the fold increase or decrease in OA sensitivity relative to a *D. melanogaster*, *w*, baseline sensitivity, represented by the horizontal line with a set sensitivity of 0. The species grouping observed on the survival curves remains. All *D. melanogaster* strains fall within the 0 baseline sensitivity indicating similar sensitivity to the reference *w* line. The *Oregan-R* and *w1118* strain had the highest resistance to OA in this group; similarly, *D. sech.08* and *D.sec.27* are the most resistant strains in the *D. sechellia* group. Although *D. mauritiana* appears most similar to *D. melanogaster*, it is significantly more sensitive to OA. *D. sechellia* have really low β, indicating they are largely more resistant to OA than *D. melanogaster*.

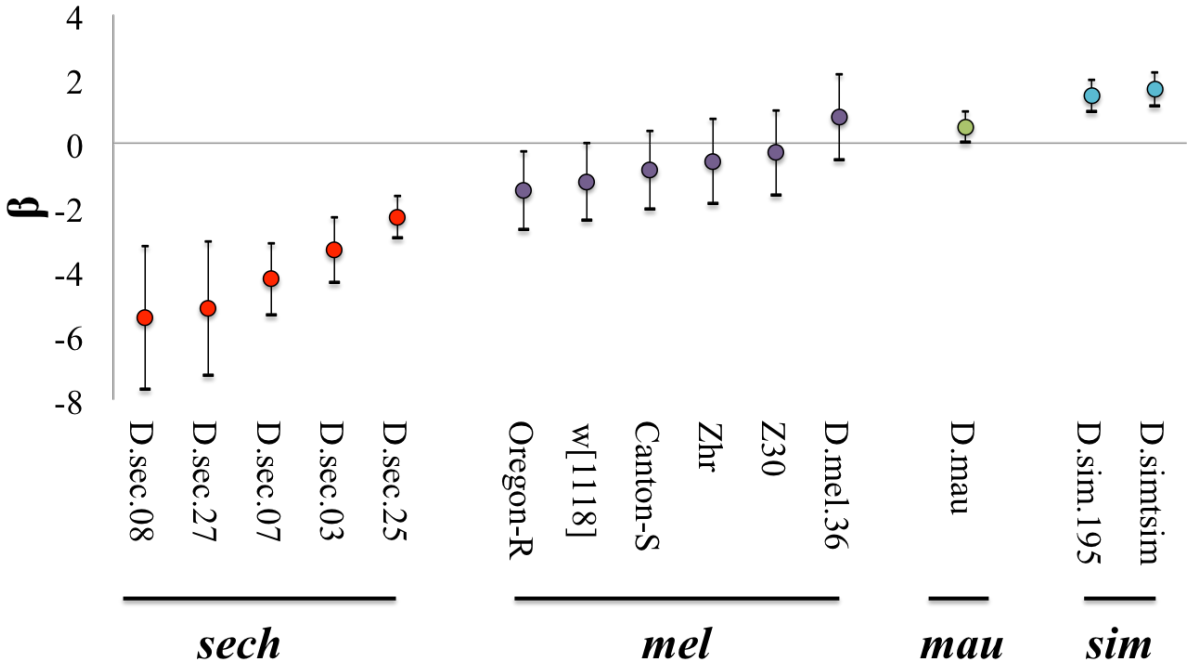
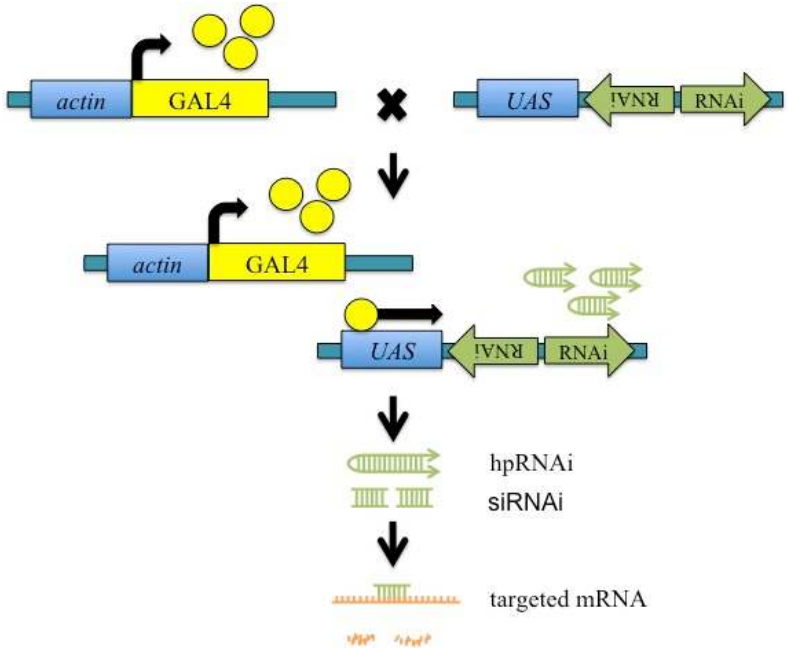


Figure 4A

(i) A schematic representation showing how the UAS/Gal4 under a *CyO* balancer works in *D. melanogaster* to knock-down each gene in the resistance locus. This system allows partial-loss-of-function by expressing an inverted-repeat hairpin RNA for RNA interference and inactivation of a specific gene's mRNA. (ii) Crossing scheme of a virgin female *actin*-GAL4/*Cyo* and male UAS-RNAi/UAS-RNAi that yields two phenotypes *actin*-GAL4/UAS-RNAi (knockdown) and UAS-RNAi/*CyO* (control).

(i)



(ii)

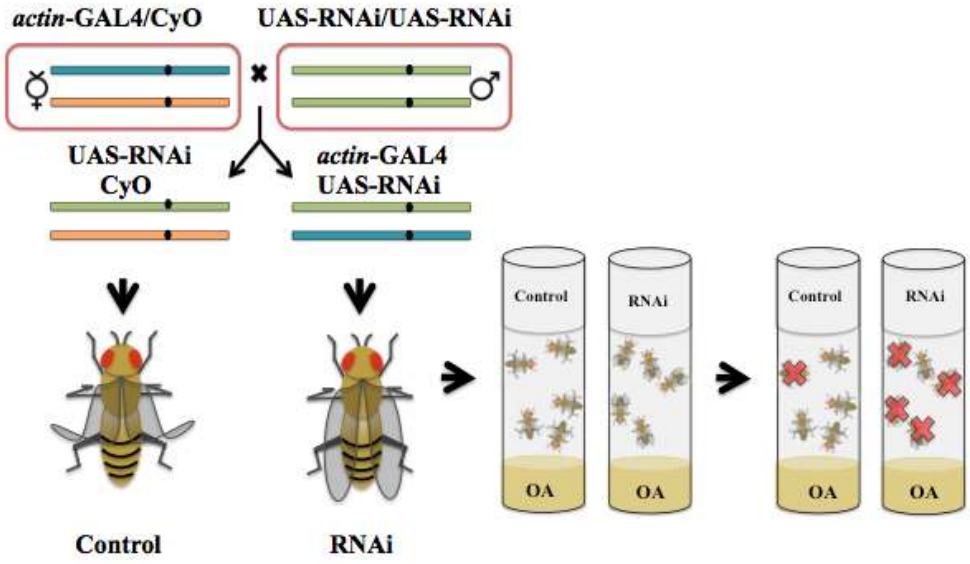


Figure 4B

Using the ubiquitous GAL4 driver to knockdown the genes in the resistance locus throughout development, knockdown of *Osi6/7* shows increases sensitivity to OA. All knockdown flies were compared against their respective UAS-RNAi/*Cyo* siblings. Error bars are represented by 2SE.

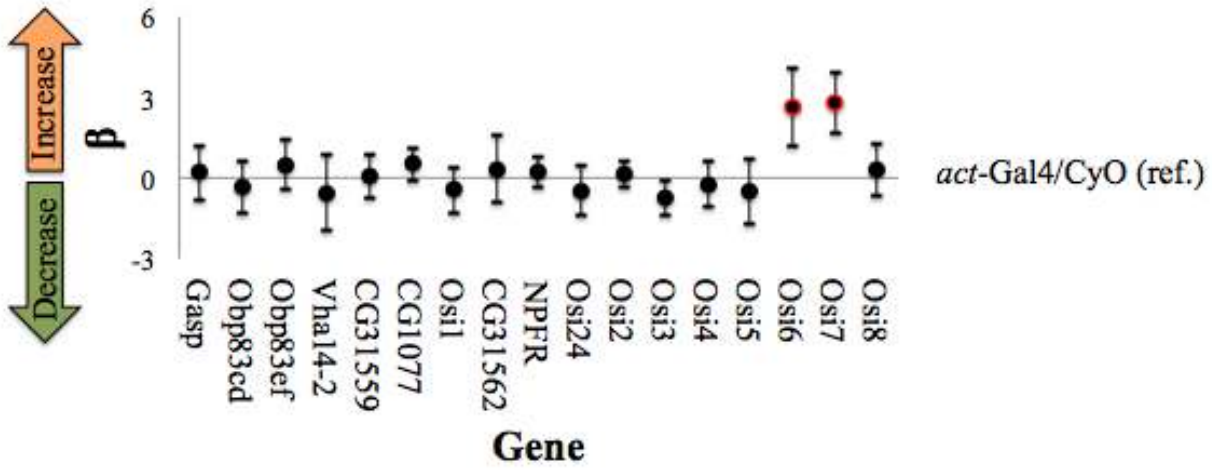
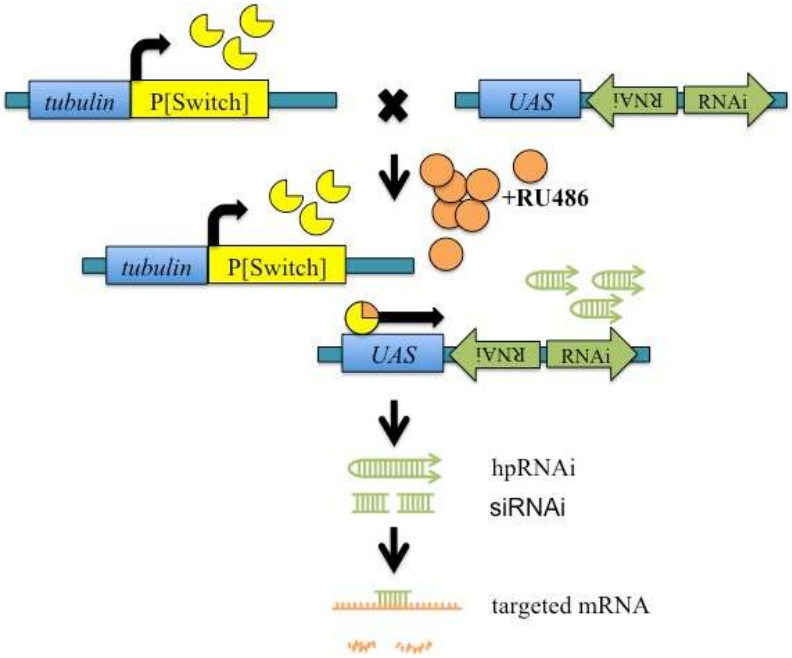


Figure 5A

(i)*Tubulin*-P[Switch] uses a modified chimeric GAL4 gene (Gene-Switch) that encodes the GAL4 DNA binding domain, the human progesterone receptor ligand-binding domain, and the activation domain form the human protein p65. The chimeric molecule only becomes active in the presence of the synthetic antiprogestin, mifepristone (RU486), and then binds to the UAS sequence to activate transcription of the RNA hairpin, knocking-down expression of that gene.
(ii) Crossing scheme for the Gene-Switch/UAS system producing only one genotype, but RU486-unexposed siblings were used as the control.

(i)



(ii)

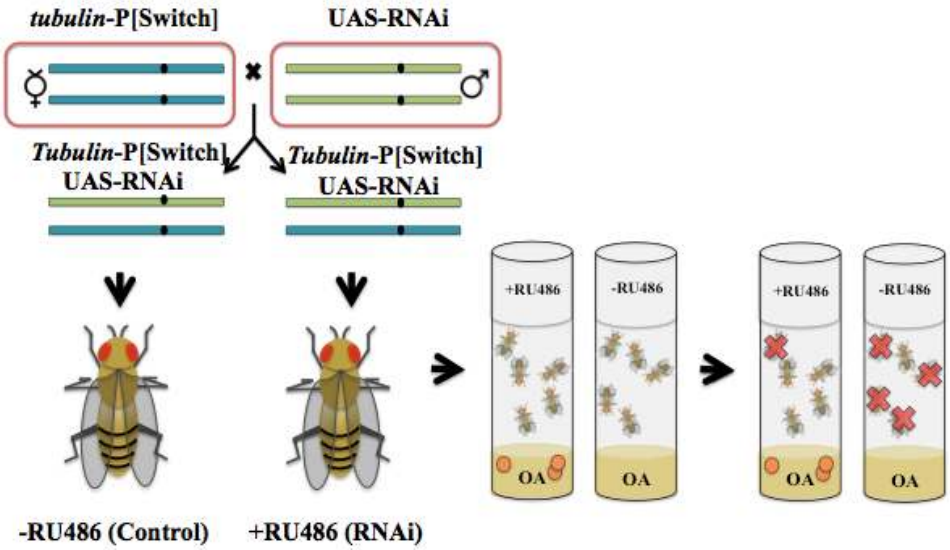


Figure 5B

Using a stage-specific GAL4 driver, activated at the adult stage when exposed to the hormone RU486 suggest that knocking-down *Osi6/8* increases sensitivity to OA. All knockdown flies were compared against their respective unexposed siblings. A *w D. melanogaster* line exposed and unexposed to the hormone was used as a control for any possible effects the drug might have one exposed individuals. Error bars are represented by 2SE.

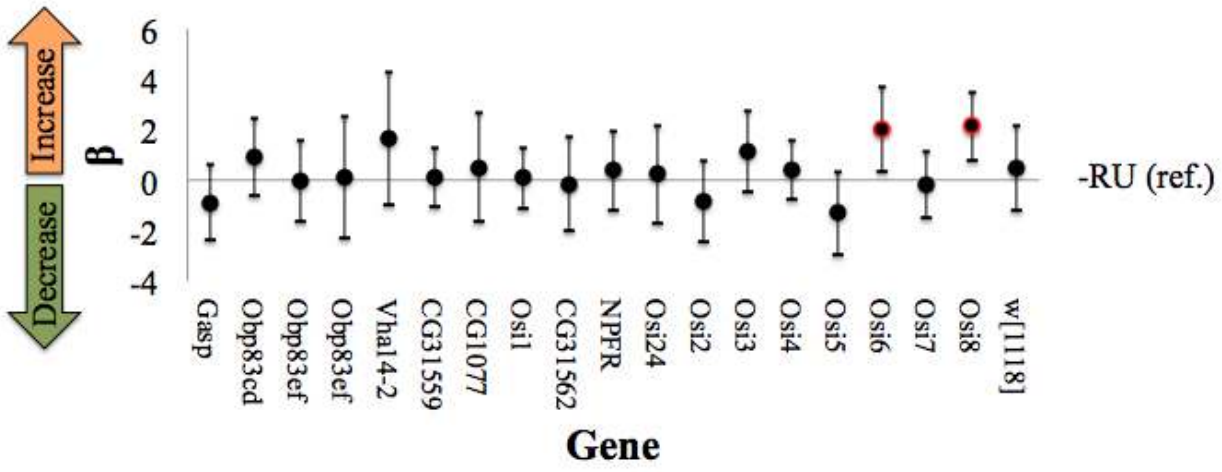


Figure 6A

Osi6/7 show parallel, high expression during the first 24hr of development, and appears to have a cyclical pattern during the beginning and end of each developmental stage. Peak expression of *Osi6/7* at 16 hours, L2 larvae stage, and 2-day post wondering pre-pupae stage (2d WPP). Little expression of *Osi6/7/8* is observed at the adult stage. *Osi8* only shows elevated expression during the pupa stage, with peak expression at the 2d WPP stage. Data is from modENCODE using *D.melanogaste*r whole flies mRNA-seq data.

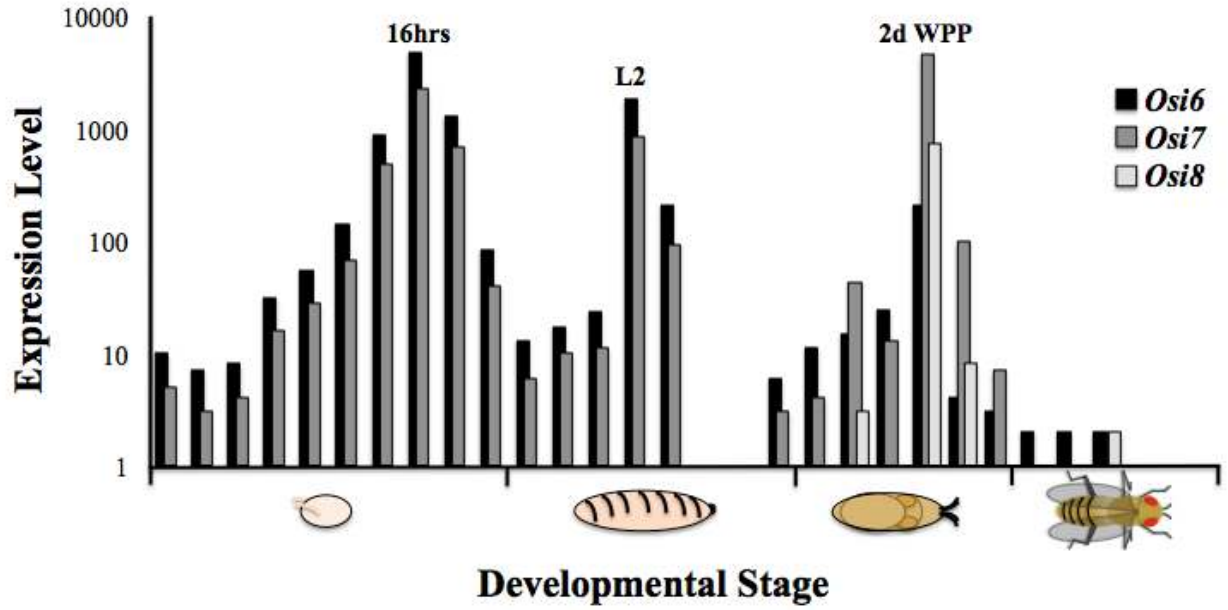


Figure 6B

Only three genes in the resistance locus show increase in mRNA expression in 4-day old flies when exposed to different environmental stressors including paraquat, a commonly used chemical found in pesticides. No data was available for *CG31562* on modENCODE. *Osi6/7* are two of three total genes in the entire region that show induced expression when treated with stressors, and expression increased with a higher paraquat dose. *Osi8* shows no induces expression for any of the treatment.

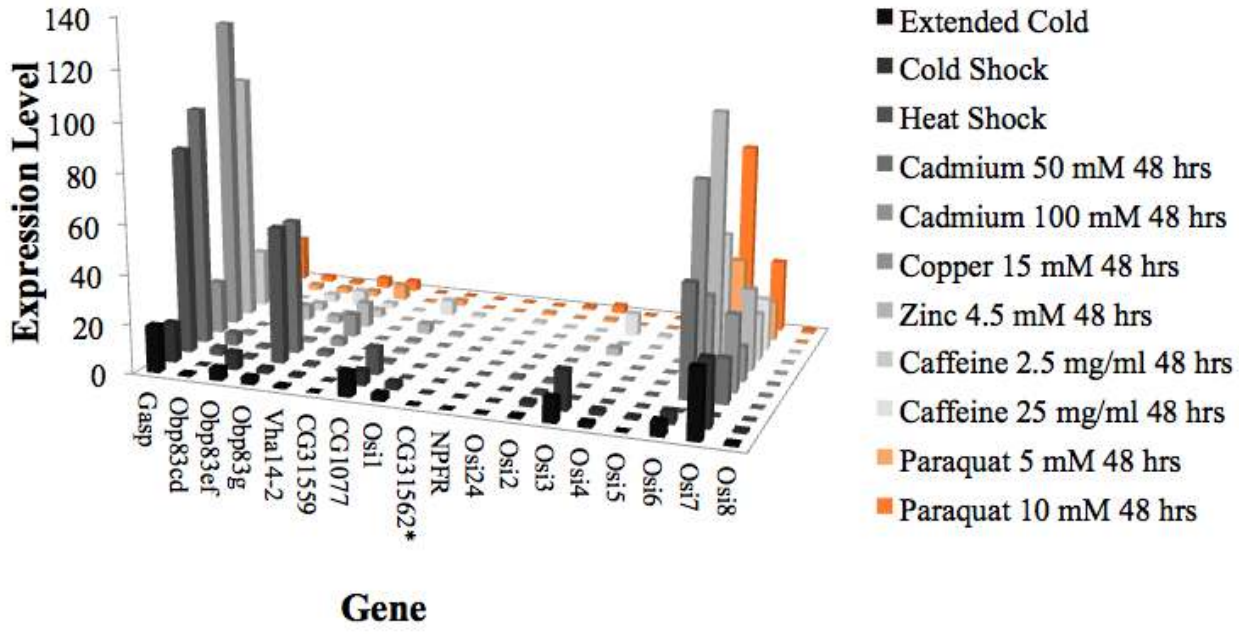


Figure 6C

Whole-bodies mRNA expression data shows *Osi6/7/8* expression primarily in the central nervous system and the fat body during the pupae stage; however, no CNS and FB data was available during the adult stage. Moderate expression of *Osi6/7* is also observed in the salivary glands and digestive system.

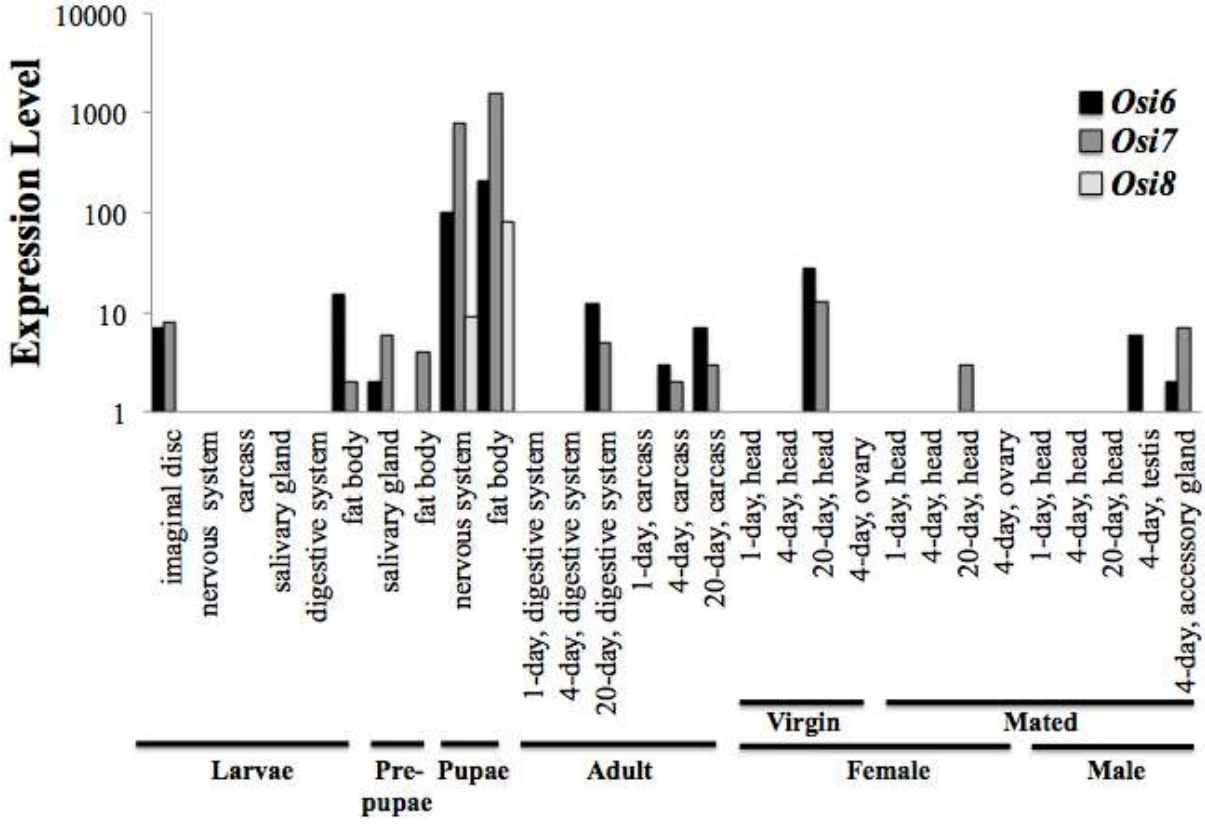


Figure 7A

Different tissue-specific GAL4 driver lines were used to knockdown *Osi6/7/8* in key tissues shown to have high expression of these genes. All knock-down flies were compared for changes in OA sensitivity against their respective GAl4 driver. Error bars are represented by 2SE. Knockdown of *Osi6/7* in fat body and salivary glands decreases sensitivity to OA. Knockdown of *Osi8* in the digestive system shows increased sensitivity to OA, and possibly an area in the salivary glands as well. The annotated "SG+FB" driver also knocks down in the malpighian tubules, trachea, dorsal head, antenna anlagen, and "Hgut" also in the ureter, malpighian tubules, protocerebrum.

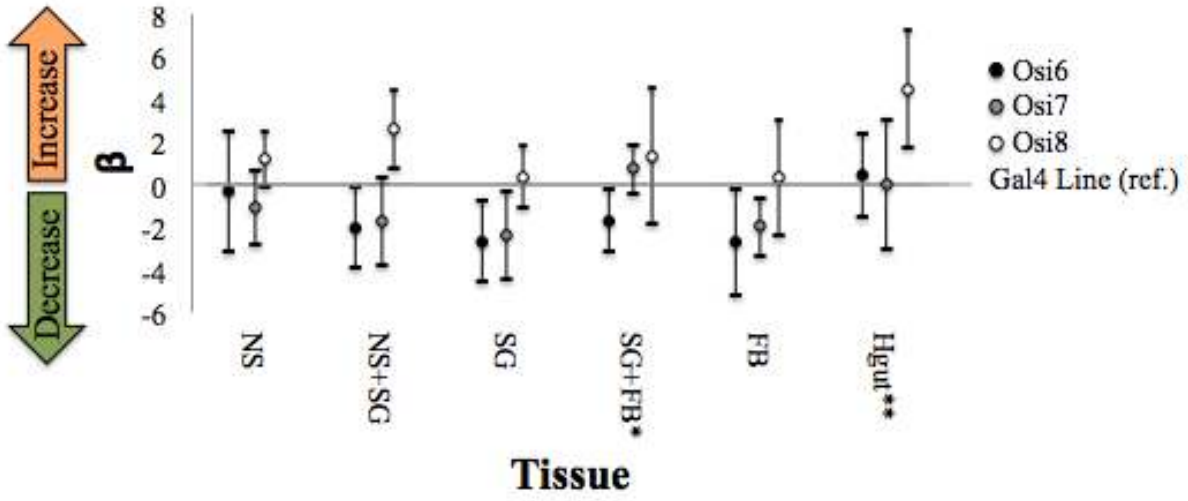


Figure 7B

Coefficients for key metabolism genes (UAS-RNAi) crossed with the *tubulin*-P[Switch] GAL4 driver known to be differentially expressed between *D. simulans* and *D. sechellia*. Knockdown was induced with RU486 at the 3-day adult stage and unexposed siblings were used as the control for each gene; no change in sensitivity was seen.

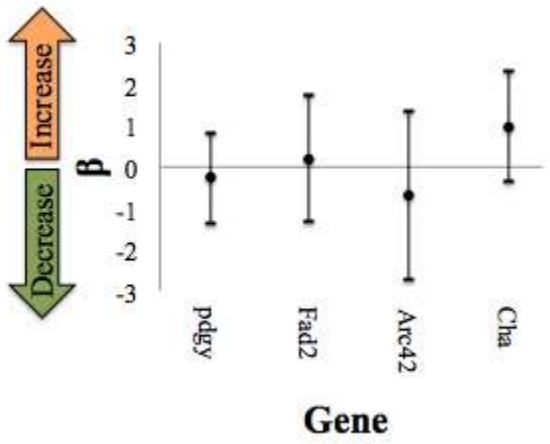


Figure 8A

To identify possible sequence differences in *D. sechellia Osi6* and *Osi7* that could have functional consequences and therefore affect OA resistance I aligned the coding sequence of the *D. sechellia* and its distantly related sister species. (i) For *Osi7*, there is a single derived nonsynonymous change (S245L) in the *D. sechellia* ortholog. This change alters an amino acid that is conserved among all other species of *Drosophila* surveyed spanning 40 million years of divergence. (ii) However, after Sanger sequencing of the other laboratory lines this nonsynonymous change in *Osi7* was not fixed in *D. sechellia*.

Figure 8B

Orthologs of *Osi6* revealed 7 derived synonymous and no derived nonsynonymous changes on the *D. sechellia* lineage. *Osi7* showed 9 synonymous and no nonsynonymous changes (with the exception of the one coding change in the *D. sechellia.25* line). *Osi8* had 4 synonymous and 2 nonsynonymous changes in *D. sechellia.*

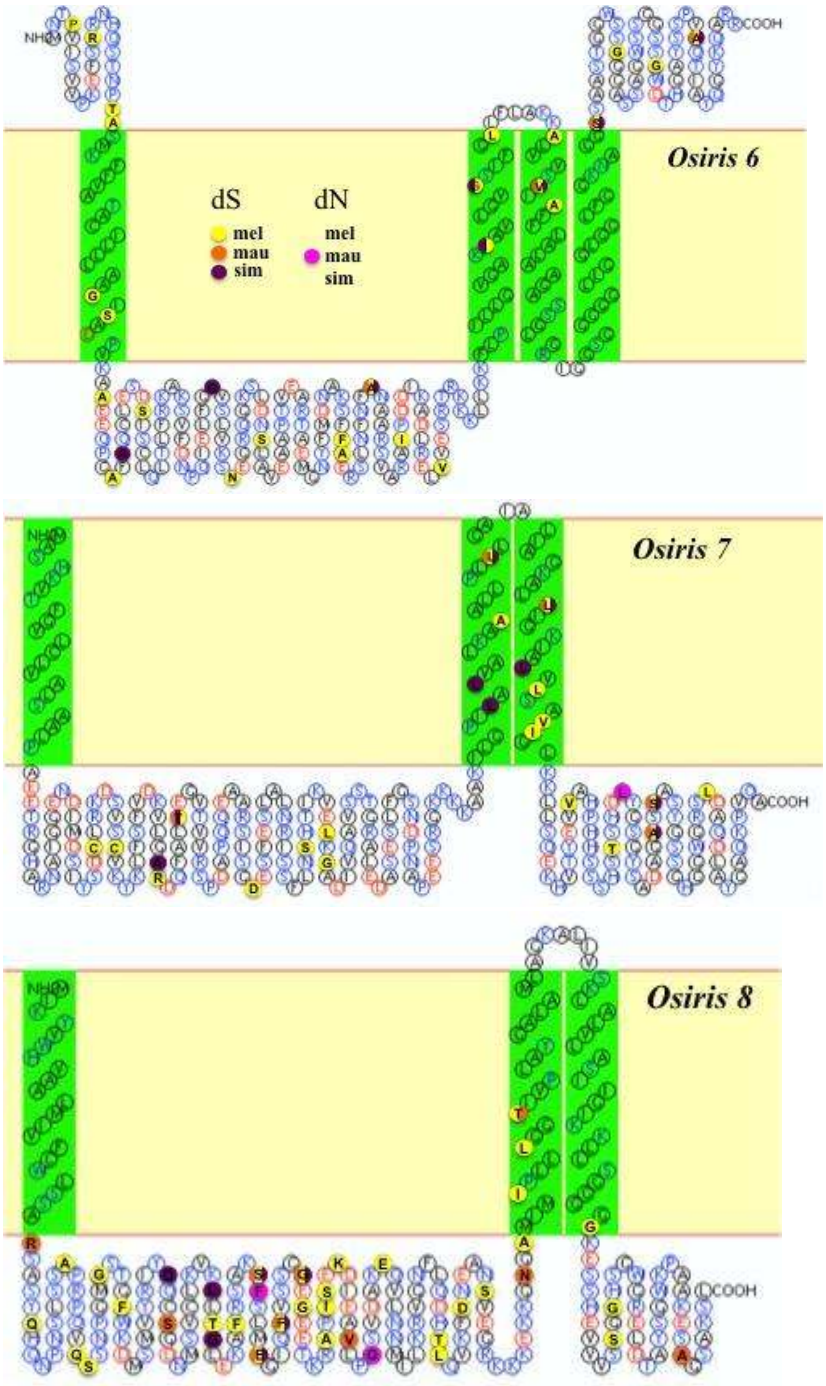


Figure 9

RNA-seq data from 7-10 day adult female flies showing changes in gene expression between *Drosophila* species. Among the 18 genes in the resistance locus, only *Osi6*/*7* show derived, reduced expression in *D. sechellia* relative to both *D. simulans* and *D. melanogaster*. *Osi8* expression is too low to make any comparisons about expression difference between species.

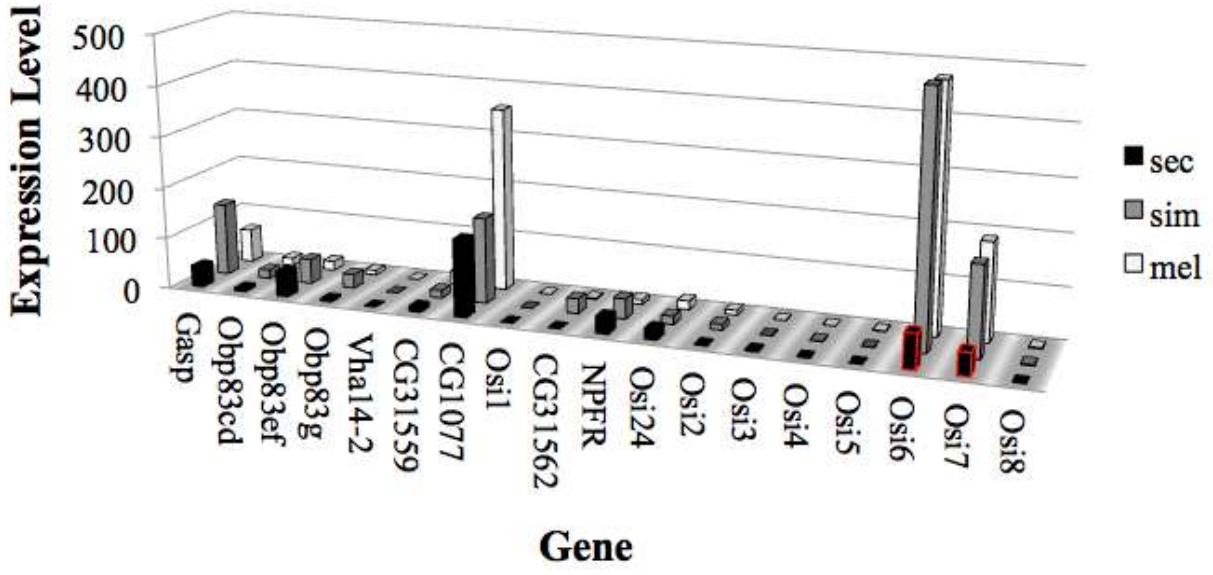


Figure S1

Schematic representation of the ripening stages of *M. citrifolia*. Levels of octanoic acid (OA) and hexanoic acid (HA), two of the predominant toxins in morinda fruit, peak in concentration during the ripe stage of the fruit. OA and HA levels start to diminish during the rotting stage, where ethanol levels increase. *D. sechellia's* resistance to high levels of OA allow it to inhabit the fruit during peak toxicity, whereas its sister species are restricted to utilizing the food source only during the rotting stage.

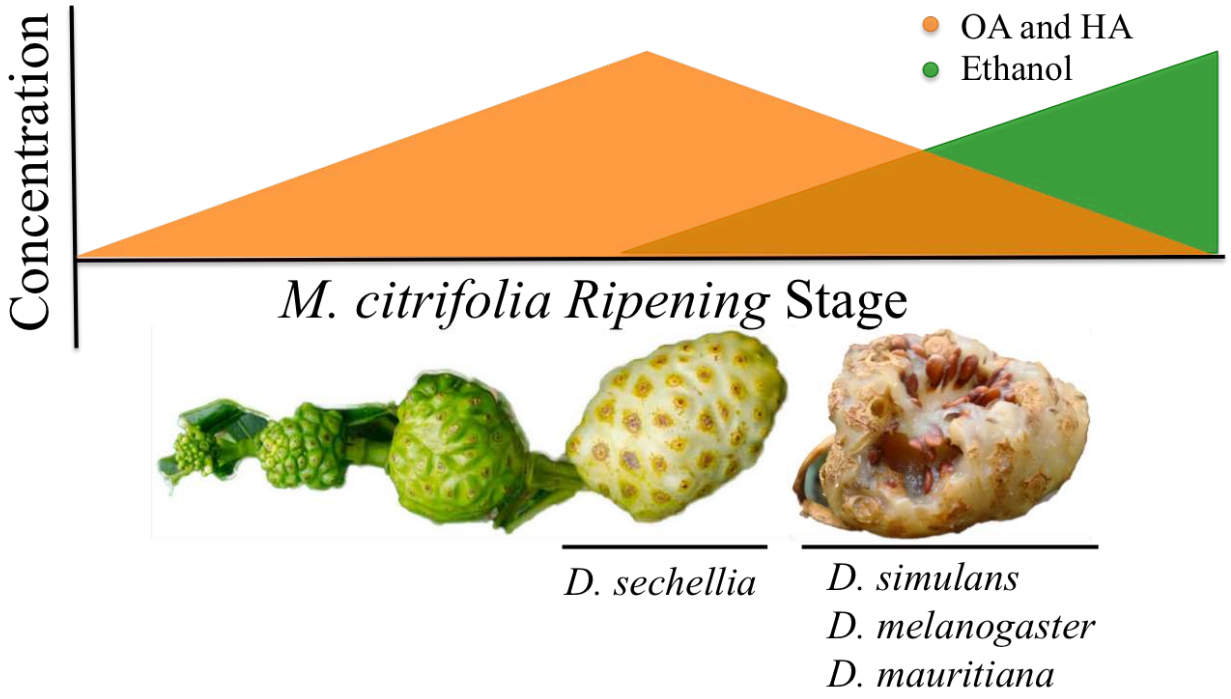


Figure S2

Electrophoresis gel showing the PCR product of reverse-transcribed RNA from the *Osi6* mutant and *Osi6* UAS-RNAi lines at different PCR cycles. Banding pattern shows the presence of *Osi6* mRNA for both the mutant and UAS line, and product size increases with the number of PCR cycles.

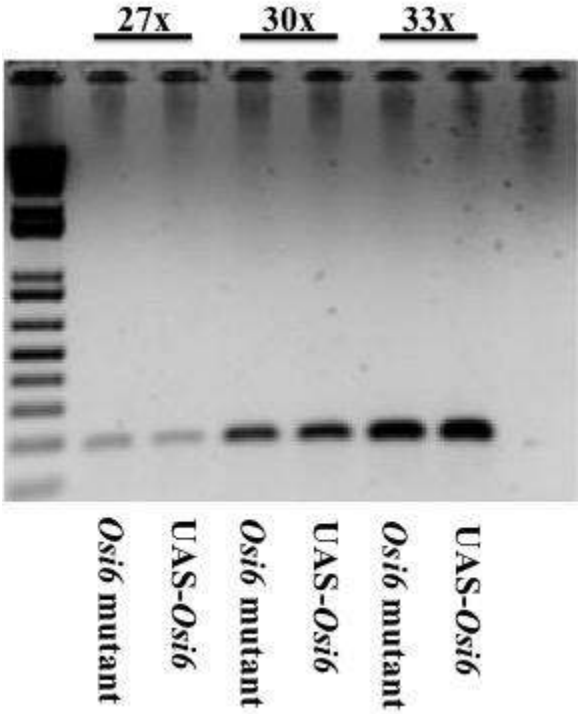


Table S1

Table showing the sample size for all *actin*-GAL4, *Tubulin*-P[Switch], and species crosses used in the mixed effect Cox regression model. The UAS/GAL4 cross for *Osi6* produces few knocked-down flies, and the 30 individuals used in the model were a result of more than 20 different crosses.

actin-GAL4	Gasp	Obp83cd	Obp83ef	Vha14-2	CG31559	CG1077	Osi1	CG31562	NPFR	Osi24	Osi2	Osi3	Osi4	Osi5	Osi6	Osi7	Osi8
UAS-RNAi	118	160	110	227	200	120	120	210	150	220	80	116	120	150	120	220	209
Gal4xUAS	100	258	30	210	200	200	170	200	260	198	190	190	258	180	30	103	230
Gal4xUAScyo	160	240	135	187	110	200	142	200	290	210	163	170	260	180	185	339	143

tubulin-P[Switch]	Gasp	Obp83cd	Obp83ef	Vha14-2	CG31559	CG1077	Osi1	CG31562	NPFR	Osi24	Osi2	Osi3	Osi4	Osi5	Osi6	Osi7	Osi8	w
Gal4xUAS(H)	90	146	120	90	90	99	170	140	91	155	70	101	149	148	129	110	86	88
Gal4xUAS	120	141	106	100	90	109	148	130	90	128	90	94	128	137	119	110	70	89

Species	D.sec.08	D.sec.27	D.sec.07	D.sec.03	D.sec.25	Oregan-R	w1118	Canton-S	Zhr	Z30	D.mel.36	D.mau	D.sim.195	D.simtsim	60000
n	80	80	137	80	130	510	729	588	160	160	159	277	223	210	296