The effect of copy number variation at the *Cyp12d1* gene on caffeine toxicity in *Drosophila melanogaster*

By

Dianarys Hernandez-Aquino

B.S., Universidad de Puerto Rico-Arecibo, 2014

Submitted to the graduate degree program in Molecular, Cellular, and Developmental Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

_____

Chair: Stuart J. Macdonald

_____

Justin P. Blumenstiel

_____

Robert L Unckless

Date Defended: 11 December 2020

The thesis committee for Dianarys Hernandez-Aquino certifies that
this is the approved version of the following thesis:

The effect of copy number variation at the *Cyp12d1* gene on caffeine toxicity in *Drosophila melanogaster*

_____

Chair: Stuart J. Macdonald

Date Approved: 18 December 2020

Abstract

Xenobiotic compounds are a constant challenge for animals. We desire to understand the mechanisms behind xenobiotic resistance by identifying loci underlying variation in the toxicity of such compounds. Previous work in *D. melanogaster* mapped several loci contributing to resistance to caffeine, a model xenobiotic compound, one of which implicated the cytochrome P450 gene *Cyp12d1*. Both through RNAi knockdown and RNAseq, *Cyp12d1* involvement in caffeine resistance was detected, moreover, an association between greater *Cyp12d1* copy number and increased caffeine resistance was seen. To further test this, we used CRISPR-Cas9 editing to generate putative loss-of-function mutations in exon 2 of *Cyp12d1*, confirming via RNAseq that mutation-carrying strains have significantly lower *Cyp12d1* expression than our single control strain. We subsequently measured caffeine resistance in all strains, observing a significant reduction in resistance in 2 out of 3 CRISPR mutant strains relative to the control strain. Interestingly, the third mutant strain exhibited higher resistance to caffeine than the control. We present evidence suggesting that this result is due to a difference in splice site mutations. This makes it challenging to compare directly among strains. To overcome this, and to explicitly enable a test of the effect of varying *Cyp12d1* copy number, we next edited a strain containing two, nearly identical copies of *Cyp12d1*. Creating a strain with one perfect copy of *Cyp12d1* while preserving the genetic background of the original genotype, as well as a control non-mutated strain with two copies of *Cyp12d1*. We observed a trend, but no significant difference in resistance was detected between the lines. Using these CRISPR lines we then performed RNAseq on guts of female flies to measure RNA levels under naïve and caffeinated conditions. To which, we were able to see a significant increase in *Cyp12d1* RNA levels correlated to an increase in gene copy number. All

together, we confirmed the involvement of *Cyp12d1* in caffeine resistance, but more work is needed to determine the role of copy number variation at *Cyp12d1*.

Table of Contents

List of Tables

List of Figures

Introduction

Insects are constantly challenged by xenobiotic compounds in nature, including chemicals produced by plants as a defense against herbivory, or by the wide variety of synthetic pesticides produced by humans. To metabolize these toxic xenobiotic compounds, insect genomes harbor a large number of detoxification genes. The genes involved in detoxification are divided into three different classes. For the phase I class, the xenobiotics' toxicity is decreased by the modifying action of cytochrome P450 monooxygenases (P450s). This superfamily of enzymes modifies the toxic substrate through the introduction of reactive and polar groups by either oxidation, hydroxylation or reduction (Chahine & O'Donnell, 2011). For the phase II class, the enzymes UDP-glucuronosyltransferases (UGTs) and glutathione-S-transferases (GSTs) conjugate large, polar side chains such as glutathione, sulphate or glucuronate to the toxins, making them less active and easier to transport across membranes (Chahine & O'Donnell, 2011). Lastly, for phase III class proteins, a membrane-bound ATP-binding cassette (ABC) transporter excretes the final product from the cell (Li et al., 2007a). Due to the constant challenge of toxins, genetic variants within populations in the detoxification cascade have been the key element that allow organisms to adapt to different environments and any sudden environmental change.

Among these variants are the well-studied single nucleotide polymorphisms (SNPs) which have been thought to be the major source of genetic diversity within populations (Mackay et al., 2012;International HapMap Consortium et al., 2007). Specifically, SNPs can cause coding mutations in genes encoding proteins involved in the detoxification cascade that can impart resistance (Li et al., 2007b; Weetman, 2018). However, sequence amplifications and differences in the copy number of genes pertaining to the detoxification cascade have made copy number variants (CNV) the main topic of study for several research groups (Faucon et al., 2015; Li et al.,

2007b; McDonnell et al., 2012). Moreover, numerous reports have shown the widespread presence of CNVs in both model organisms and humans (Chakraborty et al., 2018; Iafrate et al., 2004; Emerson et al., 2008; Levy et al., 2007; Rees et al., 2014; Sebat et al., 2004; Zarrei et al., 2015). The duplications or deletions usually involve ~1 kbp or more of identical or nearly identical sequences (Freeman et al., 2006). For the most part exons appear to be the ones affected by CNVs but full-length protein coding genes can also be subject to these structural variants (Chakraborty et al., 2018). Due to the nature of these variants and the fact that most of the techniques used to genome sequence organisms use short reads which are aligned to a reference genome, it is difficult to identify and resolve them, because the reads can get misplaced as they do not expand the entirety of the duplicated genes region in one whole read (Ku et al., 2010; Metzker, 2010).

CNVs are important in evolutionary adaptation. CNVs can segregate for functional variants that affect the efficacy of detoxification. Several studies have shown that individuals with higher copy numbers of particular detoxification genes have an increased resistance to certain xenobiotic compounds (Faucon et al., 2015; Remnant et al., 2013; Schmidt et al., 2010; Zhang et al., 2015). For example, *Spodoptera liturais* is a widely spread herbivorous insect that can attack over 100 plant species, it was discovered to possess copy number differences in multiple genes within all of the stages of the detoxification cascade. Researchers identified 34 cytochrome P450 genes, 12 GSTs genes, and 11 ABC transporter genes all with variation in copy number (Gong et al., 2019). It is believed that these CNV differences play a crucial role in the insect's ability to adapt to its diverse environment, as well as provide with resistance to insecticides (Gong et al., 2019). Another example of adaptation is the *Aedes aegypti* mosquito or better known as the dengue mosquito. This mosquito carries the dengue virus which affects around 3 million people a year. The mosquito shows different levels of resistance to the insecticide deltamethrin. When studied, 41 gene

amplifications, almost all affecting cytochrome P450s, were associated with resistance to deltamethrin (Faucon et al., 2015; Faucon et al., 2017). Finally, CNV variation is important in *Bombyx mori* the domesticated silkworm used to produce silk commercially and also utilized as a model of Lepidoptera insects. A study with the goal of detecting CNV within the species found genes not only associated with detoxification but also reproduction, immunity, and signal recognition, thus providing evidence of adaptative evolution (Zhao et al., 2014).

Work from the Macdonald lab suggests that CNV at the known detoxification gene *Cyp12d1* is associated with xenobiotic resistance in *D. melanogaster* (Najarro et al., 2015). The lab phenotyped over 1700 lines of the *Drosophila* Synthetic Population Resource (DSPR) (King et al., 2012), a series of recombinant inbred lines (RILs) derived from a multiparent intercross, to map several quantitative trait loci (QTL) contributing to resistance to caffeine, a model xenobiotic compound. Q2 is the QTL with the largest-effect, accounting for 14.4% and 5.7% of the broad-sense heritability for caffeine resistance in the DSPR pA and pB populations, respectively. Under it the gene *Cyp12d1,* which is subject to CNV within natural populations of *D. melanogaster,* was identified (Najarro et al., 2015). Ubiquitous RNAi knockdown for the two gene copies *d* and *p* of the gene *Cyp12d1* showed a markedly reduced resistance to caffeine when compared to controls in four out of five trials. Moreover, the data shows that greater caffeine resistance is significantly associated with more copies of the *Cyp12d1* gene in the two panels of DSPR lines ($p < 0.0001$), with a similar trend observed in a second set of inbred lines, the DGRP (Mackay et al., 2012) ($p = 0.065$). Furthermore, when rescanning the genome after statistically adjusting for the effect of the *Cyp12d1* CNV in the DSPR, the signal observed at Q2 disappears in both panels, suggesting that the CNV of *Cyp12d1* or a variant with a strong linkage disequilibrium (LD) with the event has a key role in caffeine resistance (Najarro et al., 2015). These data make *Cyp12d1* an ideal candidate

to further study the importance of CNV in xenobiotic resistance. In this dissertation, we hypothesize there is a causative, positive relationship between *Cyp12d1* copy number and caffeine resistance. To test this, we used CRISPR-Cas9 editing to generate putative loss-of-function mutations in addition to changing copy number (CN) in *Cyp12d1*, used RNAseq in whole body as well as gut- specific and measured caffeine resistance on multiple strains in order to test our hypothesis.

Materials and Methods

Caffeine Resistance Assay

Experimental female flies were raised and assayed at 25°C, 50% humidity on a 12h:12h light:dark cycle. One day prior to the start of the assay, female flies were collected under CO_2 and put-on standard cornmeal-yeast-molasses food in narrow *Drosophila* culture vials in groups of 10-20 flies. Flies were allowed to recover from CO_2 exposure overnight. Also, one day prior to the start of the assay we made cornmeal-yeast-dextrose media supplemented with caffeine (Sigma Aldrich, C0750), poured the media into petri dishes, and after it had solidified, pushed polycarbonate *Drosophila* Activity Monitor tubes (5mm diameter x 65mm length, TriKinetics, Inc., PPT5x65) into the media to a depth of approximately 10mm. To prevent desiccation of the media, the bottom of each activity monitor tube was dipped in molten parafin wax to seal it. On the day of the experiment 2-4-day old test females were individually loaded into media-filled activity monitor tubes under CO_2, and the tubes plugged with a small piece of foam. Tubes were then inserted into *Drosophila* Activity Monitors (TriKinetics, Inc., DAM2) and the activity of each fly was automatically recorded every minute for the duration of the experiment. Once the monitors

were taken down, we analyzed the activity data using custom code written in R (R Core Team, 2013) and report the lifespan of each individual fly in the experiment.

CRISPR-Cas9 based genome editing

We used the Wisconsin group online target finder (http://targetfinder.flycrispr.neuro.brown.edu/) (Gratz et al., 2014) to determine the best guide RNA (gRNA) target in the *Cyp12d1* gene using the *Drosophila melanogaster* reference genome sequence (Release 6,(Adams et al., 2000; Thurmond et al., 2019). At the beginning of exon 2 we identified TGAGGGCTTTCATGCCCGGTGG to be the best target sequence for the gRNA, as it was one of the sites closest to the start of the gene and it showed no off-target sites. Chromosomes can harbor two copies of *Cyp12d1*, and the two copies of the gene in the reference *D. melanogaster* strain – *Cyp12d1-d* and *Cyp12d1-p* – are targeted equivalently by this gRNA. Once identified we obtained sense and antisense oligos (sense, [Phos]CTTCGATGAGGGCTTTCATGCCCGG; antisense, [Phos]AAACCCGGGCATGAAAGCCCTCATC) to hybridize and clone into the vector. We constructed the gRNA using the pU6-BbsI-gRNA plasmid following the protocol provided at http://flycrispr.molbio.wisc.edu/protocols/gRNA. The transformations were confirmed by Sanger sequencing using a T7 primer.

The plasmid was injected into 300 embryos from the Bloomington *Drosophila* Stock Center (BDSC) fly stock 55821 (y[1] M{vas-Cas9.RFP-}ZH-2A w[1118]) by BestGene (http://www.thebestgene.com/CRISPRInfoPage.do). Prior to injection we confirmed the presence of the gRNA target site in this strain by PCR using the primers 5'-GCGATCTGTGGCGATTTACG-3' and 5'-CATTCTCAAGCCCCTGGGTT-3', additionally, we verified the CNV status of the strain using the PCR protocol described in Schmidt et al. (2010).

The conditions of the PCR cycler were as follows: 95°C for 2 min, 35 cycles of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec, with a final 2 min extension at 72°C. Once the G0 animals emerged (78 emerged), they were mated with 2 animals of the opposite sex from the balancer strain A4;Kr[If-1]/CyO; D[1]/TM3, Ser[1]. This balancer strain was created using the X chromosome of DSPR founder line A4 and the second and third chromosomes from the BDSC strain 7198 (Chakraborty et al., 2018). Once F1 animals emerged, curly-winged animals (with a second chromosome genotype of 55821/CyO) were collected in order to be mated to 2 animals of the opposite sex from the balancer strain A4;Kr[If-1]/CyO; D[1]/TM3, Ser[1]. Once eggs were laid, the F1 animals were collected to extract DNA, which was executed using the Gentra Puregene Cell Kit (Qiagen, 158388). We then used the gRNA target site PCR to amplify the region surrounding the target sequence and then sent for sequencing using the primer 5'-GCGATCTGTGGCGATTTACG-3'. Cross vials initiated with F1 animals that had a mutation as well as a few non-mutant lines were kept, and further standard crossing was done using the strain A4; CyO/Kr[If-1]; A4 to obtain homozygous strains. This balancer strain was again derived from DSPR founder A4 and BDSC stock 7198. This genotype of the final edited and control strains was A4; 55821; A4. We established two strains with the same 2-bp deletion, one with a 2-bp deletion and 1-bp substitution, and a control line that had been passed through an identical series of crosses but did not result in a CRISPR/Cas9-induced mutation. The three mutations are both predicted to cause lead to a premature STOP codon in the second exon of *Cyp12d1* and create a product 61 amino acids long (as opposed to the 521 full length wildtype product).

CRISPR-Cas9 based genome editing using a homozygous duplicated *Cyp12d1* strain

Given the segregating variations in the 55821 line, during this second CRISPR experiment we set out to develop a set of lines that were isogenic for the targeted second chromosome that contains *Cyp12d1*. We first generated an injection strain with the genotype *vasa*-Cas9; A7 using standard fly crosses with balancer chromosomes. This line has a transgenically-expressed source of Cas9 on the X chromosome, and it has the second chromosome from the A7 DSPR founder, which is homozygous duplicated for *Cyp12d1*. Since the two copies of *Cyp12d1* are nearly identical, and our gRNA (see "CRISPR-Cas9 based genome editing" section above) targets both copies, our hope was to identify mutant animals where Cas9 had cut both copies, and the two cut ends had been joined back together following deletion of the intervening segment, resulting in a line with a single, chimeric copy of *Cyp12d1*. To create these lines, we followed the same protocol and used the same gRNA previously discussed in "CRISPR-Cas9 based genome editing". We sent our injection strain and gRNA plasmid to Genetivision who injected 400 embryos with 500ng/ul of the guide RNA. Fifty-seven G0 animals emerged, and each was mated with two animals of the opposite sex from the balancer strain A4; CyO/Kr[If-1]; A4. Once F1 animals emerged we followed the same protocol described in the section CRISPR-Cas9 based genome editing, with the exception that the F1 animals were mated to A4; CyO/Kr[If-1]; A4 instead of the strain A4; CyO/Kr[If-1]; A4. We were able to create 4 different lines that possess the genotype of A4; A7; A4. We created a control line where it has a duplicated WT genotype A4; A7[2 copies]; A4. As well, we created 3 types of mutants, in all of them we create a chimera between the two copies present in the strain, meaning the *Cyp12d1* copy present has the 5' end of the "p" copy and the 3' end of "d" copy. Two of the mutants have a base deletion mutation which leads to a premature

STOP codon on the second exon and the last strain created has a "seamless" union between the two copies of *Cyp12d1* referred as to A4; A7[1 copy]; A4.
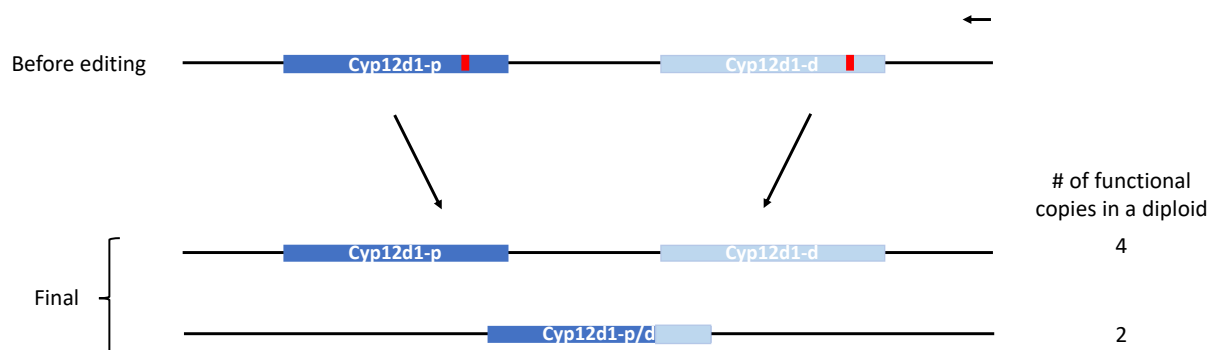

Figure 1 CNV edit mediated by CRISPR-Cas9

Here is a diagram of the CRISPR-Cas9 edit created in the A4; A7; A4 lines. On the top we have the original version of the *Cyp12d1* A7 genotype showing both the p and d copies, red boxes point to the gRNAs target site and the arrow on the top shows' direction of transcription. The bottom part of the diagram presents what was accomplished by the CRISPR-Cas9 editing. The top strand in the "Final" section shows the A4; A7 [2 copies]; A4 line where no CRISPR mutation is present. On the second strand we show the genotype of the line A4; A7 [1 copy]; A4, were we show where the seamless union between the copies p and d is made. Finally, we show the number of copies present in a diploid organism.

RNAseq of Gut Tissue from Animals with Different *Cyp12d1* Copy Number

In order to test for differences in gene expression levels between the CRISPR A4; A7; A4 animals under control and caffeine conditions, especially for the gene of interest *Cyp12d1*, we measured RNA expression in female gut tissue.

We reared the CRISPR A4; A7; A4 single and duplicated strains and collected males and virgin females from both. We set up multiple replicates of 10 female by 5 male crosses of the A4; A7; A4 single and duplicated strains, as well as a cross between CRISPR A4; A7; A4 duplicated females to A4; A7; A4 single males, creating offspring with all the genotypes needed for testing. We cleared vials once experimental flies started emerging to control the age of the flies, and two days after clearing we tipped the flies into new food in order to prepare for collection the next day.

We then collected 10 1–3-day old mated female flies per cross under CO_2 anesthesia. Flies were allowed to recuperate overnight, thus making the final experimental animals 2-4-days old on the day of the experiment. We then exposed flies to control and 1.5% caffeine food for 4-hours, transferred flies to 1.7ml microcentrifuge tubes and put-on ice in order to anesthetize them. Gut dissection was performed on 5 female flies per sample, including the midgut, hindgut, and malpighian tubules, but not the crop. Tissues of all 5 females were pooled into a TRIzol-containing (ThermoFisher Scientific, 15596018) screwtop tube and moved to the -80 freezer.

Experimental animals of the same genotype were exposed to caffeine/control conditions and were reared using the same settings as described in the section "Caffeine Resistance Assay", but in groups in vials rather than singly in activity monitor tubes. To ensure that animals were exposed for exactly 4 hours, and to allow time for dissection, the experiment was carried out over three consecutive days. Each day included one replicate sample of each genotype/treatment combination, and within each day we had 3 blocks of time that were separated by 1 hour, with each block including the control/exposed pair from a single genotype. The order in which genotypes were dissected was rotated over days in order to average over any differences in the time of dissection.

We extracted RNA from each of the 18 samples following the manufacturer's protocol for the Direct-zol RNA Miniprep kit (Zymo Research, R2050). Subsequently, we generated Illumina TruSeq stranded mRNA libraries, and sequenced on an Illumina NextSeq 550, collecting paired-end 37-bp reads (KU Genome Sequencing Core). Quality control and trimming of reads was accomplished using the fastp tool (Chen et al., 2018). We then utilized Kallisto (v0.46.1,(Bray et al., 2016)) for the pseudoalignment of reads from each sample to the *D. melanogaster* reference transcriptome (BDGP6.22). The alignment index was modified by eliminating one of the two,

nearly-identical copies of *Cyp12d1* that are present in the "iso-1" reference genome (Adams et al. 2000). We did this so that all of the reads from either copy of *Cyp12d1* will map to the same gene, allowing us to more accurately assess *Cyp12d1* expression given we know our three test genotypes have different copy number. We confirmed that it made no difference which copy was retained for analysis, moreover, a separate analysis with both copies present was completed, and the results found were approximately the same. We then made use of the program Sleuth (Pimentel et al., 2017) in R (R Core Team (2020)) to test for differential expression between the lines using the additive effects of *Cyp12d1* copy number (2, 3 or 4 copies) as the explanatory variable, between conditions and within conditions.

Comparing expression levels of *Cyp12d1* in mutant CRISPR strains using whole body RNAseq.

To measure differences in expression between the different genotypes, as well as determine if we had indeed knocked down expression of *Cyp12d1* we performed single-read RNAseq on the control and mutant lines pertaining to the A4; 55821; A4 genotype. To do this we collected 2-4-day old mated females from each strain, using 10 females from each of 2 replicate vials per strain. All flies were collected using CO_2 under naïve, caffeine-free conditions. Each set of 10 flies were then transferred to standard cornmeal-yeast-molasses food in narrow *Drosophila* culture vials and kept at 25°C, 50% humidity on a 12h:12h light:dark cycle for 24h. Then they were pooled into a TRIzol containing (ThermoFisher Scientific, 15596018) screwtop tube that was immediately dropped into liquid nitrogen then moved to a -80 freezer. We then extracted total RNA using the Direct-zol RNA MiniPrep kit (Zymo Research, R2050) following the manufacturer's protocol.

Next, we created an Illumina TruSeq mRNAseq library for each sample, pooled, and sequenced using the Illumina NextSeq HO-SR75 (KU Genome Sequencing Core).

Next, we trimmed and did quality control of reads using fastp (Chen et al., 2018). We once again used Kallisto (v0.46.1,(Bray et al., 2016)) for the pseudoalignment of reads from each sample to the modified *D. melanogaster* reference transcriptome (BDGP6.22). We then made use of the program Sleuth (Pimentel et al., 2017) in R (R Core Team (2020)) to test whether the CRISPR mutations had an effect on the expression of *Cyp12d1*.

GAL4-UAS gut RNAi

We used the Gal4/UAS RNAi system to knock down the expression of our target gene *Cyp12d1* both ubiquitously and in a tissue-specific manner. Female virgins from UAS-RNAi lines were crossed to males from Gal4 driver lines, and the female F1 Gal4-UAS-RNAi progeny of these crosses were tested for caffeine resistance. We employed the monitor system as described above, with 1% caffeine, and 1-4 day old test animals. We did separate subsets of the genotypes tested and completed a total of 4 experiments in order to test all of the lines. We used 2-3 vials per genotype per experiment for a grand total of ~192 tested animals per genotype (11-16 flies per vial). The lines we used are described in Table 1:

Table 1 GAL4-UAS RNAi lines

Type	Strain ID	Genotype	Description	Vendor	URL	Reference
Gal4	c42-Gal4	GAL4c42	Specific to principal cells of malpighian tubules in adults and larvae	Shireen Davies & Julian Dow (U Glasgow)	1	5
Gal4	uro	uro-Gal4	Specific to malpighian tubule principal cells	Shireen Davies & Julian Dow (U Glasgow)	1	6
Gal4	c724	c724-Gal4	Specific to malpighian tubule stellate cells	Shireen Davies & Julian Dow (U Glasgow)	1	7
Gal4	c710	c710-Gal4	Highly specific to Malpighian tubule stellate cells in adult and larva	Shireen Davies & Julian Dow (U Glasgow)	1	7
Gal4	6984	P{w[+mW.hs]=GawB}c754, w[1118]	Expresses in larval brain and fat body	BDSC	2	8
Gal4	30828	w[*]; P{w[+mW.hs]=GawB}Alp4[c232]	Expresses in ring neurons, Malpighian tubules, large field neurons & ellipsoid body	BDSC	2	9
Gal4	30844	w[*]; P{w[+mW.hs]=GawB}c601[c601]	Expresses GAL4 in hindgut, ureter, Malpighian tubules & protocerebrum	BDSC	2	7
Gal4	25374	y[1] w[*]; P{Act5C-GAL4-w}E1/CyO	Expresses GAL4 ubiquitously under control of Act5C promoter	BDSC	2	10
Gal4	33832	y[1] w[*]; P{w[+mC]=r4-GAL4}3	Expresses in fat body and salivary glands	BDSC	2	11
Gal4	1967	y[1] w[*]; P{w[+mW.hs]=GawB}34B	Expresses in embryonic salivary glands, posterior midgut, eye-antennal, haltere, leg and wing imaginal discs	BDSC	2	12
Gal4	43656	w[*]; P{w[+mC]=Scr-GAL4.4}1-3	Expresses strongly in the pattern of the Scr gene in the embryonic labial and T1 segments and anterior midgut	BDSC	2	13
Gal4	Myo1A	w[*]; P{w[+mW.hs]=GawB}Myo31DF[NP0001]; P{w[+mC]=UAS-CC3Ai}3	Expresses along most of midgut	BDSC	3	14
UAS	60000	w[1118]	GD RNAi line control	VDRC	4	
UAS	50507		Cyp12d1 "GD" UAS strain	VDRC	4	15
UAS	60100	y,w[1118];P{attP,y[+],w[3`]	KK RNAi line control	VDRC	4	16
UAS	109248		Cyp12d1 "KK" UAS strain	VDRC	4	17

Description of the GAL4-UAS RNAi lines used in this experiment stating the type of line used, the strain ID used by the vendors, the genotype of the strains, purpose of the line (GAL-4 lines: targeted tissues and cells ; UAS lines: control and experimental lines and type of transgene in the line), we then state the vendor and URL where they can be found and citation.

1. https://sites.google.com/site/tubulesite/Home/contacting-us---requesting-resources
2. https://bdsc.indiana.edu/
3. http://flygut.epfl.ch/patterns/1202

4. https://stockcenter.vdrc.at
5. Rosay et al. 1997
6. Terhzaz et al., 2010
7. Sozen et al., 1997
8. Hrdlicka et al., 2002
9. Yao Yang et al., 1995
10. Sedat, 2008.9.10
11. Park, 2011.1.24
12. Brand, A. 1997.6.30
13. Kaufman, 2013.1.31
14. Morgan, N.S., et al., 1994
15. Dietzl et al., 2007
16. E. W. Green et al., 2014
17. Keleman et al., 2009.8.5

The transgenes harbored by the UAS lines we employ were integrated via two different insertion methods. The GD UAS-RNAi transgenes are P-element based and are randomly inserted, while the KK lines have a defined insertion site, and transgenes are targeted to this site via the phiC31 integrase system.

Effects of *Cyp12d1* CRISPR mutations crossed to DSPR founder lines

The purpose of this experiment was to test the effects of *Cyp12d1* CRISPR mutations in the background of DSPR founders that are homozygous for alleles containing one copy (founders A4 and B2) or two copies (founders A7 and B7) of *Cyp12d1*. This experiment allows us to provide the CRISPR lines with full copies of *Cyp12d1* and ensures no animals will be homozygous for a mutation-containing *Cyp12d1* allele. Each of the three CRISPR mutant lines of A4; 55821; A4, along with the CRISPR control line, were crossed to each of the four DSPR founders. Each cross vial was initiated with 10 virgin CRISPR females and 5 founder males, and we used two replicate vials per cross. Once F1 cross progeny began to emerge, all vials were cleared, and subsequently we collected 35 females from each vial over CO_2 anesthesia. The next day, a total of 32 2-5 day old females per vial (64 per genotype) were tested for caffeine resistance using the activity monitor assay described previously. The entire experiment was repeated twice, once with 1% and once

with 1.5% caffeine media. All flies were reared and tested at 25°C, 50% humidity, and using a 12h:12h light:dark cycle.

Outbred population with a fixed copy number

To test the effect of the *Cyp12d1* CNV in an outbred background, two populations were created using inbred lines from the *Drosophila* Genetic Reference Panel (DGRP) (Mackay et al., 2012; Huang et al., 2014). Previously, a subset of the lines from the DGRP were genotyped for the *Cyp12d1* CNV (Najarro et al., 2015), allowing us to categorize them into a set of homozygous single copy lines (N = 156) and a set of homozygous double copy lines (N = 30). With this information we collected 10 virgin females and 5 males from each line and pooled flies together by CNV genotype into a pair of ½ gallon glass milk bottles containing cornmeal-yeast-molasses media. These populations were maintained for ~20 generations by transferring flies into fresh bottles every two weeks. The expected outcome is a pair of populations, each with a fixed number of copies of *Cyp12d1* in an otherwise outbred background.

To assay the populations for caffeine resistance, we collected flies from the population bottles and allowed them to lay eggs in standard narrow fly vials containing cornmeal-yeast-molasses media. After a generation in the incubator at 25°C, 50% humidity on a 12h:12h light:dark cycle, 10 virgin female and 5 male animals from both populations were collected in order to create both a heterozygous test sample using the bottle populations, as well as create samples within the bottle populations (32 vials per genotype were created). Once the crosses were created and developed, we collected 16 2-4-day old female flies per vial (total of 512 flies per genotype), we then used the monitor experiment previously described to test the difference in resistance to caffeine between the fixed populations.

Results

Cyp12d1 RNAi gut, malpighian tubes and fat body knockdown

We used the Gal4/UAS system to knock down the expression of _Cyp12d1_ in a tissue-specific manner. More specifically, we chose tissues known for detoxification in order to narrow down in what tissue the breakdown of caffeine takes place. We used the VDRC (Vienna Drosophila Resource Center) stock #109248 as our UAS-_Cyp12d1_ line and for the control non-RNAi line VDRC stock #60100. We obtained the Gal4 strains from the Bloomington Drosophila Stock Center (BDSC): #6984 (fat body), #30828 (malpighian tubules), #43656 (anterior midgut), #1967 (posterior midgut), #30844 (hindgut). Additionally, we used the malpighian tubule principal cell Gal4 drivers c42 and uro and the stellate cell drivers c710 and c724.

Knocking down _Cyp12d1_ in the adult posterior midgut (#1967) and hindgut (#30844) leads to a significant reduction in caffeine resistance (Figure 1C, t-test, $p < 0.001$), whereas knocking it down in the anterior midgut (43656) showed no significant difference. Depending on the Gal4 driver, we observe both no change (#30822, c710, uro) and increasing caffeine resistance (c42, c724, t-test, $p < 0.01$) following knockdowns in the malpighian tubules (Figure 1B), making the results a bit difficult to interpret especially since knocking down the expression of _Cyp12d1_ increased resistance. Finally, the fat body also showed a significant decrease in resistance when knocking _Cyp12d1_ down (Figure 1A, t-test, $p < 0.01$). Based on these results, the adult gut seems to be the most interesting tissue and it is suggested as a major site of caffeine detoxification via _Cyp12d1_.
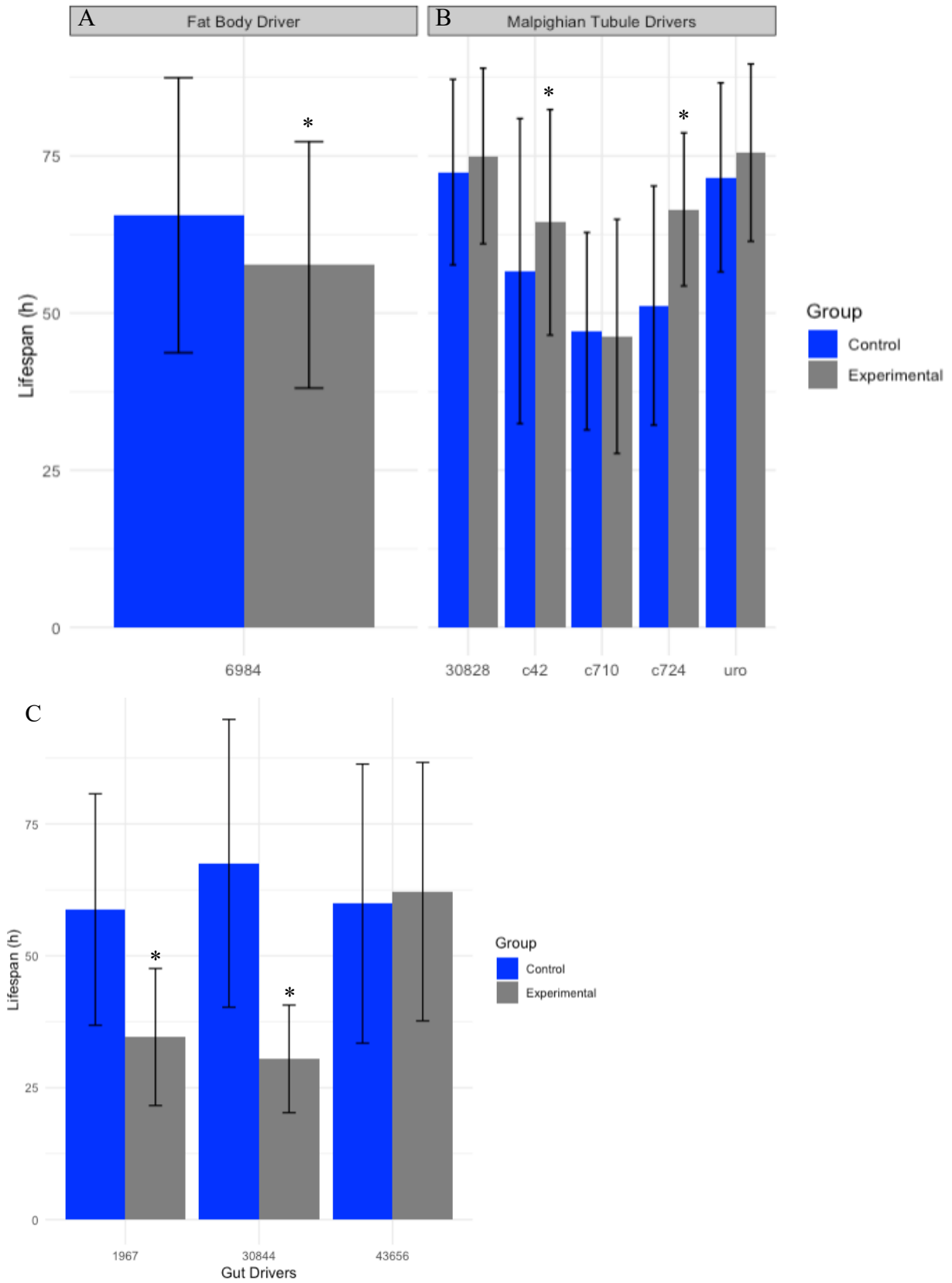
Figure 2 Caffeine resistance of RNAi knockdowns of Cyp12d1

We knockdown expression of *Cyp12d1* in 3 key detoxifying areas: the gut, the fat body and malpighian tubules. We measured the lifespan of the female progeny of Gal4-UAS crosses (64 individuals per genotype) when exposed to 1% caffeine. The bars show the mean lifespan (lifespan ± sd) of the cross. A) We show the fat body Gal4 driver #6984, B) from left to right the malpighian tubule Gal4 drivers used are #30828, c42, c710, c724 and uro, lastly C) from left to right the gut Gal4 drivers used are #1967 (posterior midgut), #30844 (hindgut), and #43656 (anterior midgut). In blue are the control individuals and in grey are the experimental samples. Asterisks signify a significant difference in resistance when compared to the control.

CRISPR-Cas9 premature stop codon mutations at *Cy12d1* in A4; 55821; A4 cause change in resistance to caffeine

Following work by Najarro et al. (2015), where the gene *Cyp12d1* was associated with caffeine resistance, we wanted to directly test the role of *Cyp12d1* on caffeine resistance. We accomplished this using CRISPR mutant lines . These lines which I refer to by the genotype of A4; 55821; A4 shows a series of different mutations. As previously mentioned, the CRISPR control line (A4; 55821[wt]; A4) has a single copy (2 total copies in diploid state) of *Cyp12d1,* whereas the CRISPR mutants have a premature STOP codon in the second exon of the only copy present (A4; 55821[Del1]; A4, A4; 55821[Del2]; A4, A4; 55821[Del3]; A4). As shown in Figure 3, we tested caffeine resistance at two different concentrations: 1% and 1.5%. When using 1% and 1.5% of caffeine we observed that A4; 55821[Del1]; A4 and A4; 55821[Del2]; A4 lived a shorter amount of time than A4; 55821[wt]; A4; we observed a significant difference between A4; 55821[wt]; A4 and A4; 55821[Del1]; A4 (tukey, p = 0.0000189). When comparing A4; 55821[Del3]; A4 with A4; 55821[wt]; A4, this one lived for a significantly longer amount of time when tested at 1% (tukey, p = 0.0000023) and showed a similar trend at 1.5% caffeine. While A4; 55821[Del1]; A4 and A4; 55821[Del2]; A4 display the expected result of reduced resistance to caffeine when compared to A4; 55821[wt]; A4 (tukey, p =0.0000000, p =0.0042209, respectively), A4; 55821[Del3]; A4 increased in resistance.

Effects of *Cyp12d1* on caffeine resistance in cross A4; 55821; A4 to DSPR founder line

We further hypothesized that since we had created lines that did not exist in nature (null copy of *Cyp12d1*), other detoxification genes might be compensating for its loss. More specifically, other cytochrome P450s might be compensating for the complete loss of *Cyp12d1* in the mutant A4; 55821; A4. We then set out to test the caffeine resistance of trans-heterozygous F1 progeny between the A4; 55821; A4 CRISPR lines and DSPR founders A4, B2, A7 and B7 to compare animals containing 1, 2, and 3 functional copies of the gene *Cyp12d1* and carrying at least one functional copy of *Cyp12d1*. Animals A4/A4; 55821[Del2]/A4; A4/A4 and A4/B2; 55821[Del2]/B2; A4/B2 contain 1 full copy of *Cyp12d1* from each founder and A4/A7; 55821[Del2]/A7; A4/A7 and A4/B7; 55821[Del2]/B7; A4/B7 have 2 full copies of *Cyp12d1* provided by the founder, all of these lines contain as well a *Cyp12d1* CRISPR copy. When comparing the mutants to the control, we see trends, but we see no significant difference using tukey testing between them (Figure 4)

A4; 55821; A4 whole body RNAseq brings light to mutant splice site

Given the results when testing the caffeine resistance in the A4; 55821; A4 CRISPR control and mutants, as well as observing no significant differences across genotypes from the DSPR founder cross, we asked ourselves if we had in fact accomplished to create a knockdown of *Cyp12d1*. To test this, we performed single-read RNAseq on A4; 55821[wt]; A4, A4; 55821[Del1]; A4, A4; 55821[Del2]; A4, and A4; 55821[Del3]; A4 and compared expression between them. *Cyp12d1* showed significantly lower expression in the CRISPR mutants and was 1 of 5 differentially expressed genes using kallisto/sleuth (10% FDR) (Figure 5). This confirmed that we indeed created a knockdown version of *Cyp12d1*; however, when looking at the reads more

closely, we observed that all of these lines exhibit a second difference within the *Cyp12d1* gene, this being a different splice site mutation. The lines A4; 55821[wt]; A4 (CRISPR control), A4; 55821[Del1]; A4, A4; 55821[Del2]; A4, (the ones that showed a reduced resistance) have a mutated splice site at the end of the third exon, whereas the A4; 55821[Del3]; A4 (the one that showed an increase resistance) does not have this mutation (Figure 6). We believe this splice site mutation difference between the lines may lead to the phenotype observed where A4; 55821[wt]; A4 showed a lower resistance than A4; 55821[Del3]; A4, since the former has splice site mutation while A4; 55821[Del3]; A4 does not. This second mutation was a result of using the injection line 55821, which is not isogenic.
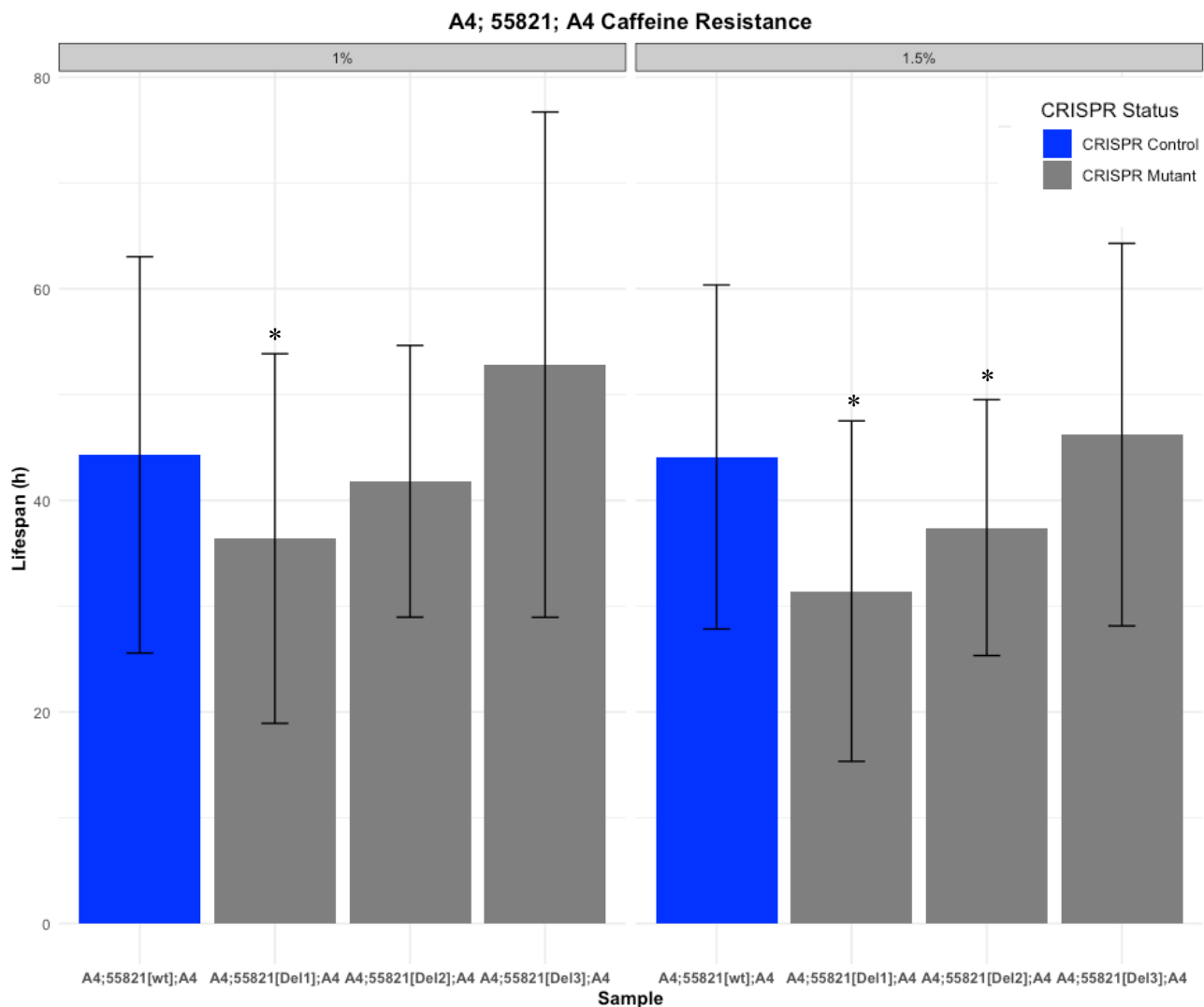
Figure 3 CRISPR-Cas9 edited lines tested for caffeine resistance

We tested CRISPR edited and control lines A4; 55821; A4 for caffeine resistance. We used both 1% (~250 female flies) and 1.5% (~128 females flies) caffeine. The bars illustrate the mean lifespan (lifespan ± sd) of each line tested. Shown from left to right the blue bar represents the CRISPR control line and the grey bars represent the CRISPR mutants. On the x-axis the samples are named. Asterisks signify a significant difference in resistance when compared to the control.
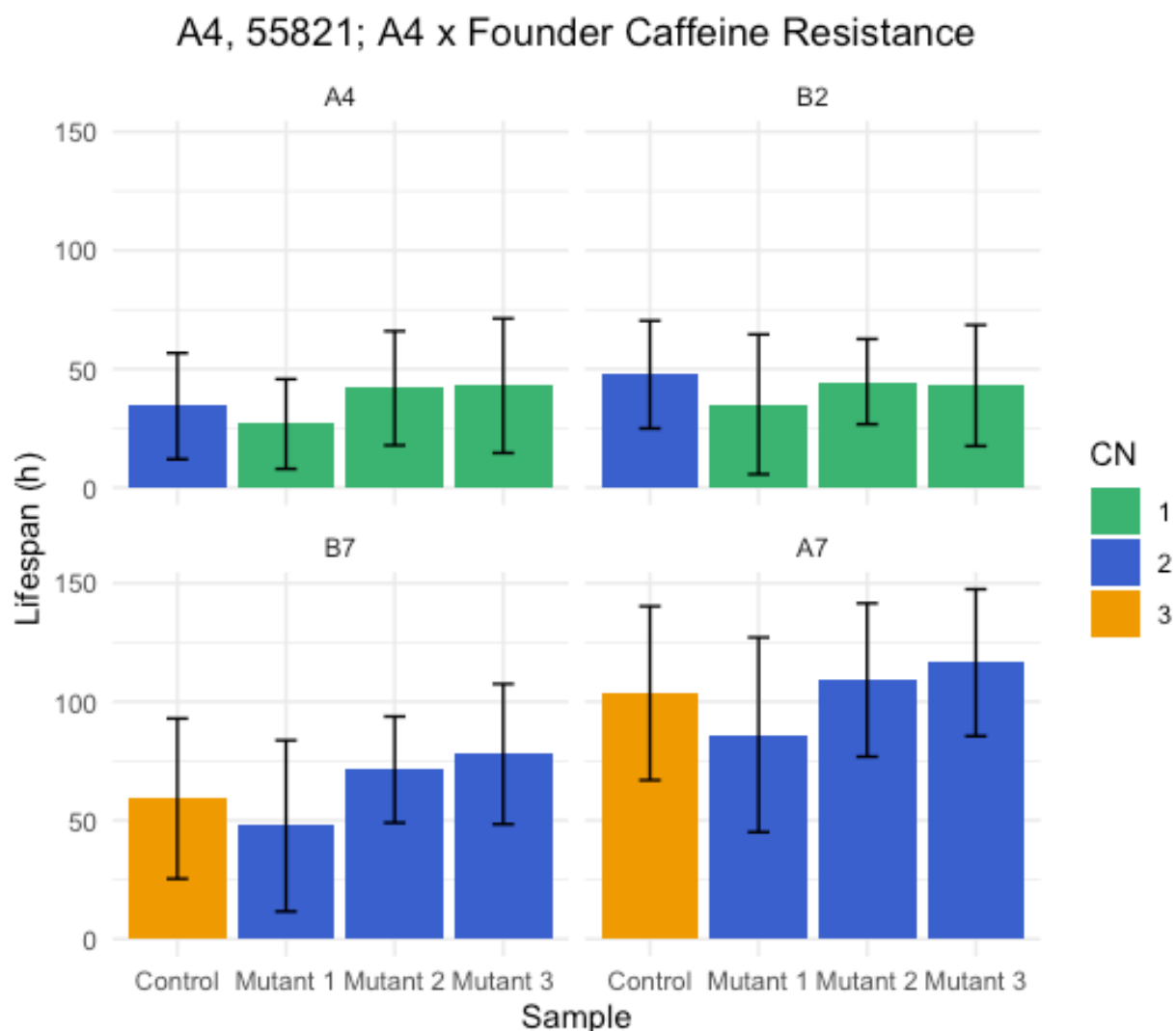

Figure 4 A4; 55821; A4 x Founder Caffeine Resistance

F1 progeny of A4; 55821; A4 CRISPR lines and DSPR founders A4, B2, A7 and B7 were tested for caffeine resistance. Separated into 4 quadrants based on the founder used in the cross. Founders A4 and B2 provided 1 copy of *Cyp12d1* to each cross and founders A7 and B7 provided 2 copies

of *Cyp12d1* to each cross. Blue bars represent individuals with 2 total copies of *Cyp12d1*, green represent 1 total copy present and orange have 3 total copies of *Cyp12d1*. Lifespan is on the y-axis and CRISPR controls and mutants on the x-axis. No significant difference was observed within each quadrant.
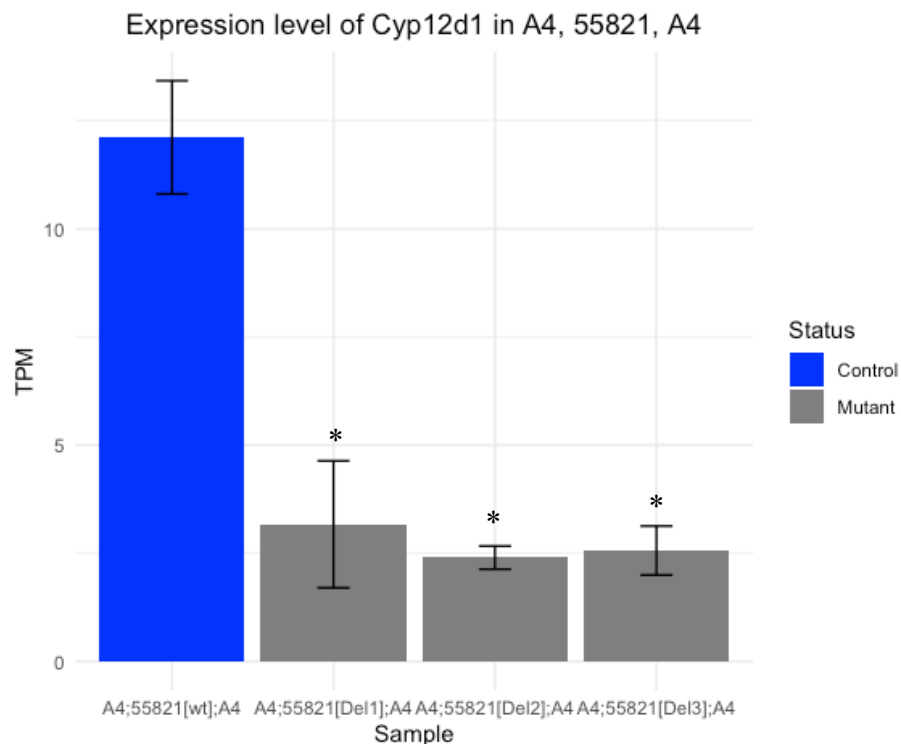

Figure 5 Cyp12d1 expression levels in whole body RNAseq

Shown are the transcript per million of the gene *Cyp12d1* of the CRISPR lines A4; 55821; A4. The x-axis shows the name of the samples. Blue represents the CRISPR control and grey the mutants. *Cyp12d1* showed significantly lower expression in the CRISPR mutants when compared to the control. Asterisks signify a significant difference in expression when compared to the A4;55821[wt];A4.
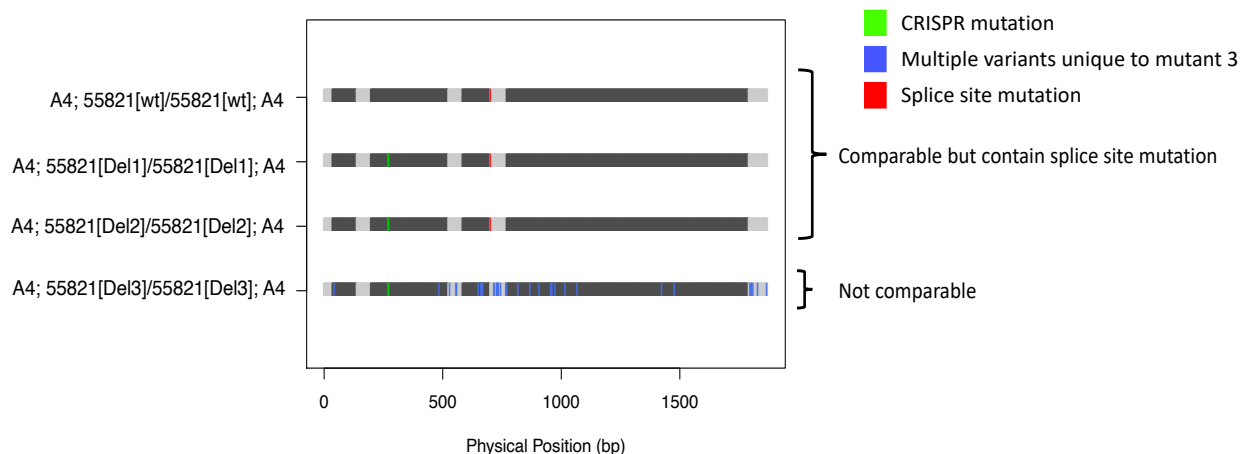
Figure 6 Splice site mutation diagram

Diagram of the gene *Cyp12d1* in the A4; 55821; A4 lines. On the Y-axis we show sample names and, on the x-axis, physical position within the gene in base pairs. Light gray squares represent introns and dark gray represent exons. The green line represents the position of the CRISPR mutations created, red the splice site variation position and in blue we show the different unique variants present in A4; 55821[Del3]/55821[Del3]; A4. Finally, we point which lines are comparable and which one is not.

Effect of *Cyp12d1* CNV on caffeine resistance in CRISPR-Cas9 isogenic mutants A4; A7; A4

Given the segregating mutations present across the A4; 55821; A4 lines, we set out to create isogenic CRISPR mutants. In order to develop this set of lines, we created an isogenic injection strain using the DSPR founder A7; with this, we were able to generate the lines A4; A7; A4. The founder A7 has an original genotype of a duplicated copy of *Cyp12d1* and because of this we were able to create true copy number variant mutants. The duplicated line A4; A7[2 copies]; A4 (total of 4 copies in diploid) has no CRISPR mutations even though it went through the CRISPR pipeline: the single line A4; A7[1 copy]; A4 (total of 2 copies in diploid) is a chimera between the two copies present in the A7 strain that were seamlessly joined. In order to test all three genotypes A4; A7[2 copies]; A4 (4 copies in diploid), A4; A7[1 copy]; A4 (2 copies in diploid) and A4; A7[1 copy]/ A7[2 copies]; A4 (3 copies in diploid) for caffeine resistance, we

crossed A4; A7[2 copies]; A4 and A4; A7[1 copy]; A4, thus creating a heterozygous line A4; A7[1 copy]/ A7[2 copies]; A4. We conducted 2 separate experiments where we tested the 3 genotypes. In experiment 1, all animals lived for a shorter amount of time than those in experiment 2; however, we saw no significant difference between the experiments. Once again, when comparing the different strains, we see no significant difference between them when conducting a tukey test. There is a slight difference between the mean lifespans where the duplicated strain is higher, but it is not significant.
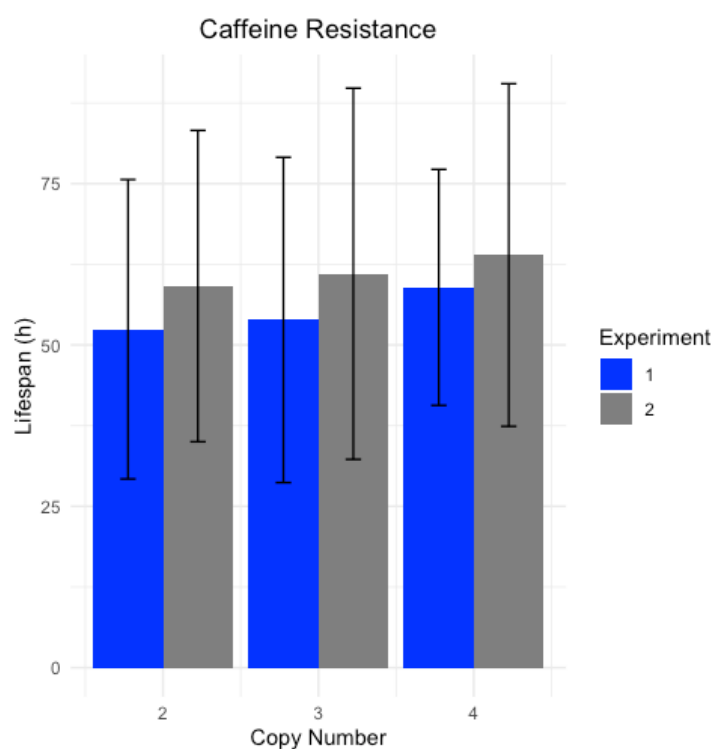

Figure 7 Caffeine resistance of CRISPR mutants A4; A7; A4 lines

A4; A7; A4 lines were tested in 2 separate experiments for caffeine resistance, experiment 1 shown in blue and experiment 2 in grey. Lifespan in hours is on the y-axis and on the x-axis are the samples divided by the copy number of *Cyp12d1* copies they possess. No significant difference was observed between the samples.

CNV seems to play a role in the expression levels of *Cyp12d1* in the gut

Taking into account the results obtained in the RNAi knockdowns, we wanted to test for differences in gene expression levels in the fly gut between the A4; A7; A4 CRISPR strains in both naïve and caffeine conditions. We used RNAseq to measure RNA expression in female gut tissue in order to test for difference in expression between animals with different CNV of *Cyp12d1*. We first tested for differential gene expression between the naïve and caffeinated conditions of the lines A4; A7[2 copies]; A4, A4; A7[1 copy]; A4 and A4; A7[1 copy]/A7[2 copies]; A4: 127 genes were found to be differentially expressed using kallisto/sleuth (5% FDR), our gene of interest *Cyp12d1,* being among the top of them. Under naïve conditions, we observed a significant increase in expression of *Cyp12d1* when comparing A4; A7[1 copy]/A7[2 copies]; A4 to A4; A7[1 copy]; A4 strain (tukey, p= 0.0170778) as well as the duplicated strain when compared to single (tukey, p = 0.0029267). We saw no significant difference in expression of *Cyp12d1* between A4; A7[1 copy]/A7[2 copies]; A4 and A4; A7[2 copies]; A4 strains when compared to each other under naïve conditions. When comparing the strains under caffeine conditions, we saw a significant increase in expression of *Cyp12d1* in A4; A7[2 copies]; A4 when compared to A4; A7[1 copy]; A4 (tukey, p= 0.0163826). Even though we weren't able to find a significant difference in *Cyp12d1* when comparing A4; A7[1 copy]/A7[2 copies]; A4 to A4; A7[1 copy]; A4 or when comparing A4; A7[1 copy]/A7[2 copies]; A4 to A4; A7[2 copies]; A4 genotype, there is a strong trend that can be seen in figure 8 (tukey, p = 0.2976407, p = 0.1205709, respectively).
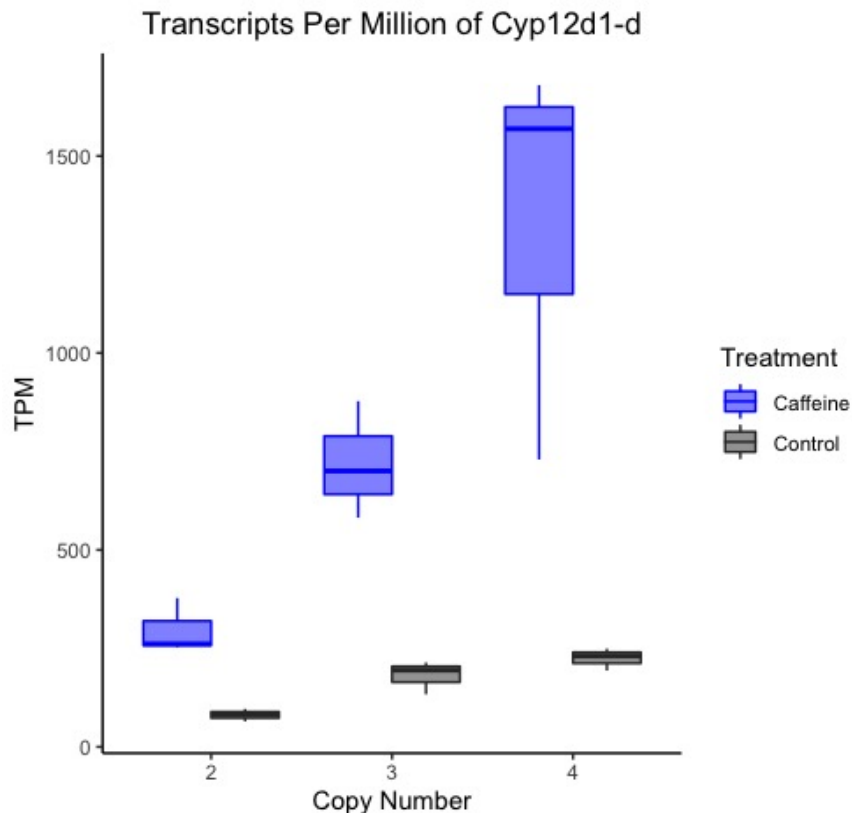
Figure 8 Expression levels of *Cyp12d1* under naïve and caffeine conditions

CRISPR lines A4; A7; A4 were exposed to caffeine and naïve conditions, after 4 hours the gut tissue was dissected out and RNA was extracted to perform RNAseq. Shown is *Cyp12d1* expression levels under naïve (grey) and caffeine (blue) conditions. The x-axis shows the copy number of *Cyp12d1* copies present in the strain; on y-axis is the transcripts per million. When compared to the animals under the same conditions a significant increase based on CNV was observed.

Testing outbred population with fix *Cyp12d1* copy number

So far, we have tested *Cyp12d1* under a fairly controlled background; thus, we speculated if we could detect a significant difference in an outbred population with a fixed copy number of *Cyp12d1*. We tested the caffeine resistance of the outbred populations made from the DGRP population. We created two separate populations based on the single or duplicated copy number status of each DGRP line. After ~20 generations, we tested the resistance to caffeine (1.5%) of the populations. The single and duplicated populations were also crossed in order to test their

offspring, thus creating heterozygous individuals. When exposed to caffeine, we observed that the heterozygous (3 total copies) individuals lived for a significantly longer amount of time than single (2 total copies) and duplicated (4 total copies) individuals (tukey test p < 0.001). We were concerned that the heterozygous individuals were more fit than the single and duplicated populations (due to increased heterozygosity). Thus, we set up a starvation resistance assay to test for fitness. Heterozygous individuals showed slightly more resistance to starvation; however, it was not significant.
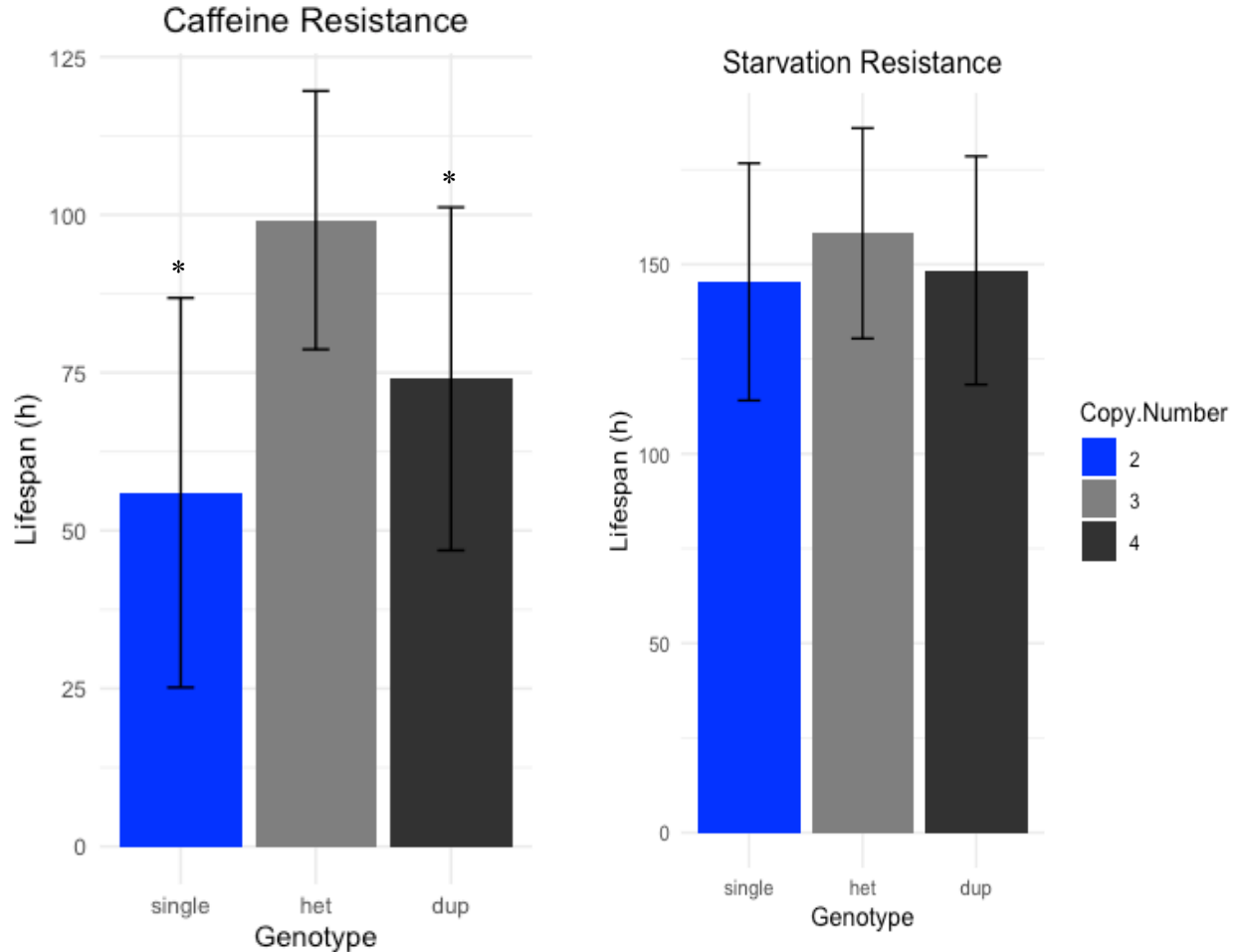

Figure 9 Outbred population with fix CNV of *Cyp12d1*

Two otherwise outbred populations were generated from strains homozygous for one or two copies of *Cyp12d1*. Female flies from each population, as well as F1 progeny from a cross between individuals from each population. were used to test for caffeine (~500 individuals per genotype) and starvation (~400 individuals per genotype) resistance. Each bar represents the mean lifespan (lifespan ± sd) of each population. The single population has two copies of *Cyp12d1*, the het or heterozygous population has three copies of *Cyp12d1*, and the dup or duplicated population has four copies of *Cyp12d1*. A) Shows resistance to caffeine, B) presents starvation resistance. The x-axis shows the copy number present in each of the population tested on blue is the single population (2 total gene copies), on light grey is the heterozygous population (3 total copies) and on dark grey is the duplicated population (4 total gene copies). On the y-axis is the lifespan in hours. Asterisks signify a significant difference in resistance when compared to the heterozygous individual.

Discussion

In previous work in *D. melanogaster,* Najarro et al. (2015) mapped several loci contributing to resistance to caffeine, one of which implicated the cytochrome P450 gene *Cyp12d1*. They supported the involvement of this gene in resistance via RNAseq and RNAi knockdowns and showed an association between greater *Cyp12d1* copy number and increased caffeine resistance. To further test this, here we used CRISPR-Cas9 to generate putative loss-of-function mutations of *Cyp12d1*, confirming via RNAseq that mutation-carrying strains have significantly lower *Cyp12d1* expression than our single control strain. We measured caffeine resistance and observed a significant reduction in resistance in 2 out of 3 CRISPR mutant strains relative to the control strain. Interestingly, the third mutant strain exhibited higher resistance to caffeine than the control. We presented evidence (Figure 6) suggesting that this result is due to heterozygosity within the strain used for editing, and more specifically, a mutant splice site. This makes it challenging to compare directly among strains. To overcome this, we edited an isogenic strain containing two nearly identical copies of *Cyp12d1*. Since our guide RNA will bring the Cas9 protein to both genes, we created a strain where we eliminate the sequence between the guide RNA sites in the tandemly duplicated gene copies. This created an edited strain with 1 functional copy

of *Cyp12d1*. We also created a non-mutant CRISPR duplicated strain; using both of these, we tested the effect of CNV of *Cyp12d1* in caffeine resistance. We were not able to detect any significant difference between the resistance to caffeine. This is an unexpected result as we believed based on previous data that we should have observed an increased in resistance based on higher copy number of *Cyp12d1*.

Testing the caffeine resistance of an outbred population with a fixed number of copies of *Cyp12d1* was performed to test for increased resistance as a function of copy number, given that the genetic background was completely segregating. Two populations were created using the lines from the DGRP. The heterozygous individuals, which were created from crossing the two populations, showed a significant increase in resistance to caffeine when compared to both the single and duplicated individuals. When seeing these results, we tested the fitness of the populations by conducting a starvation assay which showed no significant difference in fitness between the populations. Once again, these results do not match our hypothesis that higher copy number of *Cyp12d1* correlates to increased resistance.

The discrepancy that we see between the CNV of *Cyp12d1* and the phenotypic response of caffeine resistance in the A4; A7; A4 CRISPR strains as well as the outbred populations, where we don't observe a correlation between higher CNV and increased resistance, could be explained by the compensation of other cytochrome P450s. In a study done by Coelho et al., (2015), they tested caffeine metabolism in *Drosophila melanogaster*, by identifying primary caffeine metabolites in the fly. Additionally, they conducted a transcriptomic screen to detect the genes involved in this metabolism, while following up with RNAi knockdowns of candidate genes

obtained from the screen. *Cyp12d1* was one of the top genes detected in caffeine metabolism in this study. They show a 4-fold transcriptional increase in *Cyp6a8* when knocking down expression of *Cyp12d1*(Coelho et al., 2015). This suggests a potential mechanism of compensation between *Cyp12d1* and *Cyp6a8*. The differences in CNV of the A4; A7; A4 CRISPR strains and the outbred populations could be activating this compensation pathway causing the strains with a lower copy number to live for a similar or higher amount of time as the higher copy number strains.

Changes in expression based on CNV of *Cyp12d1*

Through tissue specific RNAi knockdowns, we were able to narrow down the gut as a major place of caffeine detoxification through *Cyp12d1*. We utilized this information to test RNA expression levels in the gut under naïve and caffeinated conditions. We tested the A4, A7, A4 CRISPR lines in which we were able to control not only the genetic background but also the number of copies of *Cyp12d1* present. When looking at the conditions separately, we were able to observe a significant increase in expression levels based on CNV of *Cyp12d1*, we saw this in both the naïve and caffeinated condition. Based on these results, we can say that there is a detectable difference in expression based on the CNV; however, even though there is a significant difference in the mRNA expression this does not guarantee a correlation with the protein expression of *Cyp12d1* (Greenbaum et al., 2003). mRNA post-transcriptional mechanisms are diverse and complicated and not fully understood, we can hypothesize that there is a threshold for the concentration of CYP12d1 protein made dictated by post-transcriptional factors (Greenbaum et al., 2003).

Cyp12d1 major detoxification gene

Cyp12d1 is differentially expressed in response to a varied number of xenobiotics, some of which include: 4,4′-dichlorodiphenyltrichloroethane (DDT), malathion, chlorantraniliprole, and cyantraniliprole. (Daborn et al., 2007;Goff et al., 2003; Kim et al., 2018; Schmidt et al., 2010; Battlay et al., 2018; Green et al., 2019). When testing resistance to DDT Schmidt et al. (2010) tested CNV as a possible explanation for increased resistance to DDT, they did not find an association between *Cyp12d1*'s CNV and DDT resistance (Schmidt et al., 2010). This shows that even though *Cyp12d1* is associated with increased resistance the CNV plays no detectable role in it. Similar to these studies, we have shown through RNAseq and RNAi knockdowns the involvement of *Cyp12d1* in caffeine detoxification. However, given the complicated results obtained through testing the phenotypic response to caffeine resistance, more work needs to be done in order to determine if the CNV of *Cyp12d1* increases resistance as copy number also increases. In the future, protein studies are needed in order to measure protein levels and to see if there is a correlation between RNA expression levels and protein levels.

References

Adams, M. D. (2000). The Genome Sequence of Drosophila melanogaster. *Science*, *287*(5461), 2185–2195. https://doi.org/10.1126/science.287.5461.2185

Baldi, P., & Long, A. D. (2001). A Bayesian framework for the analysis of microarray expression data: Regularized t -test and statistical inferences of gene changes. *Bioinformatics*, *17*(6), 509–519. https://doi.org/10.1093/bioinformatics/17.6.509

Battlay, P., Leblanc, P. B., Green, L., Garud, N. R., Schmidt, J. M., Fournier-Level, A., & Robin, C. (2018). Structural Variants and Selective Sweep Foci Contribute to Insecticide Resistance in the *Drosophila* Genetic Reference Panel. *G3&#58; Genes|Genomes|Genetics*, *8*(11), 3489–3497. https://doi.org/10.1534/g3.118.200619

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Chahine, S., & O'Donnell, M. J. (2011). Interactions between detoxification mechanisms and excretion in Malpighian tubules of Drosophila melanogaster. *THE JOURNAL OF EXPERIMENTAL BIOLOGY*, 7.

Chakraborty, M., VanKuren, N. W., Zhao, R., Zhang, X., Kalsow, S., & Emerson, J. J. (2018). Hidden genetic variation shapes the structure of functional elements in Drosophila. *Nature Genetics*, *50*(1), 20–25. https://doi.org/10.1038/s41588-017-0010-y

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Coelho, A., Fraichard, S., Le Goff, G., Faure, P., Artur, Y., Ferveur, J.-F., & Heydel, J.-M.

(2015). Cytochrome P450-dependent metabolism of caffeine in Drosophila melanogaster.

PloS One, *10*(2), e0117328. https://doi.org/10.1371/journal.pone.0117328

Daborn, P. J., Lumb, C., Boey, A., Wong, W., ffrench-Constant, R. H., & Batterham, P. (2007).

Evaluating the insecticide resistance potential of eight Drosophila melanogaster

cytochrome P450 genes by transgenic over-expression. *Insect Biochemistry and*

Molecular Biology, *37*(5), 512–519. https://doi.org/10.1016/j.ibmb.2007.02.008

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K.,

Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., & Dickson, B. J. (2007). A

genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila.

Nature, *448*(7150), 151–156. https://doi.org/10.1038/nature05954

Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O., & Long, M. (2008). Natural selection

shapes genome-wide patterns of copy-number polymorphism in Drosophila

melanogaster. *Science (New York, N.Y.)*, *320*(5883), 1629–1631.

https://doi.org/10.1126/science.1158078

Faucon, F., Dusfour, I., Gaude, T., Navratil, V., Boyer, F., Chandre, F., Sirisopa, P., Thanispong,

K., Juntarajumnong, W., Poupardin, R., Chareonviriyaphap, T., Girod, R., Corbel, V.,

Reynaud, S., & David, J.-P. (2015). Identifying genomic changes associated with

insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing.

Genome Research, *25*(9), 1347–1359. https://doi.org/10.1101/gr.189225.115

Faucon, F., Gaude, T., Dusfour, I., Navratil, V., Corbel, V., Juntarajumnong, W., Girod, R.,

Poupardin, R., Boyer, F., Reynaud, S., & David, J.-P. (2017). *In the hunt for genomic*

markers of metabolic resistance to pyrethroids in the mosquito Aedes aegypti: An integrated next-generation sequencing approach. 20.

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-Smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., & Lee, C. (2006). Copy number variation: New insights in genome diversity. *Genome Research*, *16*(8), 949–961. https://doi.org/10.1101/gr.3677206

Goff, G. L., Boundy, S., Daborn, P. J., Yen, J. L., Sofer, L., Lind, R., Sabourault, C., & Madi-Ravazzi, L. (2003). Microarray analysis of cytochrome P450 mediated insecticide resistance in Drosophila. *Insect Biochemistry and Molecular Biology*, 8.

Gong, J., Cheng, T., Wu, Y., Yang, X., Feng, Q., & Mita, K. (2019). Genome-wide patterns of copy number variations in Spodoptera litura. *Genomics*, *111*(6), 1231–1238. https://doi.org/10.1016/j.ygeno.2018.08.002

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J., & O'Connor-Giles, K. M. (2013). Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics*, *194*(4), 1029–1035. https://doi.org/10.1534/genetics.113.152710

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., & O'Connor-Giles, K. M. (2014). Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in *Drosophila*. *Genetics*, *196*(4), 961–971. https://doi.org/10.1534/genetics.113.160713

Green, E. W., Fedele, G., Giorgini, F., & Kyriacou, C. P. (2014). A Drosophila RNAi collection is subject to dominant phenotypic effects. *Nature Methods*, *11*(3), 222–223. https://doi.org/10.1038/nmeth.2856

Green, L., Battlay, P., Fournier-Level, A., Good, R. T., & Robin, C. (2019). *Cis—And trans - acting variants contribute to survivorship in a naïve Drosophila melanogaster* population exposed to ryanoid insecticides. *Proceedings of the National Academy of Sciences*, *116*(21), 10424–10429. https://doi.org/10.1073/pnas.1821713116

Greenbaum, D., Colangelo, C., Williams, K., & Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, *4*(9), 117. https://doi.org/10.1186/gb-2003-4-9-117

Hrdlicka, L., Gibson, M., Kiger, A., Micchelli, C., & Schober, M. (n.d.). *Analysis of twenty-four Gal4 lines in Drosophila melanogaster*. 7.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A. M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R. F., Magwire, M. M., Blankenburg, K., Carbone, M. A., Chang, K., Ellis, L. L., Fernandez, S., Han, Y., Highnam, G., Hjelmen, C. E., … Mackay, T. F. C. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Research*, *24*(7), 1193–1208. https://doi.org/10.1101/gr.171546.113

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W., & Lee, C. (2004). Detection of large-scale variation in the human genome. *Nature Genetics*, *36*(9), 949–951. https://doi.org/10.1038/ng1416

International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., … Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851–861. https://doi.org/10.1038/nature06258

Kim, J. H., Moreau, J. A., Zina, J. M., Mazgaeen, L., Yoon, K. S., Pittendrigh, B. R., & Clark, J. M. (2018). Identification and interaction of multiple genes resulting in DDT resistance in the 91-R strain of Drosophila melanogaster by RNAi approaches. *Pesticide Biochemistry and Physiology*, *151*, 90–99. https://doi.org/10.1016/j.pestbp.2018.03.003

King, E. G., Macdonald, S. J., & Long, A. D. (2012). Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics*, *191*(3), 935–949. https://doi.org/10.1534/genetics.112.138537

Ku, C. S., Agus Salim, K. S. C., & En Yun Loy. (2010). The discovery of human genetic variations and their use as disease markers: Past, present and future. *Journal of Human Genetics*, 13.

Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., Lin, Y., MacDonald, J. R., Pang, A. W. C., Shago, M., Stockwell, T. B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S. A., … Venter, J. C. (2007). The diploid genome sequence of an individual human. *PLoS Biology*, *5*(10), e254. https://doi.org/10.1371/journal.pbio.0050254

Li, X., Schuler, M. A., & Berenbaum, M. R. (2007a). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, *52*, 231–253. https://doi.org/10.1146/annurev.ento.51.110104.151104

Li, X., Schuler, M. A., & Berenbaum, M. R. (2007b). Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. *Annual Review of Entomology*, *52*(1), 231–253. https://doi.org/10.1146/annurev.ento.51.110104.151104

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt, R. R. H., Barrón,

M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., … Gibbs, R. A. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature*, *482*(7384), 173–178. https://doi.org/10.1038/nature10811

McDonnell, C. M., King, D., Comeron, J. M., Li, H., Sun, W., Berenbaum, M. R., Schuler, M. A., & Pittendrigh, B. R. (2012). Evolutionary Toxicogenomics: Diversification of the Cyp12d1 and Cyp12d3 Genes in Drosophila Species. *Journal of Molecular Evolution*, *74*(5–6), 281–296. https://doi.org/10.1007/s00239-012-9506-3

Metzker, M. L. (2010). Sequencing technologies—The next generation. *Nature Reviews Genetics*, *11*(1), 31–46. https://doi.org/10.1038/nrg2626

Najarro, M. A., Hackett, J. L., Smith, B. R., Highfill, C. A., King, E. G., Long, A. D., & Macdonald, S. J. (2015). Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. *PLoS Genetics*, *11*(11), e1005663. https://doi.org/10.1371/journal.pgen.1005663

Pimentel, H., Bray, N. L., Puente, S., Melsted, P., & Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, *14*(7), 687–690. https://doi.org/10.1038/nmeth.4324

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rees, E., Kirov, G., Sanders, A., Walters, J. T. R., Chambert, K. D., Shi, J., Szatkiewicz, J., O'Dushlaine, C., Richards, A. L., Green, E. K., Jones, I., Davies, G., Legge, S. E., Moran, J. L., Pato, C., Pato, M., Genovese, G., Levinson, D., Duan, J., … Owen, M. J. (2014). Evidence that duplications of 22q11.2 protect against schizophrenia. *Molecular Psychiatry*, *19*(1), 37–40. https://doi.org/10.1038/mp.2013.156

Remnant, E. J., Good, R. T., Schmidt, J. M., Lumb, C., Robin, C., Daborn, P. J., & Batterham, P. (2013). Gene duplication in the major insecticide target site, Rdl, in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14705–14710. https://doi.org/10.1073/pnas.1311341110

Schmidt, J. M., Good, R. T., Appleton, B., Sherrard, J., Raymant, G. C., Bogwitz, M. R., Martin, J., Daborn, P. J., Goddard, M. E., Batterham, P., & Robin, C. (2010). Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genetics*, *6*(6), e1000998. https://doi.org/10.1371/journal.pgen.1000998

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., … Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.)*, *305*(5683), 525–528. https://doi.org/10.1126/science.1098918

Sozen, M. A., Armstrong, J. D., Yang, M., Kaiser, K., & Dow, J. A. T. (1997). Functional domains are specified to single-cell resolution in a Drosophila epithelium. *Proceedings of the National Academy of Sciences*, *94*(10), 5207–5212. https://doi.org/10.1073/pnas.94.10.5207

Terhzaz, S., Finlayson, A. J., Stirrat, L., Yang, J., Tricoire, H., Woods, D. J., Dow, J. A. T., & Davies, S.-A. (2010). Cell-specific inositol 1,4,5 trisphosphate 3-kinase mediates epithelial cell apoptosis in response to oxidative stress in Drosophila. *Cellular Signalling*, *22*(5), 737–748. https://doi.org/10.1016/j.cellsig.2009.12.009

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., Matthews, B. B., Millburn, G., Antonazzo, G., Trovisco, V., Kaufman, T. C., Calvi, B.

R., the FlyBase Consortium, Perrimon, N., Gelbart, S. R., Agapite, J., Broll, K., Crosby, L., Santos, G. dos, … Baker, P. (2019). FlyBase 2.0: The next generation. *Nucleic Acids Research*, *47*(D1), D759–D765. https://doi.org/10.1093/nar/gky1003

Weetman, D. (2018). Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African Anopheles gambiae. *SCIENTIfIC RePortS*, 12.

Yao Yang, M., Armstrong, J. D., Vilinsky, I., Strausfeld, N. J., & Kaiser, K. (1995). Subdivision of the drosophila mushroom bodies by enhancer-trap expression patterns. *Neuron*, *15*(1), 45–54. https://doi.org/10.1016/0896-6273(95)90063-2

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nature Reviews. Genetics*, *16*(3), 172–183. https://doi.org/10.1038/nrg3871

Zhang, X., Nie, M., Zhao, Q., Wu, Y., Wang, G., & Xia, Q. (2015). Genome-wide patterns of genetic variation among silkworms. *Molecular Genetics and Genomics: MGG*, *290*(4), 1575–1587. https://doi.org/10.1007/s00438-015-1017-7

Zhao, Q., Han, M.-J., Sun, W., & Zhang, Z. (2014). Copy number variations among silkworms. *BMC Genomics*, *15*(1), 251. https://doi.org/10.1186/1471-2164-15-251