A BULKED SEGREGANT ANALYSIS OF PYRETHROID RESISTANCE IN *MUSCA DOMESTICA*

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Jamie Catherine Freeman

December 2020

ABSTRACT

Pyrethroid insecticides are commonly used to control house flies, which are an important pest of animal production facilities; however insecticide resistance threatens our ability to control fly populations. While historically the field of insecticide resistance has focused on genes that are tractable to study, forward genetic methods in non-model organisms have the potential to advance our knowledge of insecticide resistance and insect toxicology in general. I first collected house flies from around the USA to assess how insecticide resistance levels have changed in the last decade. Next, using the most resistant population collected in the monitoring survey, I use a genetic mapping scheme, bulked segregant analysis, combined with short and long read whole genome sequencing and differential expression analysis to investigate the molecular mechanisms behind its extreme pyrethroid resistance phenotype. The known pyrethroid resistance locus, the target-site the *voltage-sensitive sodium channel* was recovered in the analysis, along with a second major peak on chromosome 5, encompassing a cluster of cytochrome P450 genes. This cluster contains extensive structural variation, which are potentially responsible for the overexpression of *CYPs* at the locus.

BIOGRAPHICAL SKETCH

Jamie Catherine Freeman was born and raised in central New Jersey. She graduated with a degree in entomology from Cornell University in 2015. After a short stint in the ecotourism industry in the beautiful cloud forests of Monteverde, Costa Rica, she returned to Ithaca in January 2016 to work as a technician in the lab of Dr. Jeffrey G. Scott at Cornell. This started an interest in the evolution of insecticide resistance that led her to pursue her MS in entomology in the Scott lab.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Jeff Scott for his constant support and mentoring, and for believing I would figure it all out, even when I wasn't sure I would. I thank my committee members, Ping Wang and Phillip Messer, for their advice on experimental design and analysis. I thank the expert staff of the Cornell BRC Genomics and Bioinformatics facilities who made almost every experiment in this thesis possible, especially Peter Schweitzer and Qi Sun. I thank the members of the Scott lab past and present for always being willing to act as sounding boards, look at a really cool piece of data I just *needed* to show someone right away, help out with an overwhelming experiment, or just chit chat about ridiculous things. I thank the Scott lab undergrads for cheerfully keeping the lab going, and whose absence was sorely felt during the holidays and COVID shutdown.

My friends here in Ithaca and spread around the country have always reminded me to take a breath and have some fun. Their humor and love always brightens my days. I thank my mother, Loretta Brennan Boyle, and the rest my family for their love and support, and for always letting me know they're proud of me, even if they're not quite sure why I want to spend so much time in the company of flies. I especially thank Zach, for his support and encouragement and for his meals, without which I may have survived, but I surely would have been many times unhappier.

TABLE OF CONTENTS

CHAPTER 1

INSECTICIDE RESISTANCE MONITORING OF HOUSE FLY POPULATIONS FROM THE UNITED STATES[1]

Authors: Jamie C. Freeman, Douglas H. Ross, and Jeffrey G. Scott

Abstract

Insecticide resistance in house fly populations is a major problem faced by livestock producers worldwide. A survey of insecticide resistance levels and pyrethroid resistance allele frequencies in the United States was conducted in 2008-09, but little is known about how resistance levels have changed over the last 10 years. In addition, new target-site pyrethroid resistance alleles that confer high levels of resistance have been recently identified in the *voltage-sensitive sodium channel*, and their frequencies in field populations are unknown. Our aim in this study was to reassess the resistance status of house flies from select locations in the United States by examining resistance levels against commonly used insecticides and frequencies of known resistance alleles. House flies were collected from animal production facilities in five different states between 2016 and 2018. Resistance levels to three insecticides (permethrin, tetrachlorvinphos, and methomyl), representing three classes of insecticides (pyrethroids, organophosphates and carbamates) varied geographically and were lowest in the population collected from New Mexico, intermediate in the population collected from Utah, and greatest in the population from Kansas. The recently identified *1B* pyrethroid resistance allele increased dramatically in frequency compared to previous reports, most notably in populations from Kansas and Maryland, indicating that it may already be widespread around the United States. Based on comparison

[1] Published: Freeman, J. C., Ross, D. H. & J. G. Scott. 2019. Insecticide resistance monitoring of house fly populations from the United States. *Pesticide Biochemistry and Physiology*. 158:61-68.

with historical data, the population collected from Kansas represents one of the most highly permethrin resistant populations ever sampled. If the alleles responsible for this level of resistance spread, pyrethroids may be of limited use for house fly control in the United States in the near future.

1. Introduction

The house fly, *Musca domestica*, is a worldwide livestock pest that is a mechanical vector of pathogens. House flies can transmit over 200 human and animal pathogens assisted by their close association with microbe-rich animal waste and wide ranging adult mobility [1]. The numerous avenues of house fly damage to livestock production make it difficult to assess the precise economic losses that occur due to house fly activity. For example, in poultry facilities house flies can vector two major diseases of chickens: colibacillosis caused by avian pathogenic *Escherichia coli* and necrotic enteritis caused by the bacterium *Clostridium perfringens*. Colibacillosis has been estimated to kill about 6% of the hens at egg facilities annually, with an annual financial cost to the facility of about $1.15 M (assuming 450,000 hens) [2]. Necrotic enteritis results in lesions in the chicken's intestine and can lead to flock mortality of 1% per day (clinical NE), with its estimated cost totaling $2.5 billion per year in the US [3]. In addition to the costs of chemical and other forms of control, house flies can decrease production through animal stress (especially to calves and chickens). House fly activities result in lowered levels of egg and milk production and reduced feed conversion [4].

Animal production facilities have relied extensively on insecticides (primarily pyrethroids and secondarily organophosphates) to control house flies and prevent disease transmission, but resistance has developed to all available insecticides and is a global problem. Effective resistance management strategies require two fundamental pieces of information. The first is knowledge about the levels of resistance to available insecticides. This data provides highly relevant information about which insecticides remain effective and

which ones should be avoided. However, in some cases this information is only assessed at a single (diagnostic) concentration (or dose), and determination of the most informative diagnostic concentration is not a trivial matter. The second necessary piece of information is understanding the frequency of insecticide resistance alleles in a population. Such data provides valuable information about the molecular evolution of resistance that cannot be obtained from diagnostic concentrations but requires that resistance alleles are known and periodically reconfirmed. Mutations in insecticide target site genes have been identified that confer resistance to pyrethroids (*Voltage sensitive sodium channel* (*Vssc*)) [5] and organophosphates/carbamates (*acetylcholinesterase* (*ace-2*) [6, 7]). However, resistance due to metabolic detoxification is also an important mechanism of resistance, and we have not identified all of the mutations responsible for this trait. Thus far, in house flies only the resistance allele of *CYP6D1* has been identified [8, 9]. Resistance due to decreased cuticular penetration has been described in house flies, but the levels of resistance conferred are low (<3-fold) [10], and the mutation responsible has not been identified.

Levels of insecticide resistance in house fly populations in the United States were surveyed in 2008-09 [11], but little new information has been published in the last decade. In addition, two new *Vssc* resistance alleles were identified since the last survey of resistance was conducted, bringing the total number to five: [12, 13]: *kdr-his* (L1014H), *kdr* (L1014F), *super-kdr* (M918T+L1014F), *Type N* (D600N+M918T+L1014F) and *1B* (T929I+L1014F). The levels of pyrethroid resistance these alleles confer is generally *kdr-his* < *kdr* < *Type N* ≤ *super-kdr* ≤ *1B* ≤ Type D [12]. However, the frequencies of these alleles in field populations are not well defined. *Type N* and *1B* alleles have only been recently identified from individuals in Kansas, so little is known about their geographic distribution.

While organophosphate and carbamate insecticides are also widely used against house flies, comparatively little is known about target-site resistance alleles in *ace-2*. Only one study has assessed the frequency of *ace-2* resistance alleles in the United States, using populations collected from Florida and New York in 2002 [14]. The incongruence of the organophosphate resistance levels of the populations studied with the frequency of *ace-2* resistance alleles indicated that, at least in those populations, other mechanisms are likely important in organophosphate resistance. Target-site mutations in *ace-2* have been confirmed to confer resistance *in vitro* [6, 7], but little is known about the contribution of each mutation to resistance *in vivo*, so monitoring of both the resistance levels and frequency of known mutations provides important information.

In this study, we were interested in knowing what changes had occurred relative to the last survey of house fly insecticide resistance in the United States in 2008-09 [11]. We did this first by characterizing the levels of resistance in house flies from animal production facilities to three insecticides that have been commonly used (permethrin, tetrachlorvinphos, and methomyl). Second, we evaluated the frequencies of known resistance alleles to three major classes of insecticides used for house fly control, namely *Vssc* and *CYP6D1* pyrethroid resistance alleles and *ace-2* organophosphate/carbamate resistance alleles**.** Our results suggest that permethrin resistance levels have become insurmountable in some populations and, that without the introduction of new insecticides, control of house fly populations could become problematic.

2. Materials and methods

2.1. House flies

Field collected house fly populations used in this study are shown in Table 1. Three populations (KS17, NM17, and UT17) were collected live to be used in both bioassays and genotyping, while three additional populations (KS18, MD17, and NE16) were collected and preserved in ethanol for genotyping. The IsoCS strain [15] was used as the susceptible strain for all bioassays. House flies were maintained at 45% relative humidity, with a 12:12 h light:dark and 29˚C:28˚C cycle. Adult house flies were provided powdered milk + granulated sugar (1:1 by volume) and water *ad libitum*. House fly larvae were reared on medium containing 2.3 L of water, 0.5 kg calf manna (Manna Pro Corp., St. Louis, MO), 90 g bird and reptile litter wood chips (Northeastern Products Corp., Warnersberg, NY), 50 g dried active baker's yeast (Lesaffre Yeast Corp., Milwaukee, WI), and 0.8 kg wheat bran (Star of the West Milling Corp., Churchville, NY).

Table 1

House fly collections used in this study.

Name	Collection location	Facility type	Collection date	Collection method	Used for bioassays?
KS17	Riley Co., KS	Poultry	June 2017	Live	Yes
KS18	Riley Co., KS	Dairy	August 2018	Preserved	No
MD17	Wicomico Co., MD	Poultry	June 2017	Preserved	No
NM17	Dona Ana Co., NM	Dairy	June 2017	Live	Yes
NE16	Lincoln, Co., NE	Feedlot	September 2016	Preserved	No
UT17	Millard Co., UT	Poultry	June 2017	Live	Yes

2.2. Insecticides and bioassays

Three insecticides, each representing a different class, were used in this study: a pyrethroid (permethrin (97%, Syngenta, Wilmington, DE)), an organophosphate (tetrachlorvinphos (99.5%, Chem Service Inc.)) and a carbamate (methomyl (99%, Chem Service Inc., Westchester, PA)).

For permethrin and tetrachlorvinphos, insecticide resistance was evaluated using a residual contact assay with diagnostic concentrations historically used for resistance monitoring: 234 ng/cm^2 for permethrin, and 670 ng/cm^2 for tetrachlorvinphos [16]. Diagnostic concentrations are commonly set at ~LC_{99} or 3 X LC_{99} of a susceptible strain, which is appropriate when trying to observe the initial evolution of resistance in a population. However, as resistance levels increase over time a higher monitoring concentration may become more useful in observing changes that better correlate with

control problems. As such, 10-fold higher monitoring concentrations were also used for permethrin (2340 ng/cm^2) and tetrachlorvinphos (6700 ng/cm^2) to gain additional information about the levels of resistance. Flies from the first through the sixth generation were used for bioassays. The IsoCS strain was used to confirm that the diagnostic concentrations were killing 100% of the susceptible strain as expected.

Residual contact bioassays were conducted as follows: 230 mL glass jars ((VWR, Radnor, PA) internal surface area is 180 cm^2) were coated with 1 mL of insecticide solution (in ACS reagent grade acetone (VWR)) or 1 mL of acetone for the controls. The acetone was allowed to evaporate for at least 30 min on a hot dog rolling machine (Gold Medal Products Co., Cincinnati, OH, USA, and 20 3–5-day-old females were transferred to each treated jar. A 2.5 cm dental wick (Richmond Dental, NC, USA) saturated with 15% (w/v) sugar water solution was supplied in each jar. Methomyl is commercially available as a bait formulation; therefore, resistance was evaluated via a feeding assay. Twenty 3-5 day-old female flies were held without access to food or water for two h and then placed in a 230 mL glass jar with two dental wicks saturated in a 15% sugar-water solution containing the desired concentration of methomyl (42 μg/mL).

For each diagnostic concentration of an insecticide, a minimum of 260 flies were tested, using flies from at least two different cages and over two or more days. Treated flies were held in a chamber kept at 25°C and 40% relative humidity with a 12:12 (L:D) photoperiod. Mortality was assessed after 48 h. Ataxic flies were considered dead. Percent survival data for populations was analyzed using logistic regression with a Tukey post-hoc test as implemented in the R package ‘multcomp’[17, 18].

For permethrin, we were also interested in comparing resistance of these new field-collected populations to other previously collected resistant populations assessed topically (as in [12, 19, 20]), so topical diagnostic dose bioassays were also conducted for permethrin. A 0.5-µl drop of insecticide in acetone solution (VWR) was applied to the thoracic notum of 3- 5-day-old female flies using a Hamilton PB-600 repeating dispenser equipped with a 25-µl Hamilton syringe (Hamilton, Reno, NV, USA). Controls were treated with acetone only. Each bioassay consisted of 20 flies per dose. All strains were initially tested against a series of 10-fold dilutions spanning a 10,000-fold concentration range, and at least three doses from this series were used for further bioassays on each strain. Treated flies were held in a chamber kept at 25°C and 40% relative humidity with a 12:12 (L:D) photo period. Each dose was replicated a minimum of three times using flies from at least two cages, over a minimum of 2 days. Flies received 15% (w/v) sugar water solution from a saturated 2.5 cm piece of cotton dental wick. Mortality, defined as flies that were ataxic, was assessed after 24 h.

2.3. Genotyping of resistance alleles

DNA was extracted from single hind legs of 32 individual female house flies from each population using an alkaline extraction method. Legs were placed in wells of a 0.2 mL 96-well plate (Laboratory Product Services, Rochester, NY, USA) with three 2.3-mm-diameter zirconia silica beads (Biospec Products, Bartlesville, OK, USA) and 10 µL of 0.2 M sodium hydroxide. The samples were homogenized on a vortex mixer for 1 min and then incubated at 70°C for 10 minutes. Next, 90 µL neutralization buffer (40 mM Tris-HCl, pH 7.5 and 1.1 mM EDTA) was added to each sample, followed by a final vortex mixing for 30

sec and a centrifugation for 30 sec to collect solution in the wells. Extracted DNA was stored at -20ºC.

PCR was carried out using 12.5 µL of GoTaq 2x (Promega, Madison, WI, USA), 9.5 µL nuclease free water, 1 µL template DNA, and 1 µL each of the 10 µM forward and reverse primers (Table 2). Flies were first genotyped for the 1014 site of the VSSC using the primers kdrFL and MdSCR7. For individuals with the L1014F mutation, additional regions containing the sites necessary to identify the *skdr*, *1B*, and *Type N* alleles were also sequenced using the primers listed in Table 2. For all three reactions, thermal cycler (Bio-Rad T100 (Bio-Rad, Hercules, CA)) conditions were as follows: 95 ºC for 3 min, followed by 35 cycles of PCR (95 ºC for 30 s, 55 ºC for 20 s and 72 ºC for 20s) and a final extension of 5 min at 72 ºC.

A fragment of *CYP6D1v1* was amplified with the primers Md6D1F1 and Md6DR2 to check for the presence of a 15-bp insert in the 5' UTR of the *CYP6D1* gene which has been associated with increased transcription of *CYP6D1* [8]. A fragment was amplified from *ace-2* with the primers MdAceF1 and MdAceR2 to examine 5 sites, V260L, A316S, G342A/V, and F407Y. For both reactions, thermal cycler conditions were as follows: 95 ºC for 3 min, followed by 35 cycles of PCR (95 ºC for 30 s, 55 ºC for 30 s and 72 ºC for 30s) and a final extension of 5 min at 72 ºC.

PCR product size was confirmed on a 1.5% agarose gel. The remaining unconsumed primers and dNTPs in the PCR mixture were removed by incubating 5 µL of PCR product with 1 µL of a reaction mixture containing 0.5 µL FastAP alkaline phosphatase (Thermo Fisher Scientific, Waltham, MA, USA), 0.05 µL Exonuclease I (Thermo Fisher Scientific), 0.05 µL Exo I buffer (Thermo Fisher Scientific) and 0.4 µL nuclease free water at 37ºC for

30 min, followed by 15 min at 85ºC. Sanger sequencing of the PCR products was performed by Cornell's Biotechnology Resource Center using the primers indicated in Table 2. Electropherograms were manually inspected for resistance mutations and edited using Chromas Lite ver. 2.01 (Technelysium Pty Ltd, South Brisbane, Australia) and the EditSeq and MegAlign applications of Lasergene (DNA Star, Madison, WI, USA). For *Vssc* and *CYP6D1*, genotype frequencies were assessed against those expected in Hardy-Weinberg equilibrium using a χ^2 test (α=0.05).

Table 2

Primers used in this study.

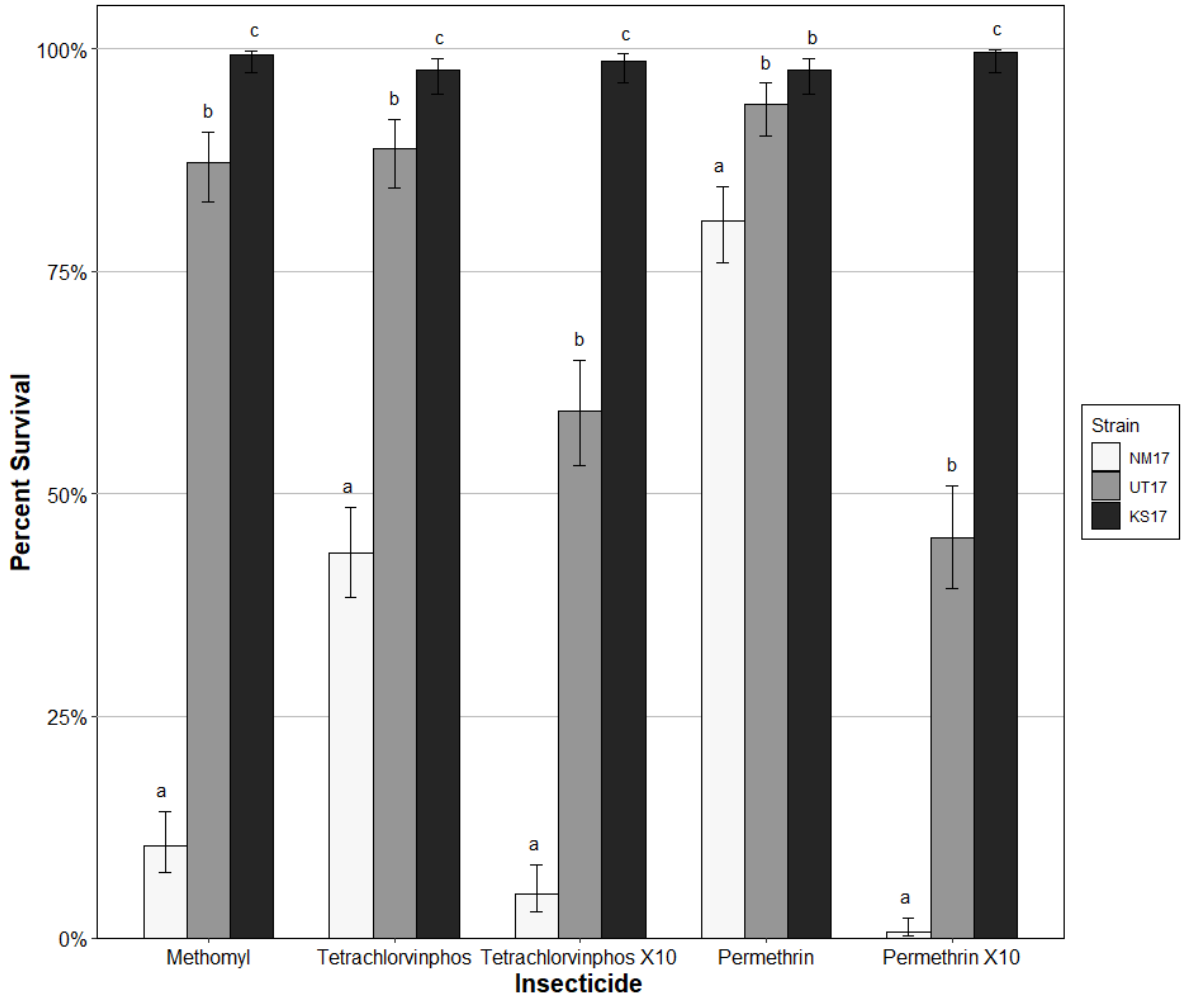
Gene	Sites genotyped	Primer name	Sequence	Sequencing primer?
Vssc	L1014H/F	kdrFL	TCGCTTCAAGGACCATGAATTACCGCGCTG	Yes
	L1014H/F	MdSCR7	TGGTATCATTGTCGGCAGTC	
	M918T, T929I	MdSCF52	GCAAAATCATGGCCCACACT	
	M918T, T929I	MdSCR3	GTTCTTTCCGAAAAGTTGCATTCC	Yes
	D600N	MdSCF61	AATACGAAATGGGCGTGGAC	Yes
	D600N	MdSCR62	CATTCTCTTCGGACATTGGTG	
CYP6D1	5’ UTR deletion	Md6D1F1	CCGTCATTTACAACGCATTAGG	Yes
	5’ UTR deletion	Md6D1R2	ACCTTCTCGTGGCATTTGTC	
Ace	V260L, A316S, G342A/V, F470Y	MdAceF1	CCGTCATTTACAACGCATTAGG	Yes
	V260L, A316S, G342A/V, F470Y	MdAceR1	ACCTTCTCGTGGCATTTGTC	

3. Results

3.1. Residual diagnostic concentration bioassays

Based on the diagnostic concentrations for all three insecticides, the lowest resistance levels were found in the NM17 population and higher levels were found in UT17 and KS17 (Figure 1). The NM17 strain is much more susceptible to methomyl and tetrachlorvinphos having only 10.3% and 43.3 % survival, respectively, while the UT17 and KS17 strains showed >85% survival to both insecticides (with the highest percent survival in KS17). For permethrin, the populations responded similarly with > 75% survival and no significant difference between UT17 and KS17 using the historical diagnostic concentration. The 10X diagnostic concentrations for permethrin and tetrachlorvinphos made results between the populations more stratified and resulted in a clear distinction between all three strains. Survival of the KS17 populations did not dramatically change between the 1X and 10X concentrations of either insecticide, but NM17 and UT17 survival decreased 8.6- and 1.6-fold for tetrachlorvinphos and 98- and 2.2-fold for permethrin.

Fig. 1. Levels of resistance to three insecticides in house flies from three populations in the USA. For tetrachlorvinphos and permethrin, an additional diagnostic dose at 10X the original concentration was also tested. Bars with different letters are significantly different ($p \leq 0.05$) within each insecticide (generalized linear model with logit link function followed by Tukey's HSD test). Bars represent the mean survival at the diagnostic concentration and error bars represent 95% confidence intervals. At least 260 individuals were tested for each population and insecticide concentration.



3.2. Topical permethrin diagnostic dose bioassays

Similar to the diagnostic concentration assays, the permethrin topical bioassays revealed different permethrin resistance levels across the populations when different diagnostic doses were used (Table 3). At a dose of 0.1 µg/fly all three field collected populations appeared equally resistant. However, at the higher doses it became clear that the populations were very different, with extremely high permethrin resistance in the KS17 strain (with 97% survival at 10 µg/fly).

Table 3

Mean percent survival ± SE for permethrin topical bioassays. At least 3 replicates (20 individuals per replicate) were tested for each strain at each dose.

	Permethrin dose (µg/fly)					
Strain	Control	0.001	0.01	0.1	1.0	10
IsoCS	98.8 ± 0.3	97.0 ± 0.5	54.0 ± 1	0.0 ± 0	*	*
KS17	100 ± 0	*	*	96.7 ± 0.6	100 ± 0	97.0 ± 0.7
UT17	96.7 ± 0.3	*	*	98.0 ± 0.2	85.0 ± 1	23.0 ± 1
NM17	100 ± 0	*	99.0 ± 0.2	96.0 ± 0.2	26.0 ± 1	0.0 ± 0

*Strain not tested at this dose.

3.3. Genotyping of resistance alleles

The frequencies of the *Vssc* pyrethroid resistance alleles varied widely between collections (Figure 2), while the *CYP6D1v1* resistance allele was at a frequency of >75% in

all populations (Figure 3). In terms of *Vssc* alleles, NM17 had the highest frequency of susceptible alleles (71.9%), with lower frequencies of *kdr-his* (12.5%) and *kdr* (15.6%). NE16 had a high frequency of susceptible alleles (46.9%), intermediate frequencies of *kdr-his* (28.1%) and *kdr* (18.8%) and a low frequency of *skdr* (6.3%) and *1B* (1.6%). The predominant allele in UT17 alleles was *kdr* (76.6%), with *kdr-his* making up the rest of the population (23.4%). In MD17 the most common allele was *1B* (45.3%), followed by *kdr-his* (28.1%), *kdr* (21.9%), and *skdr* (4.7%). The predominant alleles in KS17 were *skdr* (56.3%) and *1B* (35.9%), with a low frequency of *kdr* (7.8%). No susceptible alleles were detected in UT17, MD17, or KS17. The KS18 collection, like KS17, showed high frequencies of *skdr* (44.6%) and *1B* (19.6%) and a low frequency of *kdr* (8.9%). Unlike KS17, the KS18 collection had an intermediate frequency of *kdr-his* (19.6%), and a low frequency of susceptible alleles (7.1%). The *Type N* allele was not detected in any population. The genotype frequencies for *Vssc* did not differ significantly from those expected under Hardy-Weinberg equilibrium (p-values: NM17 0.54, NE16 0.08, UT17 0.08, MD17 0.51, KS17 0.14, KS18 0.43). For the populations tested in the bioassays, the *Vssc* genotyping results are in strong agreement with the relative levels of permethrin resistance.

The frequency of the resistance allele *CYP6D1v1* was uniformly high in all populations tested (Figure 3), with frequencies ranging from 75.0% in NE16 to 98.3% in KS17. The genotype frequencies for *CYP6D1v1* did not differ significantly from those expected under Hardy-Weinberg equilibrium (p-values: NM17 0.49, NE16 0.35, UT17 0.46, MD17 0.63, KS17 0.72). MD17 had greater frequency of the *CYP6D1v1* allele than UT17 (92.2% versus 76.6%) although the level of permethrin resistance in MD17 was less than that of UT17.

Fig. 2. *Vssc* allele frequencies in populations from the US collected in 2016-18. The D600N allele was also examined, but not detected in any of the flies genotyped (32 in each population).

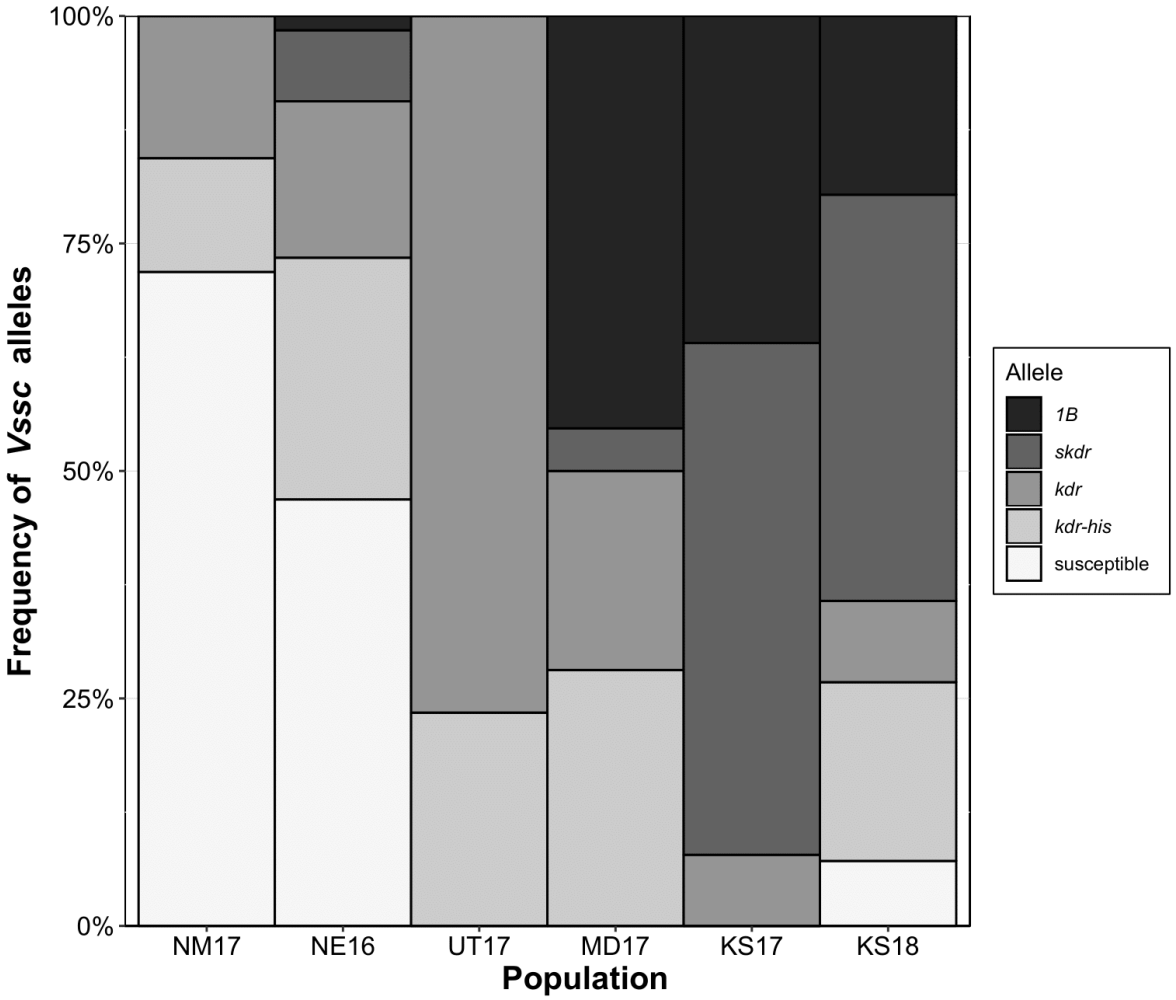
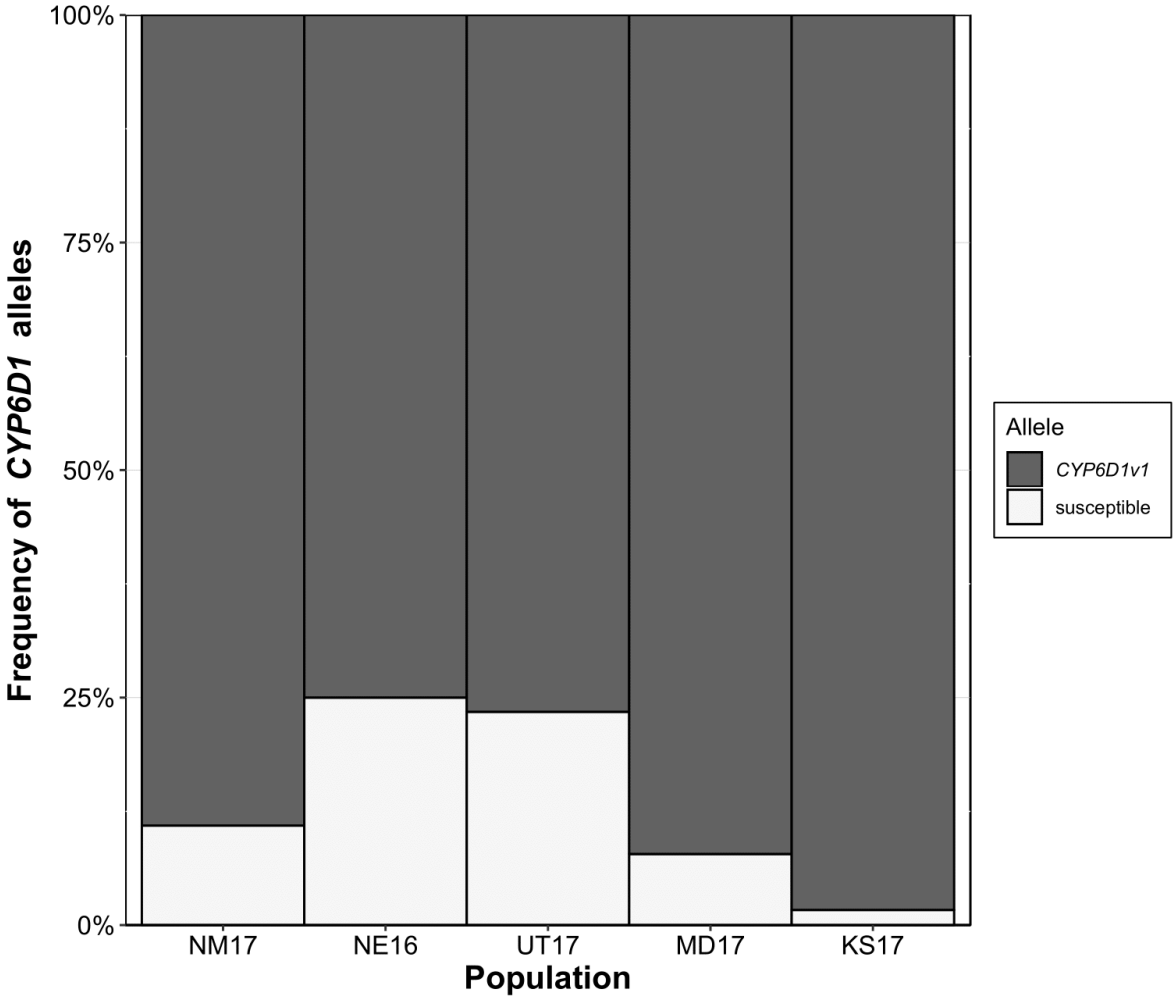


Fig. 3. *CYP6D1v1* allele frequencies across 5 populations collected in 2006-07. From each site, 32 flies were genotyped.



The *skdr*, *1B*, and D600N alleles include L1014F plus additional mutations. The M918T mutation of the *skdr* allele has never been found without L1014F, and this held true; all 44 individuals with M918T in this study also had the L1014F mutation. Similarly, the T929I mutation of the *1B* allele was always found with L1014F, but never M918T (n=53). L1014H has never been found in combination with any other mutation. Of the 58 *kdr-his*

alleles, 44 were checked for the M918 and T929 sites, and no other mutations were observed in this region of those alleles.

Sequencing results from *ace-2* were ambiguous due to the observation of three nucleotides for some sites, which made calculations of allele frequencies impossible with the sequencing data from this study. In house flies, as in *Drosophila melanogaster*, only one *ace* gene (*ace-2*) is believed to be present [21]. Four individuals, three from NM17 and one from UT17 had three different amino acids at site 342: the susceptible G342, and the resistance alleles G342A and G342V (Supplementary Fig. 1a). One additional individual from UT17 had 3 nucleotides for a single site in an intronic region (Supplementary Fig. 1b). Our results suggest a possible duplication of *ace-2*, but future studies will be necessary to resolve this issue. While allele frequencies could not be calculated, the number of observations of each resistance allele is reported for the populations sequenced (Table 4).

Table 4

Presence of individuals having *ace-2* alleles with known resistance mutations in three in populations of house flies, with allele counts. Individuals with three nucleotides at any site were excluded from allele counts.[a]

	Mutation (frequency)				
Population	V260L	A316S	G342A	G342V	F470Y
NM17	No	No	Yes (1/58)	Yes (3/58)	Yes(4/58)
UT17	No	Yes (30/60)	Yes (30/60)	No	Yes (35/60)
KS17	No	Yes (37/60)	Yes (56/60)	No	Yes (56/60)

[a]Three *ace-2* alleles were found for five individuals (see section 3.3), indicating a possible duplication of *ace-2* (see section 3.3).

4. Discussion

A comparison of the percent survival at the standard diagnostic concentrations in the flies collected in 2017 from New Mexico and Kansas (NM17 and KS17) relative to what was reported for collections made in 2008-09 (NM08 and KS09, [11]) revealed some interesting differences. Compared to the 2008-09 season, percent survival to tetrachlorvinphos was greater in 2017, increasing from 30.6% in NM08 to 43.3% in NM17 and from 78.4% in KS09 to 97.8% in KS17. Similarly, the percent survival to permethrin was also greater in 2017 relative to the 2008-09 collections, increasing from 67.0% in NM08 to 78.0% in NM17 and from 93.5% in KS09 to 98.0% in KS17. Methomyl survival was high in both KS09 and KS17 (95.2% and 99.3%, respectively), but interestingly, in New Mexico, methomyl survival dropped from 70.3% in NM09 to 10.3% in NM17. This may reflect a decreased use of methomyl in New Mexico resulting from the registration of house fly baits utilizing other active ingredients (imidacloprid in 2004, cyantraniliprole in 2014, and dinotefuran in 2015). This would be consistent with an observation of a reduction in resistance to methomyl found in a previous study of house flies from California where decreased methomyl usage for fly control has been reported anecdotally since the mid-1990s [22]. However, this is speculative as records of insecticide use were not available from the facilities we studied. Since this is the first report of the insecticide resistance status of house flies from Utah, we could not compare it to previous years. However, comparing the percent survival of UT17 against all three insecticides to the overall results from 2008-09 puts it among the states with the greatest percent survival against each insecticide.

Diagnostic concentrations have been useful in monitoring the initial evolution of resistance, but as the levels of resistance increase it is important to reassess monitoring methodology. The initial evolution of resistance to a novel compound is commonly monitored using a diagnostic concentration of about 3 x the LD_{99} of the susceptible strain. In the case of permethrin this was spectacularly successful as populations were easily observed to change following its introduction ([23] vs [16]). However, survival at this diagnostic concentration need not reflect the levels of resistance, nor issues of control failure. This can be readily seen in Figure 1. The three populations are clearly different when the diagnostic concentration is increased 10-fold, while the differences are quite muted at the standard diagnostic concentration. A similar result was noted for the two concentrations of tetrachlorvinphos. These results clearly demonstrate the need to periodically re-evaluate the diagnostic concentrations that are being used for resistance monitoring so that the most valuable information is obtained.

Diagnostic concentration (or dose) bioassays are a quick and simple method of assaying change over time that require less animals than the determination of the population LD_{50}. Monitoring high levels of resistance through residual contact methods can require prohibitively expensive amounts of insecticides, so topical bioassays may be necessary. Going forward, the levels of permethrin resistance in the United States make topical application methods more cost effective.

Although the residual contact diagnostic dose assays indicated a modest increase in permethrin resistance levels since the 2008-09 survey, a larger number of comparisons are available using the permethrin topical diagnostic dose assays. Table 4 shows a summary of percent survival of female flies from previous studies where other field-collected strains

have been evaluated by topical application with permethrin. Comparing the KS17 results (Table 3) with previous results (Table 4), it is clear that KS17 represents one of the most highly permethrin resistant populations of house flies ever sampled. Though high levels of permethrin resistance have been found in other field-collected strains, highly resistant individuals generally make up only a small proportion of the population, and laboratory selection is often necessary to detect them. The frequency of highly resistant individuals within the KS17 populations (without any laboratory selections) poses real questions as to the current and future efficacy of pyrethroid insecticides against house flies.

For the NM17, UT17, and KS17 populations, the permethrin survival in both the residual contact and topical bioassays agreed with the *Vssc* genotyping results, but *CYP6D1v1* allele frequencies did not predict resistance levels. The frequency of *CYP6D1v1* was lower in UT17 than in NM17 (76.6% versus 89.1%), although the percent survival against permethrin was greater in UT17. Resistance due to the *CYP6D1v1* allele is nearly dominant [24], so *CYP6D1v1* heterozygotes are not at a severe disadvantage in the presence of permethrin and all the populations surveyed may have approximately the same levels of *CYP6D1* mediated resistance. It is important to note that the *CYP6D1v1* allele was only partially responsible for increased expression of *CYP6D1* in the LPR strain and a second *trans* regulatory factor on autosome 2 remains unidentified [25, 26]. In addition, there is significant evolutionary plasticity in the CYPs that evolve to confer insecticide resistance [27], and other CYPs (house flies have 146 *CYPs* [28]) are likely to be involved in resistance in some populations. Furthermore, it is possible that new resistance alleles have evolved in genes not currently known to be involved in resistance. These are important areas requiring future study.

The frequencies of *Vssc* resistance alleles were variable between sites, as can be seen in Figure 2, although *CYP6D1v1* frequencies were less variable (Figure 3). *Vssc* susceptible alleles were only found in three populations: NM17, NE16, and KS18. Every population except KS17 contained *kdr-his* alleles, and *kdr* was the only allele present in all populations. NE16, MD17, KS17, and KS18 contained *Vssc* alleles that confer the greatest level of resistance (*skdr* and *1B* [13]), and these "super" resistance alleles made up at least half of the genotyped alleles in MD17, KS17, and KS18. The third known "super" resistance allele, *Type N*, was not present in any of the individuals genotyped for this study. These results represent a shift in *Vssc* allele frequencies as compared to 2008-09 (Supplementary Figure 2). In contrast to the considerable shift in *Vssc* allele frequencies, *CYP6D1v1* frequencies did not exhibit much change between 2008-09 and 2016-17, which is congruent with previous data from FL, NY, and NC showing relatively stable frequencies from 2002-03 to 2008-09 [11, 29].

Given the high frequencies of *1B* and *skdr* resistance alleles in the KS17 strain collected from a poultry house, we wished to examine if this was limited to that facility or not, as previous studies have suggested house flies at poultry houses can be genetically isolated from those in surrounding areas [30]. Therefore, flies were collected from a dairy in the same county in Kansas in 2018 (KS18). The KS18 flies had a lower frequency of *Vssc* resistance alleles than the KS17 population (Figure 2). However, the KS18 population also had high frequencies of *skdr* and *1B* (44.6% and 19.6% respectively), which indicates that although the poultry house originally sampled for the KS17 flies does represent an extreme case, the regional population also has frequencies of the *1B* and *skdr* alleles that are higher than previously seen.

The *1B* allele was recently discovered (after permethrin selection of a population collected in Kansas in 2013 [12]) and has only been found previously at low frequencies. Its frequency was assessed in 88 individuals from Riley Co., Kansas in 2015 and only one individual heterozygous for the *1B* allele (1.1%) was found [12]. Upon reexamination of the 2008-09 *Vssc* sequencing data it was also found to be present in one individual from New York [12], which indicated it may have already have wide geographic spread in the US. This is supported by the high frequency of the *1B* allele in MD17 (45.3%). Resistance to permethrin (and several other pyrethroids) did not vary significantly between congenic strains with the *skdr* and *1B* alleles [13], so the rise in prominence of the *1B* allele over the *skdr* allele presents an interesting puzzle. Such an extreme shift in allele frequency over just a handful of years would require an intense selective pressure. This pressure does not seem to be due to its ability to confer resistance alone, though potentially lower relative fitness costs of *1B* relative to *skdr* would hasten its spread.

Overall, there is considerable variation in resistance levels and *Vssc* resistance alleles between populations. The frequencies of highly resistant individuals in KS17 represent a potential loss of utility of pyrethroid insecticide sprays for control of house flies in some areas. Monitoring the spread of newly discovered "super" *Vssc* resistance allele *1B*, and investigations into its potential fitness benefits over *skdr* will provide understanding into the impressive shift of some populations toward the *1B* allele. New Mexico seems to be maintaining susceptible alleles despite prolonged insecticide use and areas of concentrated dairy production. Further work to identify the landscape of population frequencies of resistance alleles in New Mexico as compared to more resistant neighboring states may provide insight into strategies for controlling the spread of insecticide resistance alleles.

Table 5

Examples of dose of permethrin and percent survival of female flies for previous field-collected strains of house fly. For studies where >2 populations from the same county or province were tested, data are shown for only two: the populations with the maximum survival and a population having the median survival for the collections tested.

Strain/ Collection	Collection year	Collection location	Facility type	Permethrin dose (µg/fly)	Female % survival	Reference
Learn	1980	USA: Chemung Co., NY	Dairy	0.04	50%	[31]
Georgia	1982	USA: Barrow Co., Georgia	Poultry	0.181	50%[a]	[32]
J1	1984	USA: San Bernardino Co., CA	Dairy	9.25	50%[a]	[33]
IX	1990	Hungary: ?	Swine	12.9	50%[a]	[34]
V	1990	Hungary: ?	Swine	0.3	50%[a]	[34]
YPER	1997	Japan: Third Yumenoshima Island	Dump	0.5	46%	[35]
NG98-u	1998	USA: Newton Co., GA	Poultry	13.3	50%[a]	[20]
ALHF	1998	USA: Marshall Co., AL	Poultry	3	15%	[36]
NY-pre04	2003	USA: Schuyler Co., NY	Dairy	10	2.2%[b]	[19]

Strain/ Collection	Collection year	Collection location	Facility type	Permethrin dose (µg/fly)	Female % survival	Reference
FL-late03	2003	USA: Alachua Co., FL	Dairy	10	0.36%[b]	[19]
GD	2009	China: Guangzhou, Guangdong Province	Dump	0.072	50%[a]	[37, 38]
SH	2009	China: Feng Xian, Shanghai Municipality	Swine	0.051	50%[a]	[37, 38]
SD	2009	China: Jinan, Shandong Province	Dump	0.051	50%[a]	[37, 38]
BJ	2009	China: Chaoyang, Beijing Municipality	Dump	0.035	50%[a]	[37, 38]
JL	2009	China: Changchun, Jilin Province	Dump	0.019	50%[a]	[37, 38]
KS13	2013	USA: Riley Co., KS	Dairy	1.28	48%	[12]
LHR	2017	Pakistan: Punjab, Lahore	Urban area	0.20[c]	50%[a]	[39]
ISB	2017	Pakistan: Punjab, Islamabad	Urban area	0.11[c]	50%[a]	[39]

Acknowledgements

We thank D. Boxler and L. Zurek for collections of house flies used in this study. This project was supported by Bayer CropScience (Research Triangle Park, NC, USA).

REFERENCES

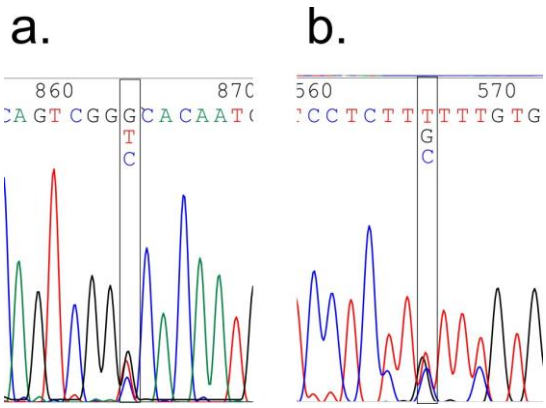
[1] D. Nayduch, R.G. Burrus, Flourishing in filth: House fly–microbe interactions across life history, Ann. Entomol. Soc. Amer., 110 (2017) 6-18.
[2] Zoetis, Poulvac, 2018.
[3] B. Wade, A. Keyburn, The true cost of necrotic enteritis, 2015.
[4] C.J. Geden, J.A. Hogsette, Research and extension needs for integrated management programs for livestock and poultry, 2001.
[5] J.G. Scott, Evolution of pyrethroid resistance in *Musca domestica*, Pest Manag. Sci., 73 (2017) 716-722.
[6] T. Kozaki, T. Tomita, K. Taniai, M. Yamakawa, Y. Kono, Expression of two acetylcholinesterase genes from organophosphate sensitive- and insensitive- houseflies, *Musca domestica* L. (Diptera: Muscidae), using a baculovirus insect cell system, Appl. Enotmol. Zool., 37 (2002) 213-218.
[7] S.B. Walsh, T.A. Dolden, G.D. Moores, M. Kristensen, T. Lewis, A.L. Devonshire, M.S. Williamson, Identification and characterization of mutations in housefly (*Musca domestica*) acetylcholinesterase involved in insecticide resistance, Biochem. J., 359 (2001) 175-181.
[8] J. Gao, J.G. Scott, Role of the transcriptional repressor *mdGfi-1* in *CYP6D1v1*-mediated insecticide resistance in the house fly, *Musca domestica*, Insect Biochem. Mol. Biol., 36 (2006) 387-395.
[9] J. Seifert, J.G. Scott, The *CYP6D1v1* allele is associated with pyrethroid resistance in the house fly, *Musca domestica*, Pestic. Biochem. Physiol., 72 (2002) 40-44.
[10] R.F. Hoyer, F.W. Plapp, Jr., Insecticide resistance in the house fly: identification of a gene that confers resistance to organotin insecticides and acts as an intensifier of parathion resistance, J. Econ. Entomol., 61 (1968) 1269-1276.
[11] J.G. Scott, C.A. Leichter, F.D. Rinkevich, S.A. Harris, C. Su, L.C. Aberegg, R. Moon, C.J. Geden, A.C. Gerry, D. Taylor, R.L. Byford, W. Watson, G. Johnson, D. Boxler, L. Zurek, Insecticide resistance in house flies from the United States: Resistance levels and frequency of pyrethroid resistance alleles, Pestic. Biochem. Physiol., 107 (2013) 377-384.
[12] S. Kasai, H. Sun, J.G. Scott, Diversity of knockdown resistance alleles in a single house fly population facilitates adaptation to pyrethroid insecticides, Insect Mol. Biol., 26 (2017) 13-24.
[13] H. Sun, S. Kasai, J.G. Scott, Two novel house fly *Vssc* mutations, D600N and T929I, give rise to new insecticide resistance alleles, Pestic. Biochem. Physiol. , 143 (2017) 116-121.
[14] T. Kozaki, S.G. Brady, J.G. Scott, Frequencies and evolution of organophosphate insensitive acetylcholinesterase alleles in laboratory and field populations of the house fly, *Musca domestica* L., Insect Mol. Biol., 95 (2009) 6-11.
[15] R.L. Hamm, J.-R. Gao, G.G.-H. Lin, J.G. Scott, Selective advantage for III^M males over Y^M males in cage competition, mating competition and pupal emergence in *Musca domestica* L. (Diptera: Muscidae), Environ. Entomol., 38 (2009) 499-504.
[16] P.E. Kaufman, J.G. Scott, D.A. Rutz, Monitoring insecticide resistance in house flies (Diptera: Muscidae) from New York dairies, Pest Manag. Sci., 57 (2001) 514-521.
[17] T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models, Biom J, 50 (2008) 346-363.

[18] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2018.
[19] F.D. Rinkevich, R.L. Hamm, C.J. Geden, J.G. Scott, Dynamics of insecticide resistance alleles in two different climates over an entire field season, Insect Biochem. Mol. Biol., 37 (2007) 550-558.
[20] S. Kasai, J.G. Scott, Overexpression of cytochrome P450 CYP6D1 is associated with monooxygenase-mediated pyrethroid resistance in house flies from Georgia, Pestic. Biochem. Physiol., 68 (2000) 34-41.
[21] E. Huchard, M. Martinez, H. Alout, E.J. Douzery, G. Lutfalla, A. Berthomieu, C. Berticat, M. Raymond, M. Weill, Acetylcholinesterase genes within the Diptera: takeover and loss in true flies, Proc Biol Sci, 273 (2006) 2595-2604.
[22] A.C. Murillo, A.C. Gerry, N.T. Gallagher, N.G. Peterson, B.A. Mullens, Laboratory and field assessment of cyantraniliprole relative to existing fly baits, Pest Man. Sci., 71 (2015) 752-758.
[23] J.G. Scott, R.T. Roush, D.A. Rutz, Insecticide resistance of house flies from New York dairies (Diptera: Muscidae), J. Agric. Entomol., 6 (1989) 53-64.
[24] N. Liu, J.G. Scott, Inheritance of CYP6D1-mediated pyrethroid resistance in house fly (Diptera: Muscidae), J. Econ. Entomol., 90 (1997) 1478-1481.
[25] N. Liu, J.G. Scott, Increased transcription of *CYP6D1* causes cytochrome P450-mediated insecticide resistance in house fly, Insect Biochem. Molec. Biol., 28 (1998) 531-535.
[26] G.G.-H. Lin, J.G. Scott, Investigations of the constitutive overexpression of CYP6D1 in the permethrin resistant LPR strain of house fly (*Musca domestica*), Pestic. Biochem. Physiol., 100 (2011) 130-134.
[27] J.G. Scott, S. Kasai, Evolutionary plasticity of monooxygenase-mediated resistance, Pestic. Biochem. Physiol., 78 (2004) 171-178.
[28] J.G. Scott, W.C. Warren, L.W. Beukeboom, D. Bopp, A.G. Clark, S.D. Giers, M. Hediger, A.K. Jones, S. Kasai, C.A. Leichter, M. Li, R.P. Meisel, P. Minx, T.D. Murphy, D.R. Nelson, W.R. Reid, F.D. Rinkevich, H.M. Robertson, T.B. Sackton, D.B. Sattelle, F. Thibaud-Nissen, C. Tomlinson, L. van de Zande, K.K.O. Walden, R.K. Wilson, N. Liu, Genome of the house fly (*Musca domestica* L), a global vector of diseases with adaptations to a septic environment, Genome Biol. , 15 (2014) 466.
[29] F.D. Rinkevich, L. Zhang, R.L. Hamm, S.G. Brady, B.P. Lazzaro, J.G. Scott, Frequencies of the pyrethroid resistance alleles of *Vssc1* and *CYP6D1* in house flies from the eastern United States, Insect Mol. Biol., 15 (2006) 157-167.
[30] J.G. Scott, T.G. Alefantis, P.E. Kaufman, D.A. Rutz, Insecticide resistance in house flies from caged-layer poultry facilities, Pest Manag. Sci., 56 (2000) 147-153.
[31] J.G. Scott, G.P. Georghiou, Rapid development of high-level permethrin resistance in a field-collected strain of house fly (Diptera: Muscidae) under laboratory selection, J. Econ. Entomol., 78 (1985) 316-319.
[32] D.L. Horton, D.C. Sheppard, Nolan, Jr., M.P, R.J. Ottens, J.A. Joyce, House fly (Diptera: Muscidae) resistance to permethrin in a Georgia caged layer poultry operation, J. Agric. Entomol., 2 (1985) 196-199.
[33] J.A. Meyer, G.P. Georghiou, M.K. Hawley, House fly (Diptera: Muscidae) resistance to permethrin on southern California dairies, J. Econ Entomol., 80 (1987) 636-640.

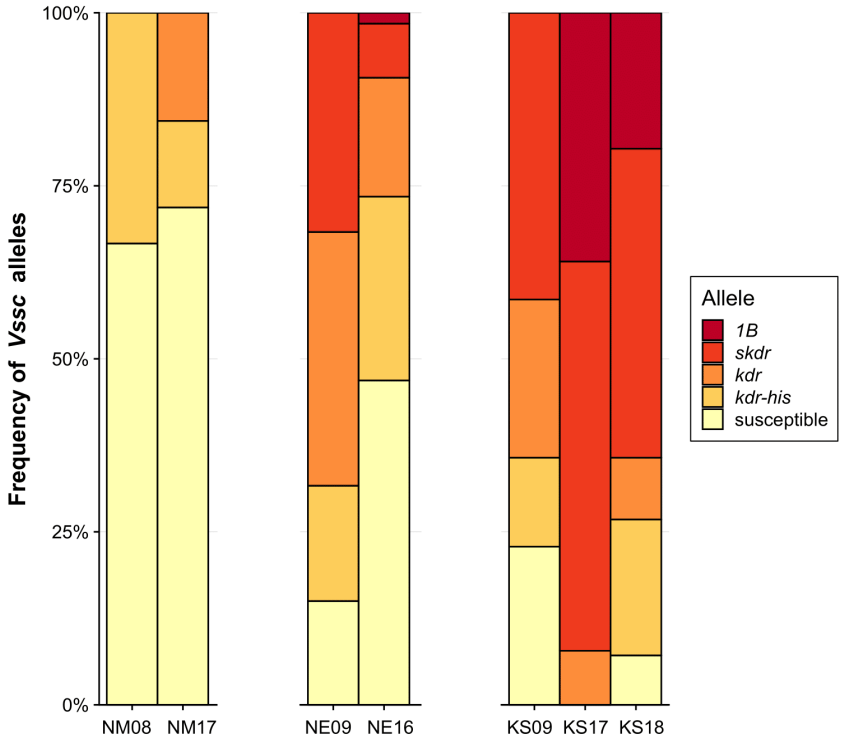
[34] L. Pap, R. Farkas, Monitoring of resistance of insecticides in house fly (*Musca domestica*) populations in Hungary, Pestic. Sci., 40 (1994) 245-258.
[35] T. Shono, S. Kasai, E. Kamiya, Y. Kono, J.G. Scott, Genetics and mechanisms of permethrin resistance in the YPER strain of house fly, Pestic. Biochem. Physiol., 73 (2002) 27-36.
[36] N. Liu, X. Yue, Insecticide resistance and cross-resistance in the house fly (Diptera: Muscidae), J. Econ. Entomol., 93 (2000) 1269-1275.
[37] Q. Wang, M. Li, J. Pan, M. Di, Q. Liu, F. Meng, J.G. Scott, X. Qiu, Diversity and frequencies of genetic mutations involved in insecticide resistance in field populations of the house fly (*Musca domestica* L.) from China, Pestic. Biochem. Physiol., 102 (2012) 153-159.
[38] Q. Gao, M. Li, C. Sheng, J.G. Scott, X. Qiu, Multiple cytochrome P450s overexpressed in pyrethroid resistant house flies (*Musca domestica*), Pestic. Biochem. Physiol., 104 (2012) 252-260.
[39] H.A.A. Khan, W. Akram, A. Fatima, Resistance to pyrethroid insecticides in house flies, *Musca domestica* L., (Diptera: Muscidae) collected from urban areas in Punjab, Pakistan, Parasitol. Res., 116 (2017) 3381-3385.

SUPPLEMENTAL INFORMATION

Supplementary Fig 1. Example chromatograms for *ace-2* highlighting individuals with three alleles at the 342 resistance site (a) and an intronic site (b).



Supplementary Fig. 2. *Vssc* allele frequencies for current and historical collections, for counties with historical data. KS17, KS18, NE16, and NM17 populations from this study and 2008-09 data from [11].



CHAPTER 2
COMPLEX STRUCTURAL VARIATION ON CHROMOSOME 5 IS LINKED TO PYRETHROID RESISTANCE IN *MUSCA DOMESTICA*

Abstract

Increased detoxification of insecticide is a common method of insecticide resistance. Despite this commonality, the molecular changes that are responsible are for are generally poorly understood. Often little is known beyond the involvement of particular enzyme families in the resistance phenotype or some number of overexpressed detoxification genes that may be involved. The availability of forward genetic studies in pest species will prove invaluable to advancing our understanding of resistance and insecticide toxicology in general. We use a quantitative trait mapping method called bulk segregant analysis (BSA) to determine the basis of a multigenic pyrethroid resistance phenotype to the known target-site of pyrethroids, the *voltage-gated sodium channel*, and a cluster of *cytochrome P450s* (*CYPs*) on chromosome 5. Structural variant calling using long read sequences indicates the area linked to resistance on chromosome 5 is structurally complicated in the resistant strain, and RNA-seq data indicates overexpression of 8 CYPs at this locus. The BSA mapping results are supported by linkage mapping of resistance with known morphological makers and synergist bioassays that indicate involvement of esterase and CYP-mediated detoxification in the resistance phenotype.

2.1 Introduction

Pyrethroid insecticides were first introduced for house fly control in the 1980s, and pyrethroid resistance was rapidly reported (partially due to cross-resistance with DDT resistance [1]). Despite this, pyrethroids are still the one of the most commonly used classes of insecticides for control at animal production facilities in the United States [2]. Previously, we reported a collection of house flies from Kansas (KS17) with previously unseen levels of pyrethroid resistance [3], indicating that at least in some locations, control of house flies with pyrethroids is no longer possible. Spread of the causal mutations could results in widespread control problems.

The two most commonly reported mechanisms of pyrethroid resistance in house flies are mutations in the target-site (the voltage-sensitive sodium channel, *Vssc*) and increased detoxification by cytochrome P450s (CYPs) [4]. Additionally, decreased penetration of pyrethroids through the exterior cuticle of the fly [5, 6] and increased detoxification by esterases [7-9] have been reported. Identification of these resistance mechanisms is facilitated by knowing the target-site gene, comparison of topical vs. injected resistance levels, and by the use of enzyme inhibitors (e.g. insecticide synergists). The identification of other mechanisms of resistance is far more difficult. By far, target-site alterations are best understood. It is clear that metabolic resistance is common mostly due to evidence provided by synergists, and while this can indicate involvement of particular metabolic gene families in resistance the particular genes involved are generally not known. There is a similar knowledge gap for other classes of insecticides and in most pest species.

Resistance due to increased detoxification by CYPs generally seems to occur through overexpression, and has been implicated in insecticide resistance in the house fly through both *cis* and *trans* regulatory changes (upregulation of *CYP6D1v1* in the Learn Pyrethroid Resistant (LPR) strain [10]. Changes in expression of *CYPs* can also result from copy number variation

[11, 12]. There is one report of a single nucleotide polymorphism (SNP) in a CYP that resulted in DDT-resistance [13]. The large number of CYP genes in most insect genomes (146 in house flies [14]), makes it difficult to identify mutations causing CYP mediated resistance.

Like CYPs, esterases have mostly been implicated in pyrethroid resistance though synergist data (with *S,S,S*-tributylphosphorotrithioate (DEF)) [7, 9]. There is some evidence that DEF has some non-specific activity against oxidases as well (it can inhibit oxidases at high concentrations *in vitro* [15] and has been shown to synergize resistance in fipronil which has no ester bonds [16, 17]). Carboxylesterase activity from crude extract of house fly abdomens increased in response to selection with permethrin or trichlorphon (an organophosphate) [18]. An allele of *MdαE7* is known to increase hydrolosis of organophosphates [19], but nothing is known about its potential effect on pyrethroid detoxification. Overexpression of 11 esterases was found in a strain with DEF-suppressible resistance relative to two unrelated susceptible strains [20], but overexpression alone is not sufficient to implicate these esterases in resistance; the overexpression must be genetically linked to the resistance phenotype.

There are also other gene families that may be involved in detoxification of insecticides, and other organismal processes whose alteration could result in resistance. In addition to altering insecticide penetration though the cuticle, detoxification, or interaction with the target-site, resistance can result from altering either distribution of insecticide within the insect or excretion of an insecticide. For example, the ATP-binding cassette (ABC) transporters are involved in efflux of a wide diversity of substrates, and RNAi knockdown of ABCB subfamily member *Mdr65* in *Drosophila melanogaster* increased toxicity of nine insecticides of various structures (but not five others), while knockdown of *Mdr49* and *Mdr50* increased mortality [21]. While there is a complex relationship between different ABC transporters and toxicity, likely dependent

on their ability to transport particular insecticides or metabolites, expression patterns in the body and in different cell types, some have been clearly implicated in insecticide resistance [22, 23]. In order to advance our understanding of both insecticide resistance and the general processes of a xenobiotic compound's journey through an insect, forward genetic approaches in pest insect species are necessary.

Bulked segregant analysis (BSA) is a technique for trait mapping by pooled sequencing of "bulks" with contrasting traits that was popularized in crop plant breeding [24]. By determining the regions of the genome that differ between bulks, the regions of the genome linked to the trait can be mapped. In contrast to more traditional backcross mapping populations, this is more economically adapted for mapping with whole-genome sequencing data as it reduces the number of libraries required drastically, but combines information from a large number of individuals to increase mapping resolution. Bulked segregant analysis has proven extremely successful in mapping monogenic traits (including pesticide resistance) in the two-spotted spider mite [25-27], as well as in other insects [22]. BSA has also been used against polygenic traits, like melanin pigmentation in *D. melanogaster* (which found 19 QTLs across 9 populations [28]) and variation in locomotion in *Drosophila mojavensis* larvae (which found 2 QTLs [29]). These two organisms have the advantage of an extremely high qaulity reference genome, which is not available for most organisms. It is clear that a highly polygenic trait architecture poses a challenge for any approach to phenotype mapping and that genome assembly fragmentation limits the usefulness of BSA to polygenic traits, but unclear how BSA might fare against a more intermediately polygenic trait in a more standard quality draft assembly.

In order to identify the loci responsible for resistance in a strain of house flies having >80,000-fold resistance to permethrin, this project had six goals. First, we characterized the

mechanisms of resistance using synergists and injection vs topical insecticide bioassays. Second, we conducted a linkage analysis to identify the autosomes involved in resistance. Third, we assembled an improved reference genome assembly for *Musca domestica* to facilitate whole-genome based trait mapping. Fourth, we conducted a whole-genome sequencing based bulk segregant analysis to map the pyrethroid resistance phenotype within the genome. Fifth, we conducted third-generation long read sequencing to identify potential structural variants at the resistance loci. Sixth, we assessed expression differences associated with the resistance. This study provides interesting candidate regions for further validation of causal genes.

2.2 Materials and methods

2.2.1. House fly strains and rearing

The KS17 strain was established from flies collected from a poultry house in Riley, Co., KS in June 2017 [3]. This strain was selected for characterization due to its extremely high levels of pyrethroid resistance. Two susceptible strains were used in experiments: IsoCS [30] and aabys (the reference genome strain) [31]. Table 1 lists the strains and lines used in this paper, along with the experiments they were used for. Resistance phenotypes were validated with permethrin bioassays (see 2.3.1) periodically over the course of all experiments (Table S1). House flies were reared as previously described [3].

Table 1: Strains and lines of house fly used in this paper. Resistant abbreviated as "res" and susceptible abbreviated as "sus."

Strain/Line	Description	Selection of KS17-R	Permethrin & synergist bioassays	Chromosomal linkage of resistance	Genome assembly	Bulked segregant analysis	Structural variant calling	RNA-seq
KS17	Field-collected res strain	X						
KS17-R	Permethrin selected res strain		X	X		X	X	X
IsoCS	Sus strain		X			X	X	X
aabys	Sus strain with morphological markers			X	X			X
SUS1-3	Sus bulk selected from BSA segregating pop					X		
UNS1-3	Unselected bulk from BSA segregating pop					X		
RES1-3	Res bulk selected from the BSA segregating pop					X		
BSAS4	RES1-3 flies were continued as a segregating pop and re-selected for resistance							X

2.2.2. Permethrin selections of KS17-R

In order to create a strain homozygous for the resistance alleles present in the parental KS17 field collection, three selections with permethrin were performed as follows. Adults were separated by sex within 8 hours of emergence (to ensure females were unmated) and held until 3-5 d old in 29.6 mL paper cups (Karat, Chino, CA, USA), and provided powdered milk + granulated sugar (1:1 by volume) and a cotton ball soaked with dH_20, which was moistened twice daily. Flies were dosed topically with permethrin (97%, Syngenta, Wilmington, DE) in 2 µL of ACS reagent grade acetone (VWR, Radnor, PA) using a 100-µL Hamilton syringe (Hamilton Co., Reno, NV). Male and female flies were held separately for 3-4 d post selection and the survivors were released into a cage. Doses and generation details for each selection are presented in Table S2.

2.2.3. Characterization of KS17-R resistance

2.2.3.1 Permethrin and synergist bioassays

To assess the level of permethrin resistance in the selected KS17-R strain, topical permethrin (98.3%, 40.5% *cis*, 57.8% *trans*, Sigma-Alrich, St. Louis, MO) bioassays were conducted. To investigate metabolic mechanisms of resistance, the suppression of permethrin resistance in conjunction with application of synergists was measured. Three synergists were tested: diethyl maleate (DEM) (97%, Sigma-Alrich, St. Louis, MO), an inhibitor of glutathione-*S* transferases (GSTs), *S, S, S*-tributyl phosphorothionate (DEF) (98%, Chem Service Inc., Westchester, PA), an inhibitor of hydrolases, and piperonyl butoxide (PBO, 90%, Sigma-Alrich, St. Louis, MO), an inhibitor of CYPs. The levels of resistance were also checked for two additional pyrethroids: deltamethrin (100%, Roussel UCLAF, Paris, France), and cyfluthrin (98.1%, Chem Service, West Chester, PA, USA).

Insecticide bioassays were performed by topical application to the thoracic notum of 3-5 d old female flies. A 0.5-µl drop of insecticide in acetone solution (VWR, Radnor, PA) was applied using a Hamilton PB-600 repeating dispenser equipped with a 25-µl Hamilton syringe (Hamilton, Reno, NV, USA). The maximum concentration of permethrin that could be solubilized was 200 µg/0.5µL, so for doses above 200 µg the amount of solution applied was increased. Synergists were applied in 0.5-µl of acetone solution 1 hour before permethrin application, at doses of 10 µg for DEM and DEF and 5 µg for PBO. These represent the maximum sublethal doses for each synergist. Controls were treated with synergist + acetone. Each bioassay consisted of 20 flies per dose and at least five doses were used per bioassay with at least three giving mortality values between 0 and 100%, expect in cases where 100% mortality could not be achieved. Treated flies were held at 25°C and 40% relative humidity with a 12:12 (L:D) photo period. Each dose was replicated a minimum of three times using flies from at least two cages, over a minimum of 2 days. Flies received 15% (w/v) sugar water solution from a saturated 2.5 cm piece of cotton dental wick (Richmond Dental, NC, USA). Mortality, defined as flies that were ataxic, was assessed after 48 h. The IsoCS strain [30] was used as the susceptible strain for all bioassays.

Probit analysis [32] using Abbott's correction for control mortality [33] was implemented in R using a custom script (https://github.com/JuanSilva89/Probit-analysis, commit 2eaaff0) to calculate LD_{50} and 95% confidence intervals. In some cases, it was not possible to obtain >50% mortality of the resistant strain due to physical limitations (the solubility of permethrin or the maximum injectable volume of acetone). In these cases, LD_{50} values are reported as ">maximum dose tested" without confidence intervals.

Resistance ratios (RRs) were calculated by dividing the LD_{50} of KS17-R by the LD_{50} of the susceptible strain IsoCS. Synergist ratios were calculated by dividing the LD_{50} of permethrin alone by the LD_{50} for permethrin + synergist.

2.2.3.2 Injection bioassays

Injection bioassays were used to evaluate altered penetration as a mechanism of resistance. A 0.22-µl drop of insecticide in acetone solution (VWR) was injected into the postscutellum of 3-5 day-old female flies using a Hamilton PB-600 repeating dispenser equipped with a 10-µl Hamilton syringe (part no. 80300) (Hamilton), filed to a point. Controls were injected with acetone. Treated flies were held at 25°C and 40% relative humidity with a 12:12 (L:D) photo period. Flies received 15% (w/v) sugar water solution from a saturated 2.5 cm piece of cotton dental wick (Richmond Dental). Mortality, defined as flies that were ataxic, was assessed after 24 h.

2.2.3.3 Chromosomal linkage of permethrin resistance in KS17-R

To determine the relative contributions of each chromosome to pyrethroid resistance in KS17-R, linkage of the heterozygous resistance phenotype to recessive physical markers was assessed, following the method of Tsukamoto [34]. aabys is an insecticide-susceptible strain with recessive morphological markers ali-curve (*ac*), aristapedia (*ar*), brown body (*bwb*), yellow eyes (*ye*), and snipped wings (*snp*) on autosomes 1, 2, 3, 4, and 5, respectively. Thirty-five KS17-R male flies were crossed to 87 aabys females, and the F_1 males were backcrossed to aabys females, resulting in 32 combinations of phenotypes in the backcross generation. The presence of the morphological marker indicated that autosome was from the aabys strain, as the recombination rate in male house flies is very low in most strains [35]. To evaluate the role of each chromosome, mortality at a diagnostic dose of

0.049 µg permethrin (97.8%, 39.3% *cis*, 58.5% *trans*, Sigma-Alrich, St. Louis, MO) was tested using a total of 8095 female flies sorted to the 32 phenotype groups. This dose was chosen because it kills 100% of the aabys strain, but none of the F_1s.

2.2.4. An improved reference genome for the aabys strain of house fly

An improved house fly reference genome for the aabys strain was assembled using long read PacBio sequences, polished using a Chromium linked reads library (10X Genomics, Pleasanton, CA), and scaffolded using Bionano optical mapping. High molecular weight DNA for all libraries was extracted using the Bionano soft tissue protocol on pools of eight flies. In short, raw PacBio subreads reads were corrected and assembled with Canu v1.7.1 [36]. Raw subreads were again used to polish the assembly with Quiver [37]. Then, 10X Chromium data was aligned with LongRanger v2.2.2 and SNPs/INDELs were corrected using Pilon v1.22 [38]. Hybrid scaffolds were generated from the polished contigs and the BioNano maps using BioNano Access v1.2.2. This assembly was manually curated using Purge Haplotigs v1.1.1 to remove allelic contigs [39]. Assembly statistics were assessed using QUAST v4.0 [40] and completeness was assessed using BUSCO v3.1.0 [41] with the Diptera odb9 gene set [42]. Scaffolds were assigned to chromosomes using homology mapping to the *D. melanogaster* Muller elements as in [43].

Gene models were predicted using MAKER v2.31.10 [44] trained with Augustus v3.3.3 [45]. Repeats were soft-masked using a custom repeat library generated with Repeat Modeler v1.0.11 [46]. RNA-seq data from various life stages, tissues, and strains was obtained from NCBI and assembled with Trinity [47] (see supplementary methods). Three rounds of prediction were run, the first used the assembled house fly transcripts and the *D. melanogaster* SwissProt set (obtained from UniProt on October 7, 2019), followed by two

additional rounds to improve the quality of the *ab initio* predictions. Manual curation of gene models was limited to regions of interest from the bulked segregant analysis results.

2.2.5. Bulked segregant analysis of permethrin resistance in KS17

2.2.5.1 Bulked segregant analysis experiment set-up

A segregating population was generated by crossing a single KS17-R female with an IsoCS male (Figure 1). After collection of eggs, both individuals were saved frozen at -70°C for DNA extraction. This population was expanded to approximately 5,000 flies at the F_2 generation, and approximately 9,000 flies from the F_3 on. Population size at each generation was estimated by counting 3 replicate sets of 200 flies and taking the mass, then using the mass of the whole population to estimate number. This line was reared in 3 m^3 cages as a single population with intermixing of the offspring at each generation.

Bulk segregant analysis has been performed by comparing two contrasting phenotype groups (in this case resistant versus susceptible) or by comparing a phenotype selected group with an unselected (resistant versus unselected). We chose to compare resistant versus susceptible flies, as the expected allele frequency differences should be larger. We also collected unselected individuals as an experimental control to monitor for potential fixation of alleles over the experiment. Three phenotypes were necessary: permethrin resistant (survivors of a high dose of insecticide), permethrin susceptible (those dead from a low dose of insecticide), and an unselected pool. At the F_6 generation, preliminary data indicated a selection for KS17-R levels of resistance would results in very few survivors, so the selection for the resistant bulks was done in two steps. For the first selection, a dose was selected to kill ~85% of the population (0.78 µg/female and 0.39 µg/male). Selections were performed on unmated flies as in Section 2.2.2.

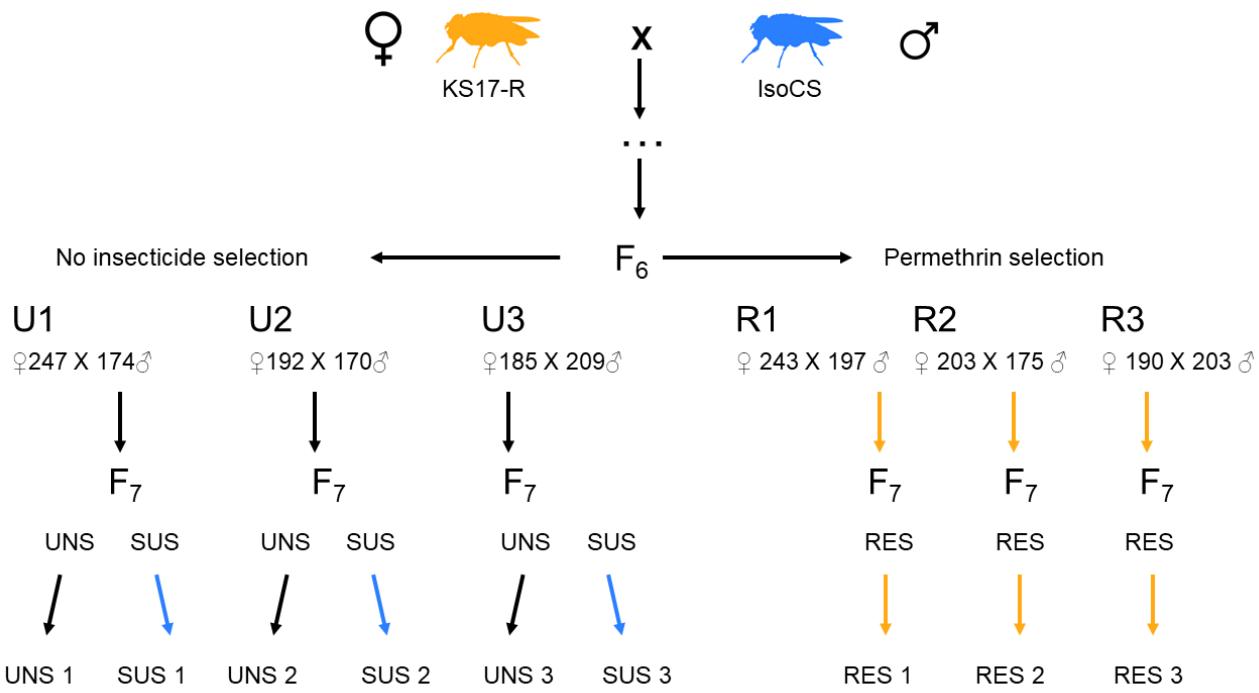


FIGURE 1: A single pair cross of a resistant and susceptible fly was used to create a segregating population for bulk segregant analysis. At the F_6, this population was separated into unselected (U1, U2 and U3) and permethrin selected (R1, R2 and R3) subpopulations of similar sizes. The permethrin selected lines were selected again at the F_7 (RES, orange arrows) to produce the RES1 – 3 lines. The unselected subpopulation was either selected for susceptibility (SUS, blue arrows) or left unselected (UNS, black arrows). BSA was conducted using nine lines UNS1-3, SUS1-3 and RES1-3.

To maintain flies suitable for the susceptible and unselected pools, unselected flies were collected in parallel with the selections, maintaining similar population sizes of approximately 600 females and 600 males between both groups (Figure 1). Flies were selected over one week, and selected flies were released into three cages to reduce stress.

To maintain equal representation of the three cage populations in the final DNA pool they were reared and selected at the next generation as discrete lines. For these final selections, collection of virgin females were not necessary as the survivors were going to be used for extraction of DNA. Doses were chosen to select the 15% most resistant females from the lines selected at the F_6 (100 µg/fly) and the 15% most susceptible females from the lines unselected at the F_6 (0.065 µg/fly). For every 100 flies selected with insecticide, 60 control flies were treated with acetone alone to monitor control mortality. For the resistant selection, mortality was assessed at 72 h and the survivors were stored at -70°C until DNA extraction. For the susceptible selection, mortality was assessed at 24 h and the dead flies were stored at -70°C.

2.2.5.2. Whole-genome sequencing and variant calling of parental strains and BSA pools

Libraries were prepared from each of the two founder individuals and one pool of five KS17-R females to obtain variant information about the parental strains. Three pools each of resistant selected, susceptible selected, and unselected were created with 100 female flies per pool. DNA was extracted from individual female flies, abdomens removed, using an isopropanol precipitation method as previously described [48]. Integrity of DNA was confirmed on a 1% agarose gel. DNA was quantified using the Qubit fluorometer (ThermoFisher Scientific) using the dsDNA Broad Range Assay Kit (ThermoFisher Scientific) according to the manufacturer's protocol. After pooling, samples were treated with Invitrogen RNAse Cocktail (ThermoFisher Scientific) (and a column cleanup was performed with the Zymo Genomic DNA Clean & Concentrator kit (Zymo Research, Irvine, CA, USA), according to the manufacturer's instructions, except that the samples were eluted with TE (pH 8.0). TruSeq DNA library preparation was performed by Cornell University's

Biotechnology Resource Center Genomics Facility. The parental strain libraries were sequenced together in one lane of an Illumina NextSeq500 with single end 75-bp reads by the Cornell Biotechnology Resource Center, and the bulk libraries were sequenced together across four lanes.

Raw reads were trimmed using Trimmommatic v0.39 [49], and then reads were mapped to the newly generated *Musca domestica* genome using bwa-mem v0.7.13 [50]. PCR and optical duplicates were marked using Picard's MarkDuplicates tool v2.19.2 [51]. Joint variant calling was performed with HaplotypeCaller in GATK v4.0.1.1 [52]. Hard filtering was done for SNPs and INDELs separately, the filtered variants were combined, and biallelic variants were selected for the BSA. Variants were annotated for predicted functional consequences using SnpEff v4.3t [53]. Scripts for variant calling are provided at https://github.com/JamieCFreeman/BioHPC_SNP_calling (commit b8ec540).

2.2.5.3. Analysis of BSA pool allele frequency data

Prior to analysis, fixed variants differing between the IsoCS and KS17-R parents were selected. A minimum depth of four reads from each parent were required to include a variant. The allele counts for the RES versus SUS pools (Figure 1) were used for analysis as they show greater differentiation relative to the UNS pools. The R package 'QTLseqR' v0.7.5.2 [54] was used to calculate the G` statistic as described in [55] and the tricube smoothed $\Delta_{SNP}$ statistic (the difference in ALT allele frequency between the two bulks). While G` and Δ_{SNP} both measure differentiation between bulks, and the G` value was used to determine loci statistically associated with the resistance, Δ_{SNP} is on a more familiar scale and is used to discuss absolute differences between bulks. To ensure precision of allele frequency measurements, variants were filtered for a minimum per sample depth of 20 and a

maximum depth of 200 (to exclude repetitive regions). QTLseqR expects that one parent strain corresponds to the REF allele of the VCF files, while the other corresponds to the ALT allele, so for variants where the IsoCS strain had the ALT allele while KS17-R had the REF, the alleles were switched so the REF allele corresponded to the IsoCS allele. As G` is calculated using the logarithm of the count values, a pseudocount of 0.1 was added to all counts for the G` analysis. A window size of 1,000,000 bp was chosen, after testing widow sizes from 500,000 to 10,000,000 bp. Preliminary analysis indicated the results were robust to changes in window size. Scaffolds containing less than 40 SNPs were excluded from the analysis. As G` is close to being log normally distributed, the statistical significance for particular variants can be assessed by estimating the null distribution of G` from the data. Magwene [55] proposes filtering of outliers using Hampel's rule, but Mansfield [54] recommends filtering using a tricube smoothed Δ_{SNP} (the difference between ALT variant allele frequencies of the two bulks). In our analysis, filtering of outlier regions using the tricube smoothed Δ_{SNP} (set to 0.25) resulted in a better fit of the estimated null distribution to the data (see Figure S1).

2.2.5. RNA-seq to identify candidate genes involved in resistance in KS17-R

Some genes (e.g. *CYPs*, esterases, nuclear receptors, transcription factors) may be involved in the resistance phenotype through a change in expression, and this overexpression may map to the genes themselves (regulation in *cis*) or to another region of the genome (regulation in *trans*). This means that while the BSA can localize genes that confer resistance through a *cis* regulatory change, it cannot identify genes that confer resistance as a result of a *trans* regulatory change (though potentially the *trans* regulatory change could be found). In order to better understand how gene expression changes were

related to permethrin resistance of the KS17-R strain (especially in the detoxification gene families implicated from the synergist bioassays (section 2.2.3.1)) an RNA-seq analysis was used to identify genes overexpressed at the transcript level in individuals from the segregating population that exhibited a resistant phenotype.

Approximately 200 mated females combined from the three BSA resistant bulks (Figure 1) were allowed to oviposit (they had mated pre-selection, so the offspring were a mix of resistant and susceptible genotypes). This mixed line was continued at a population size of ~1,000 flies, and was selected with permethrin for two generations (see Table S5). These additional generations of recombination were expected to further break down linkage of the parental haplotypes, resulting in smaller regions of the genome linked to the resistance allele. This line (BSAS4) was used to examine differentially expressed genes, in comparison to the two parental strains IsoCS and KS17-R, along with an additional susceptible strain aabys. The resistance phenotypes of the four strains were confirmed with bioassays at the time of harvesting individuals for RNA (Table S1).

The six biological replicates were reared at different times, with the four strains reared in parallel and adult females harvested at 5 d old. Three whole 5 d old females were homogenized in 1 mL of Trizol by bead beating at for 30 s at 4.5 m/s with ~10 zircona-silica beads (Biospec Products, Bartlesville, OK, USA) using a Bead Ruptor 12 bead mill homogenizer (Omni International, Kennesaw, GA, USA) and frozen at -80°C until RNA extraction (1-2 months). Cage placement in the rearing chamber and RNA processing order were randomized in blocks. RNA was extracted using a hybrid Trizol-RNeasy (Qiagen, Germantown, MD, USA) protocol. RNA was quantified using the Qubit broad range RNA kit, and RNA quality was checked using a Fragment Analyzer (Advanced Analytical).

Lexogen 3` QuantSeq FWD libraries were prepared with 500 ng per sample according to the manufacturer's protocol. This protocol produces strand-specific libraries, with one fragment amplified from the 3` end of the transcript (also called a 3` tag-seq method). Library size distribution was evaluated using a Fragment Analyzer. One library failed quality checks and was not included in the final pool, so the BSAS4 line had five replicates only. An equimolar pool of all samples was sequenced on 75-bp single end Illumina NextSeq 500 at the Cornell Biotechnology Resource Facility.

Read processing of the 3` tag-seq libraries was implemented in Snakemake v5.20.1 [56] (available at https://github.com/JamieCFreeman/3prime-tag-seq): quality check of raw reads was performed using FastQC v0.11.9 [57], reads were trimmed with bbduk v38.86 [58], and mapped using STAR v2.7.6 [59]. Multiqc v1.9 [60] was used to collate quality control statistics. While manual curation of the house fly assembly presented here did reduce the number of allelic contigs, it was unfeasible to remove all of them. The windowed analysis of G` for the BSA requires long scaffolds in order to make sense of the signal, but for the RNA-seq analysis it was more important to have a more curated gene set. Multi-mapped reads are not used for quantification in STAR, so any genes that appeared in the assembly twice would be unlikely to be quantified. For this reason the NCBI *Musca domestica* assembly v2.0.2 and annotation file were used for the RNA-seq analysis [14]. The count table is available as Supplementary File 5. Gene annotations were mapped to the new assembly for comparison to the resistance loci from the BSA using Liftoff v1.5.1 [61].

Accurate counting of transcripts from a 3` tag-seq library requires a gene annotation covering the 3` UTR, which is often poorly recovered using automated annotation methods. While the mapping rate of the RNA-seq data to the genome itself was >95% in all libraries,

the rate of reads mapped to exons was initially lower than expected, indicating an imperfect annotation. Adding a buffer region to the end of annotated transcripts can improve the assignment of mapped reads to transcripts. Initial analyses used the *Musca domestica* NCBI annotation version 102, as well as this annotation with all transcripts extended 200 bp or 400 bp on the 3` end of the gene. Extending transcript annotations by 200 bp increased the percent of reads mapped to an exon from 69.1% to 73.5% (see Supplementary Table RNAseq annotation). In comparison, extending transcripts from 200 bp downstream to 400 bp only increased the exonic mapping rate by 1%, so extensions beyond 400 bp were not assessed. The NCBI annotation version 102 with all transcripts extended 400 bp on the 3` end was used for further analysis. Differential expression analysis was conducted in R v3.6.3 [62] using DESeq2 v1.26 [63] to test for differences in expression between strains. Log-fold change shrinkage was performed with the 'ahsr' model [64]. Principle component analysis was performed on counts with the regularized log transformation in DESeq2 [63], and the transformed counts are also used for heatmap visualization in the R package pheatmap v1.0.12 [65]. *CYP* names and associated NCBI gene IDs are provided in supplementary file 4.

2.2.6. Nanopore sequencing and identification of structural variants between KS17-R and IsoCS

2.2.6.1 High molecular weight DNA extraction and MinION sequencing

One pool of KS17-R individuals and one pool of IsoCS individuals were sequenced on the Nanopore MinION to obtain long reads for structural variant (SV) calling between the strains. To minimize diversity within these pools for both strains, an inbred line from a sibling cross was used. High molecular weight DNA was extracted from pools of six

individuals using a protocol adapted for insect tissue from the Agilent RecoverEase DNA Isolation Kit (Agilent Technologies, Inc., Santa Clara, CA, USA) which includes nuclei enrichment, digestion with RNAse and proteinase K, and sample cleanup through dialysis. Major changes included increasing the time for proteinase K digestion, increasing the concentration of proteinase K, and adding a phenol:chloroform separation between digestion and dialysis (see Supplementary methods for details). Flies removed from the pupal case (pharate adults) before emergence were found to provide the best yield and purity of DNA. Fragment size was assessed using the FEMTO pulse before and after library preparation. The Circuolomics Short Read Eliminator Kit was used to deplete fragments <25 kb (Circulomics Baltimore, MD, US). Library prep was performed with the Ligation Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK), and libraries were sequenced on a MinION flow cell R9.4.1 (Oxford Nanopore Technologies) for 48 h. The IsoCS library had poorer than expected pore occupancy, so an additional AMPure bead cleanup (Beckman Coulter, Indianapolis, IN, USA) was added before library preparation for the KS17-R library.

2.6. 2. Nanopore read processing and SV calling

Guppy v4.0.11+f1071ce (Oxford Nanopore) was used for basecalling. Nanoplot v1.32.1 [66] was used to visualize run statistics. The pipeline used for structural variant calling was based on the results of [67] and was implemented in Snakemake [56]; scripts are provided at https://github.com/JamieCFreeman/nglmr_SV, commit c0c2f68. Briefly, reads were mapped to the newly developed reference genome with NGLMR v0.2.7 [68], variants were called with Sniffles v1.012-1 [68], SURVIVOR v1.0.7 was used to merge individual

sample vcfs, and the combined vcf was provided back to Sniffles to force calling all sites in all samples.

2.3 Results

2.3.1. Selection of KS17-R

The KS17 field collected population was selected with permethrin over three generations (Table S2). Resistance did not change between the second and third selections, so no further selections were performed, and the resulting strain was named KS17-R. The resistance phenotype of KS17-R was confirmed periodically with permethrin bioassays (Table S1).

2.3.2. Characterization of KS17-R resistance

2.3.2.1 Permethrin, synergist, and injection bioassays

Results for bioassays are presented in Table 2. At the maximum dose possible due to the solubility of permethrin in acetone (800 µg permethrin in 2 µL acetone solution) only 22.5% mortality was obtained for KS17-R (n=120). It is not possible to calculate the LD_{50} for KS17-R, so it is represented for the purposes of further calculations as >800 µg. This represents a resistance ratio of >84,900 relative to IsoCS. PBO reduced the resistance ratio from >84,900 to 8,610, indicating that CYP-mediated resistance is likely a mechanism of resistance. Application of DEM did not result in any measurable reduction of the KS17-R permethrin LD_{50}. DEF resulted in a reduction of the permethrin LD_{50} by at least 29-fold for KS17-R and 3.2-fold for IsoCS, indicating a potential involvement of hydrolases in resistance. The KS17-R strain was also highly resistant to deltamethrin (>40,000,000-fold) and cyfluthrin >240,000-fold) (Table 2).

Table 2: Effect of synergists and injection on permethrin resistance in KS17-R. LD_{50} given in µg per fly. Resistance of KS17-R to deltamethrin and cyfluthrin.

	IsoCS		KS17-R		
	LD_{50} (95% CI)*	SR	LD_{50} (95% CI)*	SR	RR
permethrin	9.4 (9, 9.9)	-	>800	-	>85,100
+DEF	2.94 (2.9, 3.0)	3.2	27.7 (24.1, 31.8)	>29	9,420
+DEM	7.26 (6.08, 8.67)	1.2	>800	?	>110,00
+PBO	0.732×10^{-4} (6.46, 8.3)	13	6.25 (5.4, 7.2)	>128	8,610
injected	-	-	>88	?	-
deltamethrin	0.5 (0.5, 0.6)	-	>200	-	>400,000
cyfluthrin	8.47 (7.84, 9.15)	-	>200	-	>23,600

* LD_{50}s are given in units of ng/fly for IsoCS and in µg per fly for KS17-R.

The maximum dose of permethrin that can be solubilized in 0.22-µl of acetone is 88 µg. Injection of this dose only resulted in 18% (n=160) mortality of KS17-R, compared to 9.2% mortality (n=140) in the acetone only controls. This result indicated that permethrin resistance through altered penetration in KS17-R is not present at a level measurable with this assay, so experiments were not continued with IsoCS.

2.3.3 Linkage of permethrin resistance in KS17-R

The linkage analysis revealed that 4 autosomes are involved in permethrin resistance, with the relative ranking being chromosome 3 > 5 > 2 >1 (Figure 2, Table S3). In addition, there was a greater than additive effect seen for autosomes 2+3+5. Resistance in KS17-R is polygenic, a conclusion also supported by the synergist bioassay results. All markers were

found in male and female flies in the backcross, indicating that the male-determining factor in KS17-R is on Y.

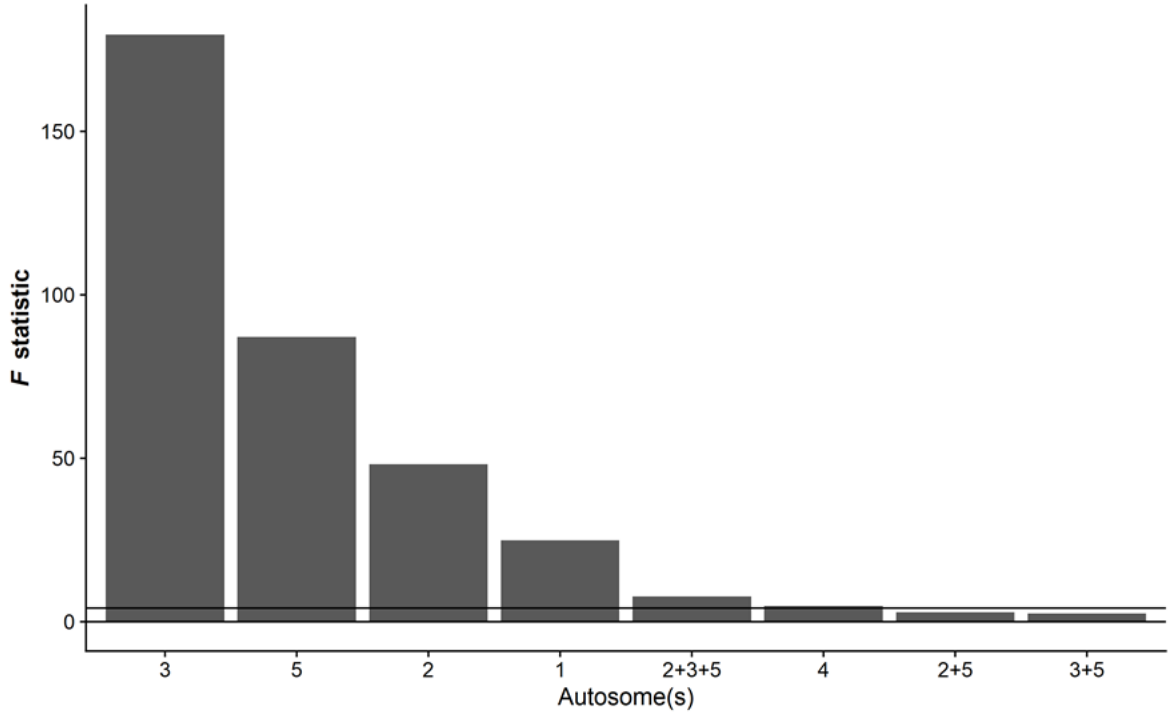


FIGURE 2: Factorial analysis of the linkage of permethrin resistance in the KS17-R house fly strain indicates resistance is linked to chromosomes 3, 5, 2, and 1. Horizontal line marks critical value for *F* distribution (df_1=1, df_2=30, α=0.01). All 31 possible combinations of chromosomes were tested (S Table Linkage Effect), but only those with top 8 largest *F*-statistic are shown.

2.3.4 Improved house fly reference genome

Approximately 5.39 Gb of PacBio data was generated using the aabys strain, with a median read length of 28,319 bp and a maximum read length of 186,540 bp. This was assembled into 2,845 contigs by Canu, with a total length of 1.22 Mb. After scaffolding, 1,684 contigs were present in the assembly with a total length of 1.37 Mb, an N_{50} of 9.1 MB,

and an L_{50} of 38 (Table 3). While the number of BUSCO genes present in the assembly was greater than those in the 2014 assembly, the rates of duplicated BUSCOs was much higher. In total 57.5% of the Diptera single-copy orthologs included in the BUSCO analysis were duplicated in the raw assembly (many of them more than once). These duplicates indicated the presence of uncollapsed alternative haplotypes in the assembly, an indication of high heterozygosity data. This posed significant challenges for variant calling against this genome, so Purge Haplotigs [39] in combination with manual curation was used to refine the assembly. Purge Haplotigs uses scaffold identity and read depth (if a scaffold has two haplotypes present in the assembly, read depth for each should decrease) to identify and remove alternative haplotypes from the assembly. By the 5^{th} round of curation, duplicated BUSCOs had dropped to 11.3%, which indicates that alternate haplotigs are still present in the assembly, but remaining haplotigs had lower percent representation in the genome and by removing them, there was a risk of removing real genes completely. Both the raw scaffolded assembly and the purged assembly are provided. The contiguity of the new genome assembly is far greater than that of the 2014 assembly (Table 3). The "purged" version is used for the BSA and the SV calling.

Table 3: Genome assembly statistics for the newly generated aabys genome assembly versus the 2014 aabys genome assembly [14]. BUSCO scores were assessed for the Diptera lineage orthologs (obd9).

	2014 genome	2020 genome v1	2020 genome purged
# contigs	20487	1684	341
Largest contig (Mb)	2.35	52.31	52.31
Total length (Mb)	750	1370	1033
GC (%)	35.11	35.06	35.05
N_{50} (Mb)	0.23	9.06	12.46
N_{75} (Mb)	0.083	1.90	5.41
L_{50}	809	38	23
L_{75}	2176	122	51
# N's per 100 kbp	7820	10399	12129
BUSCO complete (%)	0.975	0.987	0.977
BUSCO duplicated (%)	0.016	0.557	0.113

2.3.5 Bulked segregant analysis

2.3.5.1 Resistant, susceptible, and unselected flies were collected from the segregating population

The segregating population initiated by a single-pair cross of a resistant KS17-R and susceptible IsoCS fly was interbred until the F_6 generation, when a two-generation process for selecting resistant, susceptible, and unselected flies was started. For the first selection for resistance, a total of 4,046 females were selected with a mortality rate of 84.3% and

3,792 males were selected with a mortality rate of 84.8%. Survivors were spread across three cages and mated to produce the next generation (see Figure 1). For the second selection for resistance, 11,688 females were selected with 100 µg of permethrin, resulting in 85.2 % mortality over the three subpopulations. For the selection for susceptibility, 9,243 females were selected with 0.065 µg of permethrin, resulting in 85.2% survival. Control flies treated with acetone alone had low mortality (0.92% for 1,844 flies treated) indicating that the dead flies in the selection for susceptibility were killed by the insecticides and not in the assay process.

2.3.5.2 Sequencing and selection of informative variants for BSA

For the parent strains, 23.1 Gb of sequence data was generated, with 6.4X coverage of the house fly genome for the IsoCS male parent, 6.9X coverage of the KS17-R female parent, and 9.1X coverage of the KS17-R pooled sample. Three replicate pools each of resistant, susceptible, and unselected flies were sequenced over four lanes, generating 128.9 Gb of sequence data. This corresponds to 10-14X coverage of the house fly genome for each of the three resistant and three susceptible libraries. The three unselected libraries were sequenced at lower coverage (5X). Though replicate pools were sequenced for each bulk, the depth of coverage was not high enough to analyze these replicates individually. The number of reads supporting each allele were summed over replicate pools, resulting in a total of 33.6X coverage for the susceptible bulk, 39.3X coverage for the resistant bulk and 16X coverage for the unselected bulk.

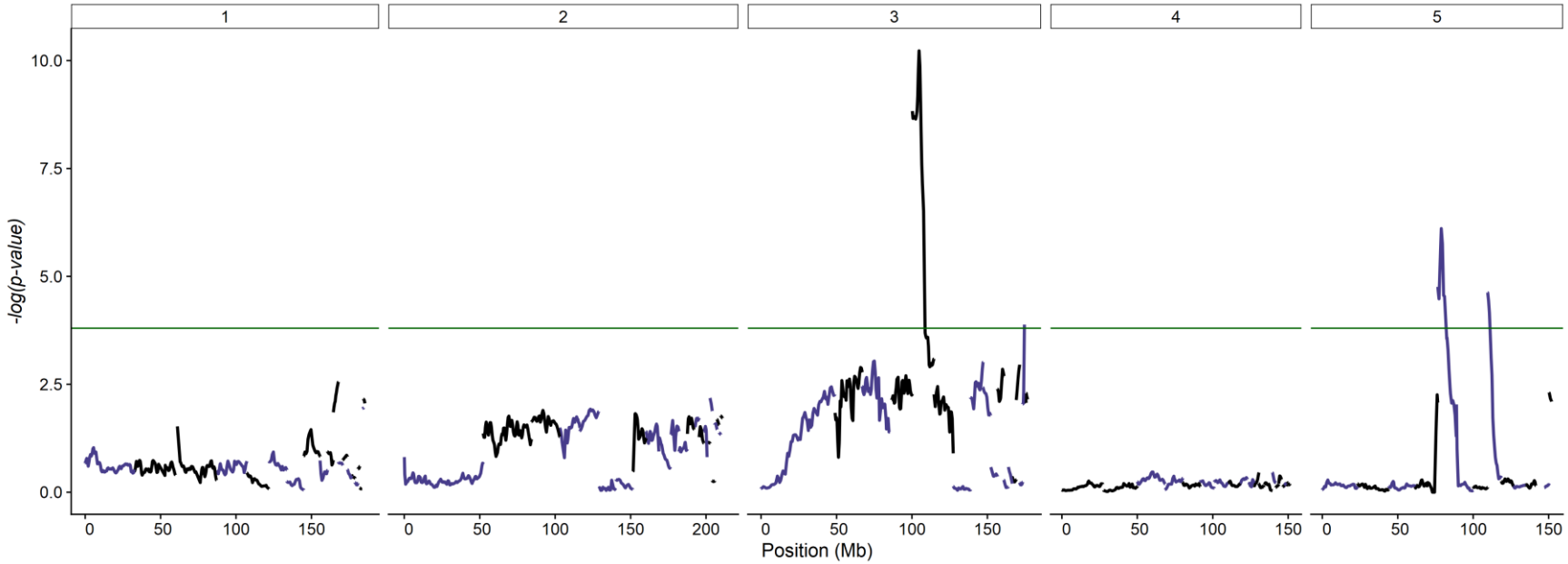
Overall 1,879,209 variants were informative for the BSA after all filters were applied (count table in Supplementary File 1). The G` statistic uses a tricube smooth function to incorporate data over a large genomic window, with weights corresponding to distance from

the focal variant. With a window size of 1 Mb, the median number of variants included in each window was 3,142, though this was highly variable (Figure S2). Variants used for the analysis had a median coverage of 52 for the resistant bulk and 44 for the susceptible bulk (Figure S3).

2.3.5.3 BSA localizes resistance loci on chromosomes 3 and 5 and more dispersed signal of resistance on chromosomes 2 and 1

Over the five house fly autosomes large variance is observed in the G` statistic (Figure 3). Even with the improved reference genome, the signal of the analysis is spread over multiple scaffolds, resulting in discontinuities between the unordered scaffolds. Clear peaks in signal are visible for chromosomes 3 and 5. As predicted by the chromosomal linkage analysis, chromosome 4 has little differentiation between the resistant and susceptible bulks. Chromosomes 1 and 2 were associated with resistance in the linkage analysis, and they do show greater differentiation than chromosome four, but they do not have a clear maximum point of differentiation.

FIGURE 3: Bulked segregant analysis of permethrin resistance in KS17-R indicates clear resistance loci on chromosomes 3 and 5. Negative $\log_{10}$ p-value from the G` analysis over scaffold predicted to each of the five house fly chromosomes. Green horizontal line represents FDR=0.01. Within chromosomes, scaffolds are ordered by length, and colors alternate to indicate breaks between scaffolds.



The most significant peak in the data is located on chromosome 3 (scaffold 20), and the peak is centered on *Vssc* (Figure 4). The point of the maximum G` value, 4,459,036 bp, falls inside the *Vssc* gene, 82 kb from the *skdr* mutation and 87 kb from the *kdr* mutation (which occur on the same haplotype). The recovery of this known causal region in the analysis indicates that the signal recovered by the BSA does represent insecticide resistance phenotype. The *kdr* sand *skdr* mutations were present in 100% of the reads for the resistant bulk, but both mutations also appeared at low frequency in the susceptible bulk (22.2% for *kdr* and 9.78% for *skdr*). This is not unexpected, as the resistance phenotype conferred by these mutations is incompletely recessive [69]. This also potentially represents some noise in the process of selecting for susceptibility. A total of 15 nonsynonymous variants were called within the *Vssc*, and if filters are applied to select variants that are homozygous for the reference allele in the IsoCS male parent and homozygous alternate for the KS17-R female parent, only three variants remain (two of which are *kdr* and *skdr*). The parental KS17 unselected strain had three known *Vssc* pyrethroid resistance alleles, *kdr* (1014F), *skdr* (918T+1014F), and *1B* (929I+1014F) [3]. As *skdr* and *1B* confer similar levels of permethrin resistance, much greater than that conferred by *kdr* [69], it was expected that both *skdr* and *1B* alleles would remain in the population after permethrin selection. While the KS17-R parent was homozygous for the *skdr* allele, the *1B* is present in the pool of KS17-R individuals at a frequency of 20% (for 15 reads total). The differentiation in allele frequencies between the resistant and susceptible bulks persists over much of chromosome 3, with a median Δ_{SNP} of 0.20. It is not possible to rule out the possibility of a second resistance locus on chromosome 3 from the data, but there are no clear signals indicating a second peak (Figure 3). While one additional scaffold hits the FDR line, it is important to

recall that the scaffolds are unordered, and we conclude this scaffold is likely just located next to scaffold 20.

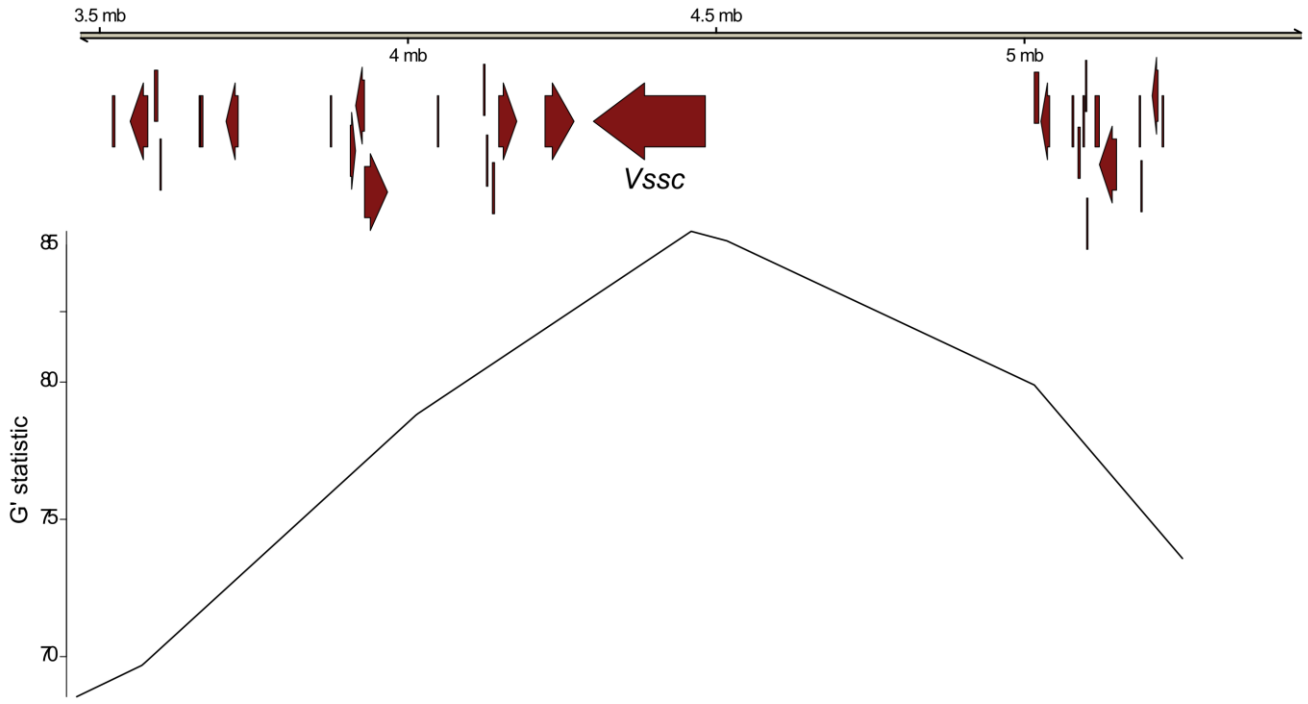


FIGURE 4: The maximum G` value for chromosome 3 is within the target-site of the pyrethroids, the *Vssc*. Gene models are represented as red arrows, with scale bar indicating position within scaffold 20 in Mb.

The second largest peak in the data is location on scaffold 25, which is predicted to be on chromosome 5. The peak G` value is at 2,343,964 (Figure 5), which is directly upstream of a cluster of *CYPs* (circled in Figure 5), and downstream of a gene predicted to be *Coagulation factor 5*. In contrast to the near fixation observed in the resistant bulk around the *Vssc* for the KS17-R alleles, the maximum smooth Δ_{SNP} value for scaffold 25 is 0.49. In looking for candidate variants, it would not make sense to rule out heterozygous

variants in this case. This leaves variants that are heterozygous or homozygous ALT in KS17-R, but homozygous REF in IsoCS, of which there are 157 in the 25 *CYPs* and none in *Coagulation factor 5*. Despite the lack of a clear candidate gene from the many *CYPs* in the cluster, the mapping data presented a region from ~ 2.37 Mb to 2.42 Mb where the depth of mapped reads dropped in the IsoCS parent, and the SUS bulks, but not in KS17-R nor the RES bulks (Figure S4). This indicated a potential structural difference between KS17-R and IsoCS directly at the peak G` value. This led us to investigate structural variants between the strains using long-read sequencing and structural variant calling (see Section 2.3.6).

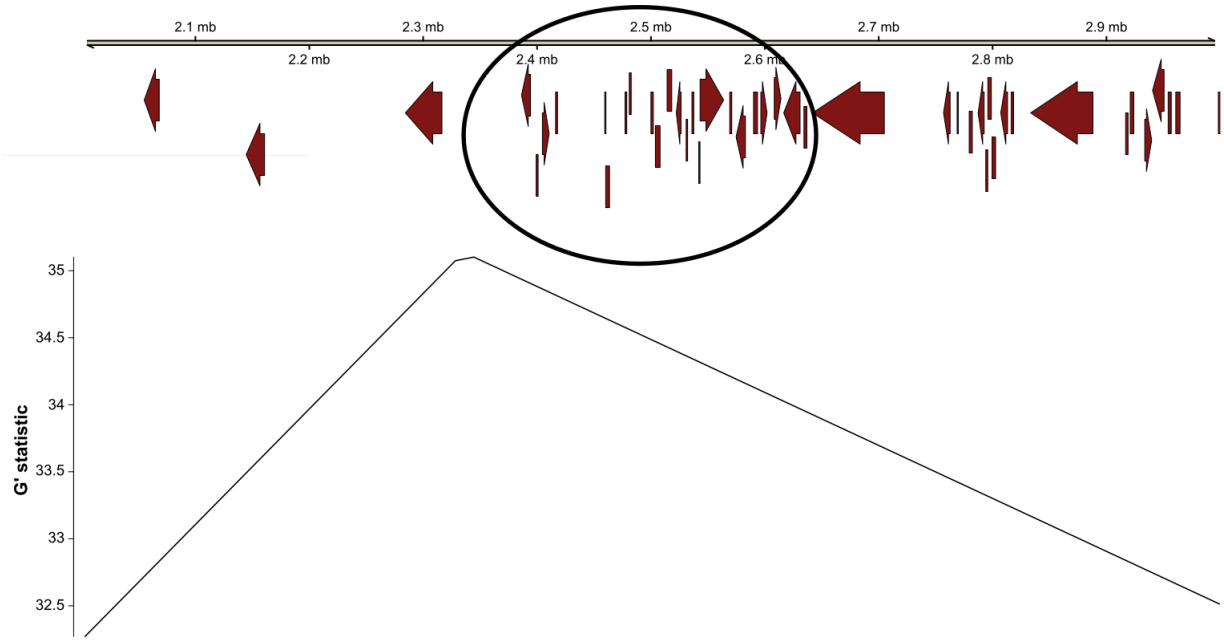


FIGURE 5: The maximum G` value for chromosome 5 is directly upstream of a cluster of *CYPs* (circled). The gene directly upstream of the cluster is *Coagulation factor V*. Gene models are represented as red arrows, with scale bar indicating position within the scaffold in Mb.

In contrast to chromosomes 3 and 5, the signal over chromosomes 1 and 2 was less clear. Chromosome 2 has relatively stable G` over most of its length. Despite the lack of a clear peak in the G` signal as seen on chromosomes 3 and 5, the resistant bulks do show greater frequency of KS17-R alleles than the susceptible bulks do, and the median Δ_{SNP} value for chromosome 2 is similar to that of chromosome 3 (Figure 6). In combination with the linkage data, this suggests there are multiple resistance loci on chromosome 2, but no particular regions stand out as potential candidates.

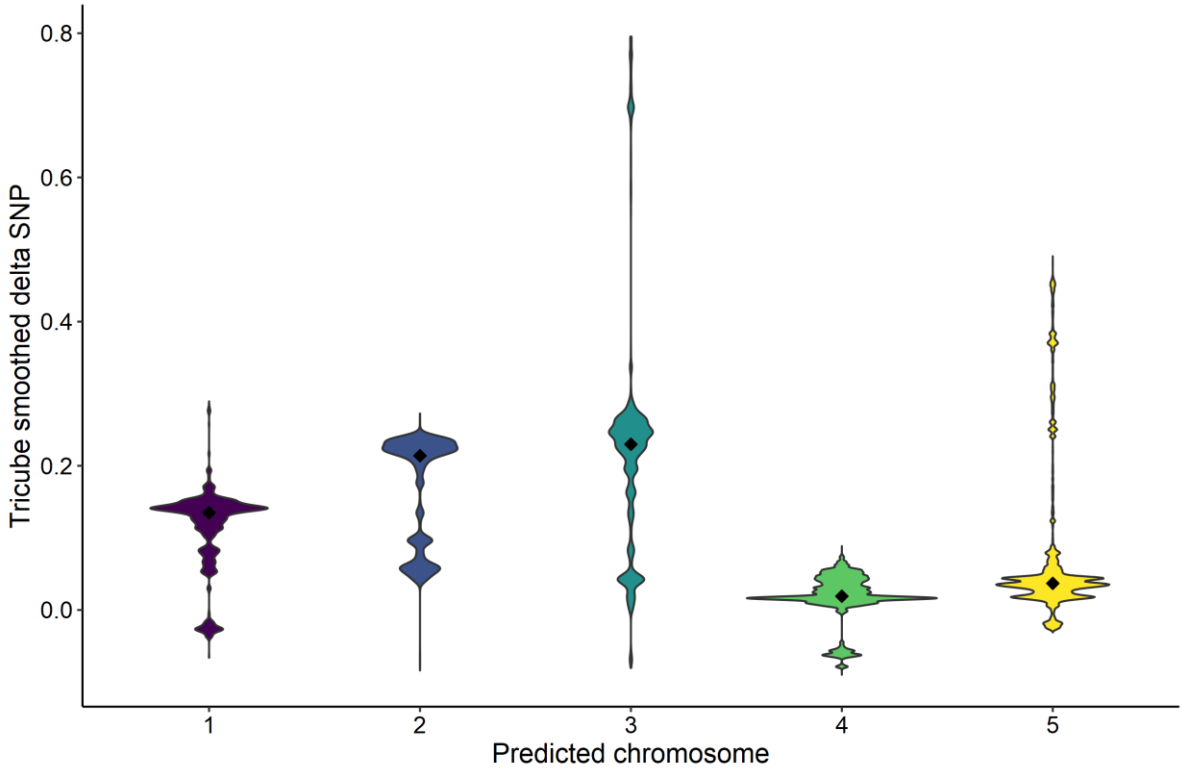


FIGURE 6: Though chromosomes 3 and 5 show the clearest signatures of association with the resistance phenotype, the median frequency of Δ_{SNP} on chromosome 2 closest to that of chromosome 3. Chromosome 1 shows greater differentiation than chromosome 4.

The overall Δ_{SNP} values on chromosome 1 are lower than those of chromosome 2, but there are clearer scaffolds that rise above the baseline level for G` for the chromosome (on scaffolds 34 and 78, Figure S5). The maximum G` on 34 lies at 4.85 Mb within a nitric oxide synthase gene, but as the scaffold has a gap right after this gene, it's unclear whether that is the true maximum of it is should lie in the gap. The only known *CYP* allele conferring resistance is the *CYP6D1v1* allele, which is on 34 ~1 Mb away from the maximum position. The G` value does not appear to peak on scaffold 78, but potentially the adjacent one, though it is difficult from the chromosome 1 data to decide which scaffold that should be.

2.3.6 Nanopore sequencing indicates complex structural variation at the chromosome 5 resistance locus

In total, 7.55 Gbp of Nanopore sequence data was obtained for IsoCS and 6.91 Gbp was obtained for KS17-R, providing ~10X coverage for both strains. Overall, longer reads were obtained for the IsoCS library which had a mean read length of 10,644 bp compared to 6,653 bp for KS17-R (see Figure S6), but more reads were obtained for KS17-R (~1 million versus ~700,000). The longest read overall was 226,325 bp (for IsoCS). This data was mapped to the newly generated reference genome, with ~10X coverage for both strains. A total of 115,895 variants were called over both strains, with 52,957 variants called for IsoCS, and 92,917 variants for KS17-R, with 19.8% overlap (Supplementary files 2 & 3). Insertions (40.1%) and deletions (54%) were the most common structural variants, with an average size of 500 bp for insertions and 2,810 bp for deletions. In comparison, duplications (2.9%), inversions (1.0%), and inverted duplications (0.49%) were much rarer. Sniffles

called 37 enormous variants >1 Mb (27 in KS17-R and 17 in IsoCS); examination of the breakpoints of a sample of these variants indicated a presence of a transposable element at one end. These variants were generally at low (<0.50) frequency in either population. Transposable element (TE) events likely make up a substantial proportion of SVs called in this set. Differences in TE presence between strains was of interest, but it was not possible to separate TE-associated and non-TE associated SVs without extensive curation of the repetitive elements present in the genome.

Examination of the chromosome 5 resistance locus indicates the presence of multiple structural variants. Present in both strains is a segregating 55 kb deletion (at a frequency of 0.5 in IsoCS, and 0.25 in KS17-R) covering the region including *CYP6A56, CYP6GU1*, and one copy each of *CYP6A24* and *CYP6A25* (which are also both duplicated outside of the region). A scaffold containing this duplication was also present in the 2020 aabys assembly before haplotig purging (scaffold 133). This explains the decreased read coverage in the IsoCS male parent and the BSA SUS bulks in this region observed in section 2.3.5.3. Based on the short read data, it appears that the IsoCS male parent was homozygous for the deletion allele, so the only variants in this region are those present in K17-R. Aside from this large deletion, IsoCS only has indels that affect intergenic regions or introns. In contrast, KS17-R has a complex series of structural variants, with a partial duplication of the CYP cluster and a series of additional segregating inversions and deletions within the duplicated region. Three reads span the duplication breakpoint, allowing for confident identification of its position in the genome (~2.62 Mb, Figure 7A) and the start of the duplicated sequence (~2.23 Mb, Figure 7A), though how much of the cluster is present in the duplicated allele is unclear. In total, 87 structural variants were called within the

duplicated region in KS17-R, relative to 11 within the same region for IsoCS (which does not show any evidence of duplication). The complexity of the SVs in this region KS17-R (Figure 7A) in comparison to IsoCS (Figure 7B) is clearly visible. The large duplication event in KS17-R is present in the reads at a frequency of 0.5, and most of the SV events within the duplicated region have frequency < 0.50 indicating that they likely are not fixed in either copy of the gene cluster. The gene directly upstream of the CYP cluster on the scaffold is also included in the duplication. Many of the SV events in KS17-R result in the loss of a functional protein (either through deletion of an entire gene or a nonsense mutation). Because the sample sequences was a pool of individuals it is not possible (with the coverage we had) to predict phased haplotypes for the region.

We also investigated the presence of structural variants affecting other *CYP*s in the genome. While there are 39 SV events that affect *CYP* exons, none are homozygous in either strain. The maximum variant frequency was 7/8 reads for a deletion on scaffold 49 that deletes the first exon of *CYP4C74* in KS17-R (though there appears to be an alternate start codon in the second exon). On chromosome 1, *CYP6D3*, *CYP6D1*, and *CYP310B2* occur next to each other. Two deletions (one 57 kb and the other 16 kb) in between *CYP6D3* and *CYP6D1* are shared between KS17-R and IsoCS (Figure 8). A duplication of 11 kb including *CYP6D1* is observed in KS17-R, supported by four reads spanning the breakpoint. Additionally, a 15-bp deletion previously linked to a *cis* regulatory change in CYP6D1 is present in the KS17-R reads (though too small to be called by Sniffles).

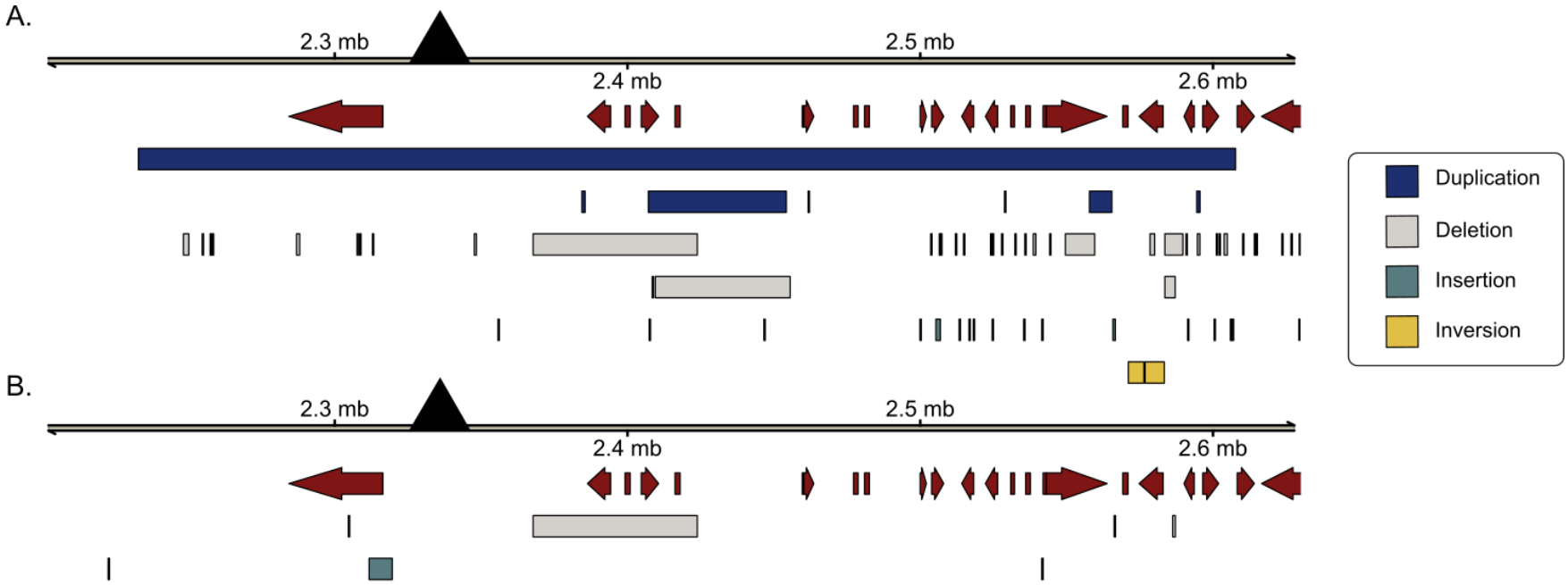


FIGURE 7: Structural variant calling with Nanopore long reads indicates complex SVs in KS17-R (A) at the chromosome 5 resistance locus relative to IsoCS (B). Gene models are represented as red arrows, with scale bar indicating position within the scaffold in Mb. Black triangle on the scale bar indicates the position of the peak G` value.

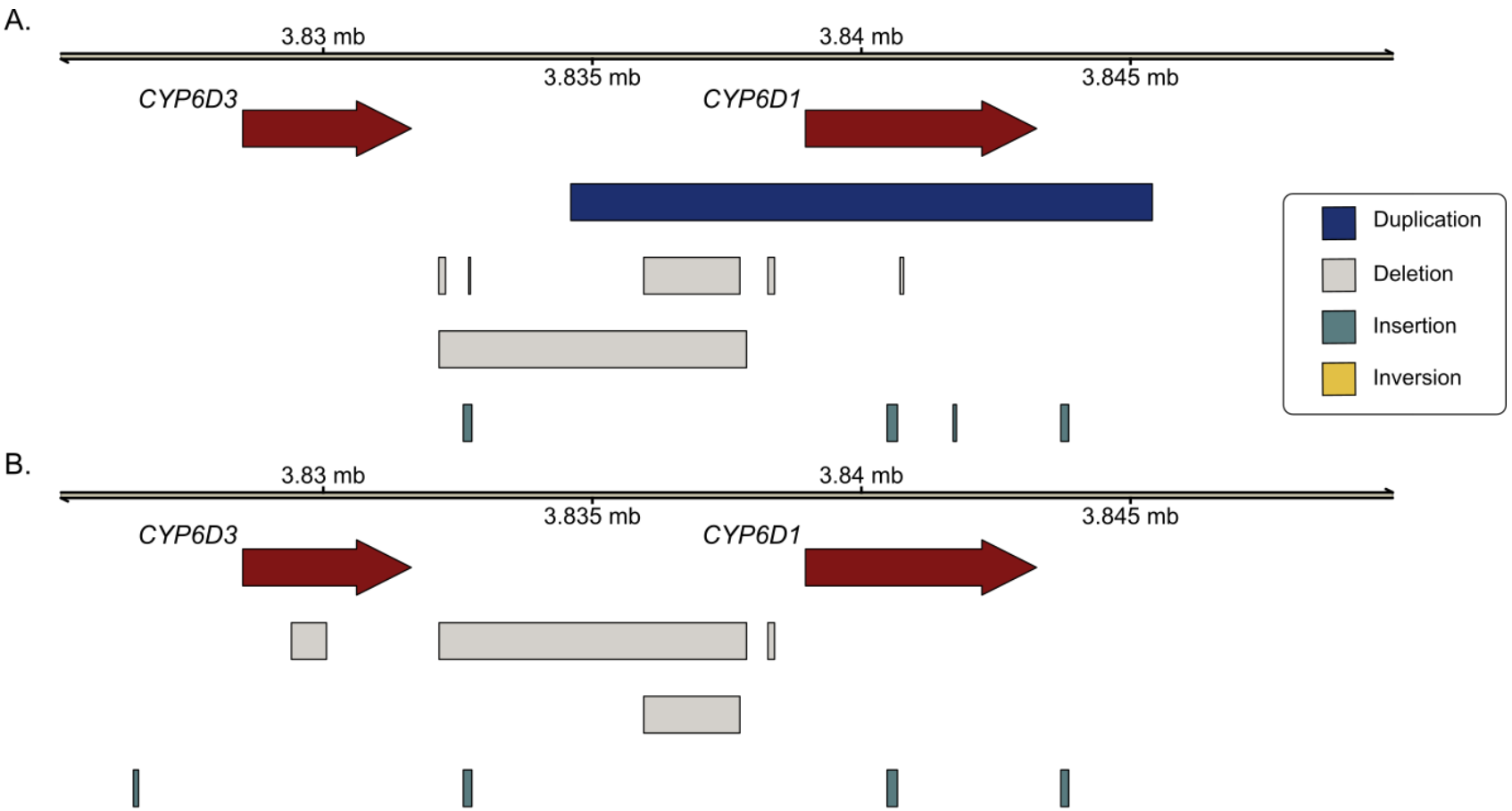


FIGURE 8: Structural variant calling with Nanopore long reads indicates SV affecting *CYP6D1* in KS17-R (A) and shared SVs in the intergenic region with IsoCS (B). Gene models are represented as red arrows, with scale bar indicating position within the scaffold in Mb.

2.3.5 RNA-seq identifies candidate genes involved in resistance in KS17-R

For the RNA-seq experiment, a total of 469 million reads were sequenced over 23 libraries. While BSAS4 is related to both IsoCS and KS17-R, it clustered more closely with IsoCS and aabys than KS17-R along the first principle component (PC1) (Figure 9). Bioassay data indicates that BSAS4 is approximately 2,600-fold more resistant than IsoCS (Table S2), so while the resistance phenotype of BSAS4 is closer to that of KS17-R, the overall gene expression patterns are more like IsoCS. In contrast, a PCA performed on 145 *CYPs* separates BSAS4 from IsoCS on PC1, which represented 38% of the variance (Figure 10). The aabys strain was separated from the three other strains along PC2 (representing 17% of the variance).

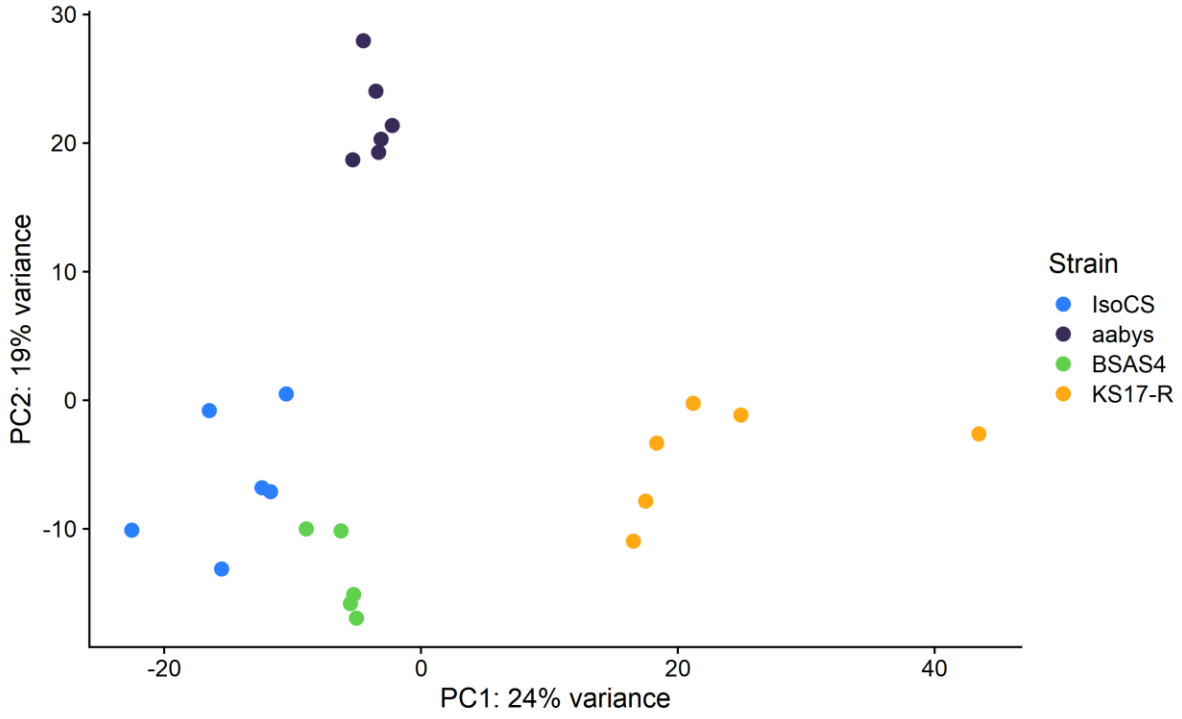


FIGURE 9: Principal component analysis of RNA-seq data from four strains indicates the gene expression of the BSAS4 strain is more similar to IsoCS than KS17-R.

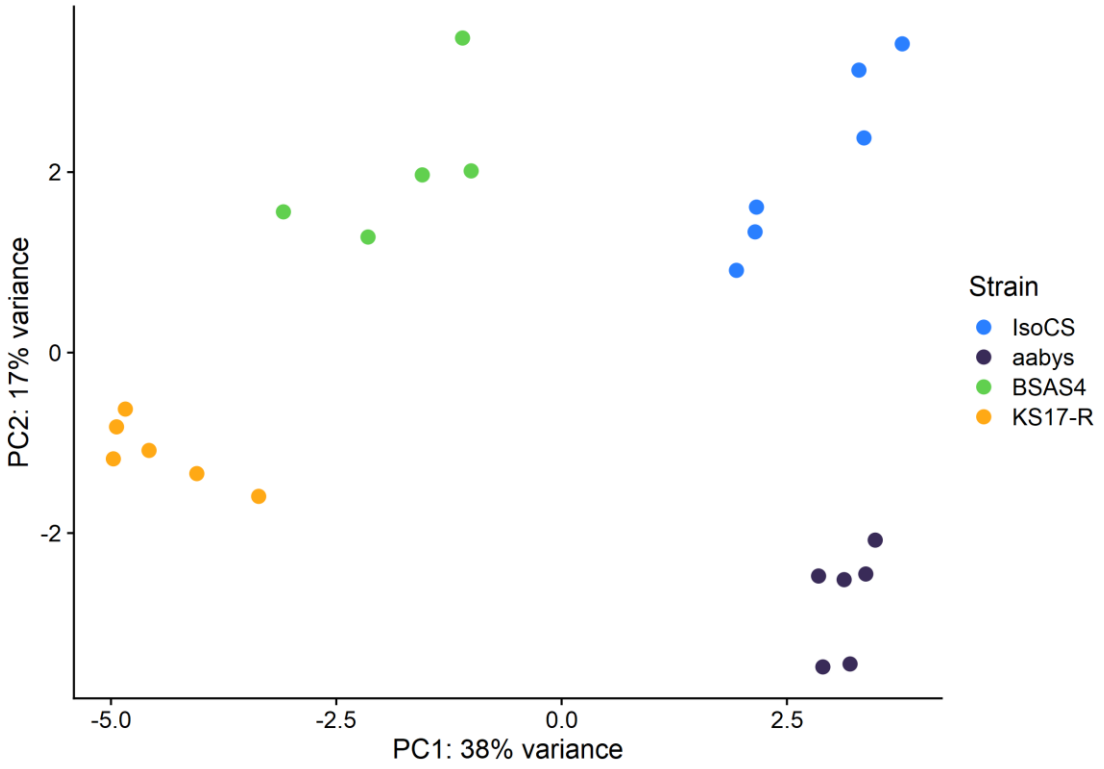


FIGURE 10: Principal component analysis of RNA-seq data for the CYP genes separates the resistant and susceptible strains along the PC1 (38% variance).

Differential expression analysis between unrelated strains generally results in many differentially expressed genes, and most will be unlinked to the phenotype of interest. The mixed BSAS4 strain was used as the comparison strain to IsoCS because of their more similar genetic background. With an FDR threshold of 0.05, 1,004 genes were called as differentially expressed between BSAS4 and IsoCS (Supplementary file 6). In comparison, 13,755 genes (13.7x more) were called as differentially expressed between KS17-R and IsoCS (Supplementary file 7), and 2,311 genes (2.3x more) were called as differentially expressed between aabys and IsoCS (Supplementary file 8).

The most significantly overexpressed gene in BSAS4 relative to IsoCS was *MdαE7*, an esterase on chromosome 2 that was 8.6 fold overexpressed. *MdαE7* is part of a cluster of 13 esterase genes, 5 of which are overexpressed in BSAS4. One or more esterases in this cluster could be responsible for the DEF suppressible resistance seen in KS17-R. Unfortunately the BSA did not yield a clear peak on chromosome 2, making it unclear if this esterase cluster was near a resistance locus or not. The second and fourth most significantly differentially expressed genes are two uncharacterized genes next to each other on chromosome 3, and both are underexpressed to a similar degree. One appears to be a derived from LTR retrotransposon *NINJA* (LOC109612838) and the other a protein of unknown function (LOC101888429) (though it appears that they may be more correctly annotated as one gene).

As KS17-R has resistance attributable to detoxification by CYPs, the relative expression of *CYPs* were of special interest. Of the 142 annotated *CYPs,* 55 are differentially expressed between KS17-R and IsoCS, and 16 of those are also differentially expressed between BSAS4 and KS17-R (Figure 11). In total BSAS4 had 21 *CYPs* differentially expressed relative to IsoCS, 18 upregulated and 3 downregulated (Figure 12). One differentially expressed *CYP* was on chromosome 1, 4 on chromosome 2, 5 on chromosome 3, and 11 on chromosome 5. Ranking the *CYPs* in order of fold-change in expression, the largest changes were observed for *CYP304A2* (+82.7-fold) and *CYP313D1* (+74.9-fold). *CYP304A2* has two copies in the genome, and while one was highly overexpressed in BSAS4 versus IsoCS, the other was slightly downregulated (0.57-fold). Interestingly, *CYP6D1*, which has been associated with resistance in house fly previously [70] does not appear to be overexpressed in KS17-R or BSAS4, even though KS17-R does have the *CYP6D1v1* allele (a 15 bp-deletion in the promoter).

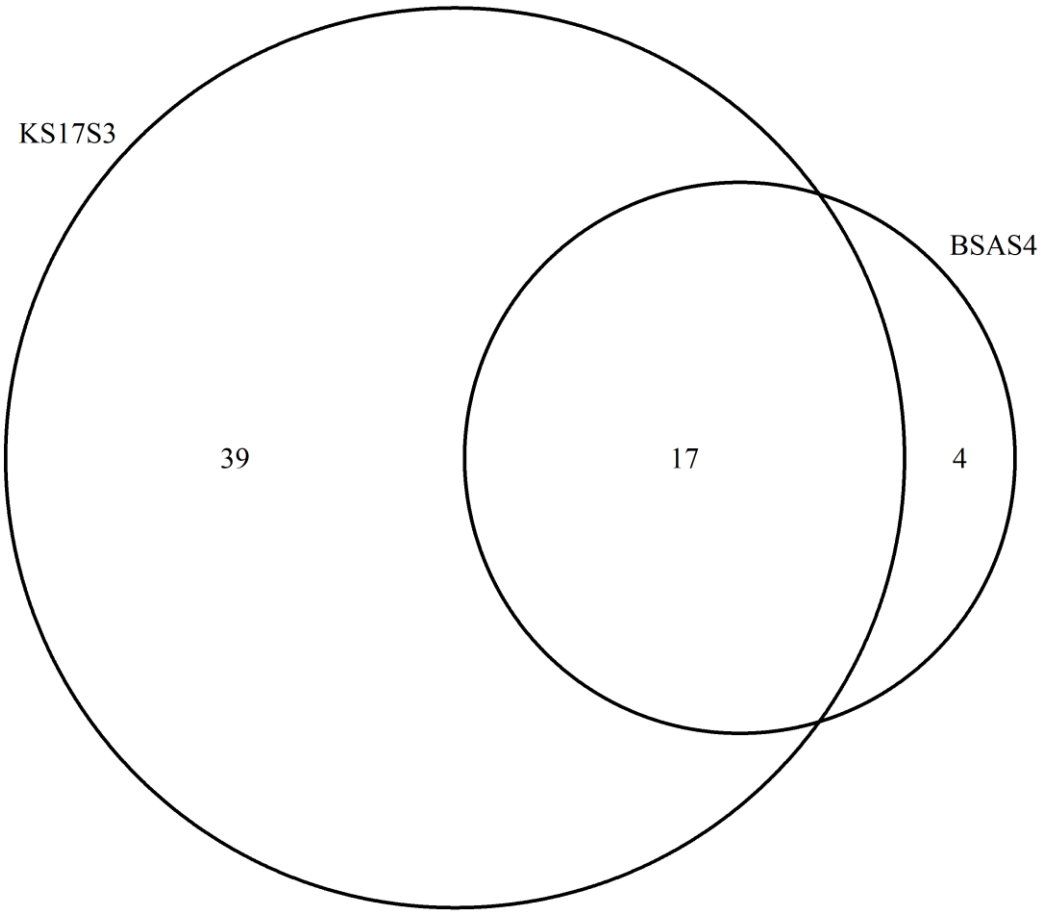


FIGURE 11: Of the 56 *CYPs* differentially expressed (p adjusted < 0.05) between KS17-R and IsoCS, 17 are also differentially expressed between BSAS4 and IsoCS.

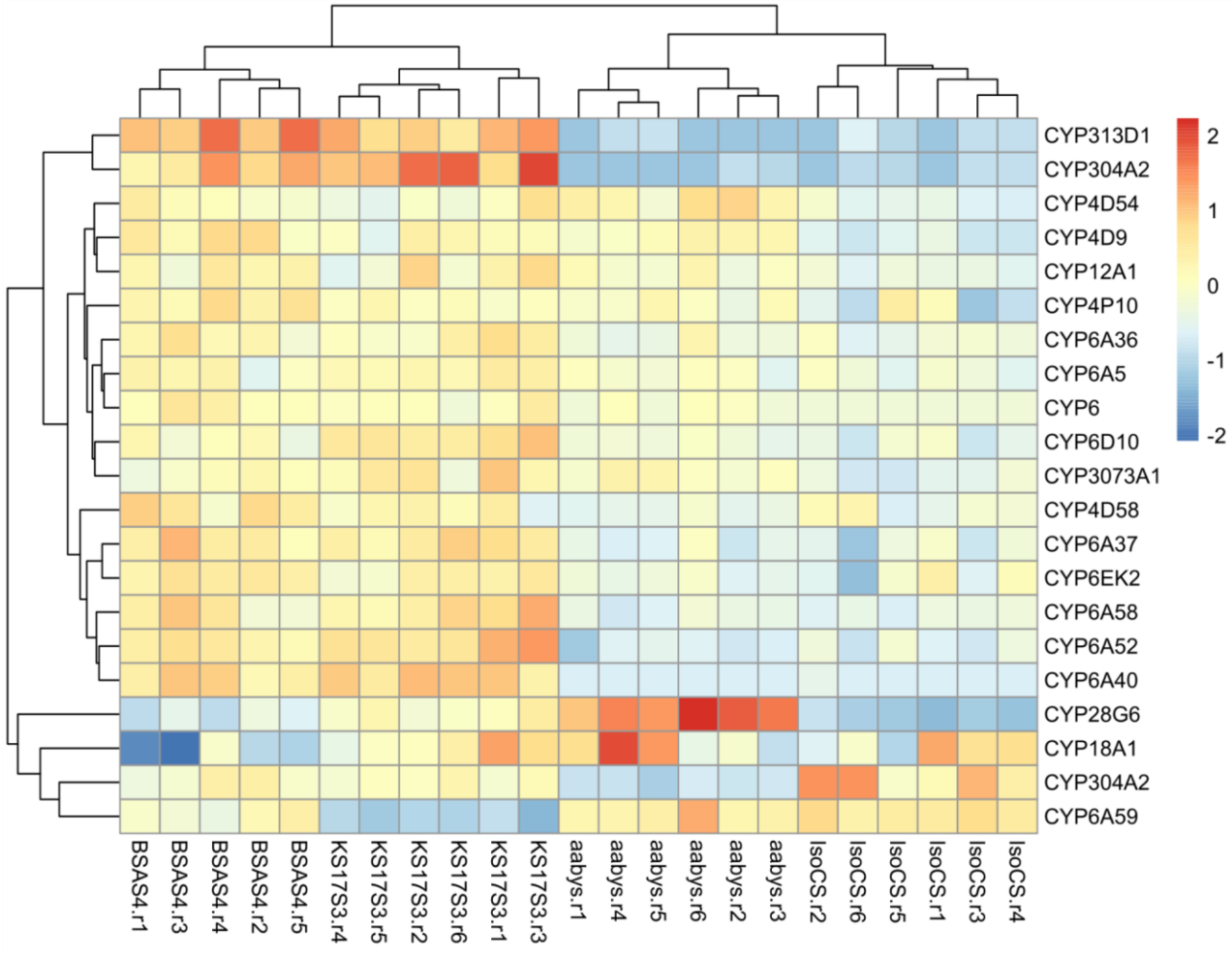


FIGURE 12: Between BSAS4 and IsoCS, 21 CYPs are differentially expressed (p adjusted < 0.05). Expression normalized within gene using the variance-stabilizing transformation from DESeq2 [63].

Of the 11 differentially expressed chromosome 5 *CYPs*, 8 of those occur at the chromosome 5 resistance locus on scaffold 25 (the *CYP6As* in Figure 12). All except one are overexpressed (*CYP6A59*). Of the *CYPs* at the resistance locus *CYP6A40* is the most significantly (~78-fold) overexpressed in BSAS4, followed by a currently nameless *CYP6* (LOC101892417) (~7 fold). This gene was annotated as a pseudogene in the NCBI annotation, but appears to have an additional start codon in the second exon that will result in a full protein.

CYP6A40 is the *CYP* closest to the right-hand breakpoint of the duplication (Figure 7), making it an intriguing candidate for involvement in the resistance.

2.4 Discussion

Bulked segregant analysis against a newly generated *Musca domestica* genome successfully mapped pyrethroid resistance in KS17-R to a known causal gene on chromosome 3 and an unknown resistance locus on chromosome 5. The relative ranking of chromosomal effect sizes from the linkage analysis (3 >5 >2 >1) can also be observed in the BSA signal (Figure 4). This provides some indication that although the BSA signal is weaker on chromosomes 2 and 1, it does represent the presence of potential resistance loci. Here, BSA was successful in identifying multiple resistance loci spread out over different chromosomes. Once multiple loci of interest are on a single chromosome, identifying regions of interest becomes much more difficult, especially when signal is spread across multiple scaffolds as was seen for chromosome 1. Another difficulty is distinguishing between a major locus and a minor locus on the same chromosome. For chromosome 3, the signal associated with the *Vssc* dominates most of chromosome 3 so a second resistance locus cannot be ruled out completely. Alternatively two loci of similar intensity on the same chromosome may result in a merging of signal, and in this case the individual peaks may no longer be visible. It seems possible that the signal observed over much of chromosome 2 represents more than one locus associated with resistance.

The contiguity of the reference genome is vital to a successful BSA experiment. The use of long read sequencing gave us an improved version of the house fly genome that was critically important in the success of our BSA. In the 2014 house fly genome, only 35 scaffolds have a length of greater than 1 Mb (the window size used for the G` analysis) [14]. However, with the

new genome we were able to see the *Vssc* peak stretched over the whole of scaffold 20 (~14 Mb), and the chromosome 5 peak can be seen across 3 scaffolds (>30 Mb) (Figure 3). Across a small number of scaffolds, the relative strength of signal between scaffolds can be compared to prioritize regions of interest, but this quickly becomes untenable.

The chromosome 5 resistance locus includes a cluster of 25 *CYP* genes, 8 of which were differentially expressed between BSAS4 and IsoCS. A duplication event over part of this cluster was observed at an allele frequency of 0.5 in the KS17-R long reads (Figure 7). The presence of a duplication itself could be causal for resistance (eg by altering gene expression of one or more of the *CYPs*), or a variant on the duplication could be causal for the resistance, but these alternatives would be difficult to functionally separate. Duplication of *CYPs* associated with pyrethroid resistance have been reported in other insects ([71, 72]. *CYP6A40* and another *CYP6* had the greatest degree of overexpression of genes in the cluster, positioning them as potentially contributing more to the resistance phenotype. It is clear from the BSA results that this region is linked to the resistance phenotype, and the SV calling results indicates that this region is structurally complicated in KS17-R. There is some indication that SVs in this region may involve TE movement (from motifs of transposable elements present around the variants). In order to resolve the role of structural variants in this region on CYP expression and resistance, further work is necessary. Potential avenues include establishing clear linkage of the chromosome 5 locus to CYP-mediated resistance with synergists, determining more clearly the sequence of the duplicated region, characterization of TEs within the region, and linking the overexpression of *CYP6A40* and the other *CYP6* to structural variation.

The many potential genomic changes that can result in changes gene expression makes resistance due to overexpression of detoxification genes a complex phenotype to map. In the

simplest cases, the overexpression will make to the gene itself. This can be attributed to *cis* regulatory changes, for example a change in the promoter that affects transcription factor binding (eg a deletion in the promoter of *CYP6D1* that prevents binding of the transcriptional repressor *Gfi-1* [73]) or potentially copy number variation of the gene itself. In the more difficult case, overexpression will map elsewhere in the genome, to a region altering the regulation of the gene *in trans*.

Detoxification genes (including some members of the CYP, GST, and esterase families) are known to be regulated by the insect ortholog of *Nrf2, Cap n' collar ortholog C (CnCC*). In normal conditions in mammals, *Nrf2* remains in the cytoplasm, bound to a ubiquitin ligase *Keap1* and marked for destruction, but under conditions of stress *Nrf2* is released and translocates to the nucleus, where it heterodimerizes with *Maf-S*, and this heterodimer binds to the promotor motifs coined antioxidant response elements (AREs) resulting in transcription of target genes [74]. Studies in a variety of insects indicate a conserved role of *CnCC*/*Maf* in regulating detoxification genes (for a summary see [75]). This pathway provides multiple clear candidates for involvement in *CYP* (or esterase) expression changes, but also some more subtle (eg proteins involved in transport of *CnCC* into the nucleus [76], proteins involved in the ubiquitin-proteasome pathway, etc.). None of these main pathway genes are differentially expressed between BSAS4 and IsoCS, though *CnCC* it is slightly (-0.8-fold) downregulated in KS17-R.

Synergist bioassays indicated hydrolase-mediated resistance in addition to the CYP-mediated resistance. The overexpression of *MdαE7*, as well as four other esterases on chromosome 2 make them potential candidate genes. *MdαE7* and *MdαE9* were also overexpressed in the pyrethroid resistant ALHF relative to two unrelated susceptible strains

aabys and CS, and the overexpression phenotype of *MdαE7* was mapped to chromosomes 2 and 5 [20]. The relatively stable G` values over chromosome 2 make it unclear whether these genes lies at a resistance locus. A line containing the chromosome 2 resistance locus within the aabys background could be used to test whether the hydrolase-mediated resistance of KS17-R maps to chromosome 2.

BSA of a multigenic pyrethroid resistance phenotype successfully identified resistance loci on chromosome 3 at the *Vssc* and on chromosome 5 at a cluster of *CYP* genes, along with more disperse signal on chromosomes 1 and 2. This was made possible by a newly generated third-generation genome assembly. The existence of resistance loci on 3, 5, 2, and 1 is supported by a chromosomal linkage analysis with morphological markers. Bioassays with insecticide inhibitors indicate increased detoxification by CYPs and esterases contribute to the resistance phenotype, and RNA-seq confirms overexpressed *CYPs* and esterases of interest. Long read Nanopore sequencing indicates the presence of complicated structural variants at the chromosome 5 resistance locus, and RNA-seq confirms overexpression of CYPs at this loci between related susceptible and resistant strains.

Data availability

The house fly strains used in this study (aabys, IsoCS, and KS17-R) are available on request. Short-read and long-read sequencing data is deposited at NCBI SRA under BioProject PRJNA684355, and the aabys genome assembly is under BioProject PRJNA681893. Supplemental data files are available at http://blogs.cornell.edu/scott/additional-information.

Acknowledgements

Sequencing was performed by the Biotechnology Resource Center (BRC) Genomics Facility at the Cornell Institute of Biotechnology. This project was supported through funding from the Cornell Biotechnology Resource Center Seed Fund, Griswold Endowment Fund, and the Sarkaria Institute of Insect Physiology & Toxicology Fellowship (Cornell Entomology Department). The aabys genome assembly was performed by Chaoxian Geng, Mike Loso, Victor Llaca, Kevin Fengler, and Achal Restogi of Corteva Agrisciences.

REFERENCES

[1] J.G. Scott, Life and death at the voltage-sensitive sodium channel: Evolution in response to insecticide use, Ann. Rev. Entomol., 64 (2019) 243-257.
[2] J.G. Scott, C.A. Leichter, F.D. Rinkevich, S.A. Harris, C. Su, L.C. Aberegg, R. Moon, C.J. Geden, A.C. Gerry, D. Taylor, R.L. Byford, W. Watson, G. Johnson, D. Boxler, L. Zurek, Insecticide resistance in house flies from the United States: Resistance levels and frequency of pyrethroid resistance alleles, Pestic. Biochem. Physiol., 107 (2013) 377-384.
[3] J.C. Freeman, D.H. Ross, J.G. Scott, Insecticide resistance monitoring of house fly populations from the United States, Pestic. Biochem. Physiol., 158 (2019) 61-68.
[4] J.G. Scott, Evolution of pyrethroid resistance in *Musca domestica*, Pest Manag. Sci., 73 (2017) 716-722.
[5] J.G. Scott, G.P. Georghiou, The biochemical genetics of permethrin resistance in the Learn-PyR strain of house fly, Biochem. Genet., 24 (1986) 25-37.
[6] B.C. Fine, P.J. Godin, E.M. Thain, Penetration of pyrethrin I labelled with carbon-14 intosusceptible and pyrethriod resistant houseflies, Nature, 199 (1963) 927-928.
[7] N. Liu, X. Yue, Insecticide resistance and cross-resistance in the house fly (Diptera: Muscidae), J. Econ. Entomol., 93 (2000) 1269-1275.
[8] A.W. Farnham, K.E. O'Dell, I. Denholm, R.M. Sawicki, Factors affecting resistance to insecticides in house-flies, *Musca domestica* L. (Diptera: Muscidae). III. Relationship between the level of resistance to pyrethroids, control failure in the field and the frequency gene *kdr*, Bull. ent. Res., 74 (1984) 581-589.
[9] L. Zhang, X. Gao, P. Liang, Beta-cypermethrin resistance associated with high carboxylesterase activities in a strain of house fly, *Musca domestica* (Diptera: Muscidae), Pestic. Biochem. Physiol., 89 (2007) 65-72.
[10] T. Tomita, N. Liu, F.F. Smith, P. Sridhar, J.G. Scott, Molecular mechanisms involved in increased expression of a cytochrome P450 responsible for pyrethroid resistance in the housefly, *Musca domestica*, Insect Mol. Biol., 4 (1995) 135-140.
[11] K. Itokawa, O. Komagata, S. Kasai, Y. Okamura, M. Masada, T. Tomita, Genomic structures of *Cyp9m10* in pyrethroid resistant and susceptible strains of *Culex quinquefasciatus*, Insect Biochem. Mol. Biol., 40 (2010) 631-640.
[12] A.M. Puinean, S.P. Foster, L. Oliphant, I. Denholm, L.M. Field, N.S. Millar, M.S. Williamson, C. Bass, Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*, PLoS Genetics, 6 (2010) e1000999.
[13] J.-B. Berge, R. Feyereisen, M. Amichot, Cytochrome P450 monooxygenases and insecticide resistance in insects, Phil. Trans. R. Soc. Lond. B, 353 (1998) 1701-1705.
[14] J.G. Scott, W.C. Warren, L.W. Beukeboom, D. Bopp, A.G. Clark, S.D. Giers, M. Hediger, A.K. Jones, S. Kasai, C.A. Leichter, M. Li, R.P. Meisel, P. Minx, T.D. Murphy, D.R. Nelson, W.R. Reid, F.D. Rinkevich, H.M. Robertson, T.B. Sackton, D.B. Sattelle, F. Thibaud-Nissen, C. Tomlinson, L. van de Zande, K.K.O. Walden, R.K. Wilson, N. Liu, Genome of the house fly (*Musca domestica* L), a global vector of diseases with adaptations to a septic environment, Genome Biol. , 15 (2014) 466.
[15] J.G. Scott, Investigating mechanisms of insecticide resistance: methods, strategies and pitfalls, in: R.T. Roush, B. Tabashnik (Eds.) Pesticide Resistance in Arthropods, Chapman and Hall, New York, 1990, pp. 39-57.

[16] S.M. Valles, P.G. Koehler, R.J. Brenner, Antagonism of fipronil toxicity by piperonyl butoxide and *S*,*S*,*S*-tributyl phosphorotrithioate in the German cockroach (Dictyoptera: Blattellidae), J. Econ. Entomol., 90 (1997) 1254-1258.
[17] Z. Wen, J.G. Scott, Genetics and mechanisms of fipronil cross-resistance in the LPR strain of house fly, *Musca domestica.*, Pestic. Sci., 55 (1999) 988-992.
[18] R.M. Sawicki, A.L. Devonshire, A.W. Farnham, K.E. O'Dell, G.D. Moores, I. Denholm, Factors affecting resistance to insecticides in house-flies, *Musca domestica* L. (Diptera: Muscidae). II. Close linkage on autosome 2 between an esterase and resistance to trichlorphon and pyrethroids, Bull. ent. Res., 74 (1984) 197-206.
[19] C. Claudianos, R.J. Russell, J.G. Oakeshott, The same amino acid substitution in orthologous esterases confers organophosphate resistance on the house fly and a blowfly, Insect Biochem. Mol. Biol., 29 (1999) 675-686.
[20] X. Feng, M. Li, N. Liu, Carboxylesterase genes in pyrethroid resistant house flies, Musca domestica, Insect Biochem Mol Biol, 92 (2018) 30-39.
[21] H. Sun, N. Buchon, J.G. Scott, *Mdr65* decreases toxicity of multiple insecticides in *Drosophila melanogaster* Insect Biochem. Molec. Biol. , 89 (2017) 11-16.
[22] X. Yang, W. Chen, X. Song, X. Ma, R.O. Cotto-Rivera, W. Kain, H. Chu, Y.R. Chen, Z. Fei, P. Wang, Mutation of ABC transporter ABCA2 confers resistance to Bt toxin Cry2Ab in *Trichoplusia ni*, Insect Biochem Mol Biol, (2019) 103209.
[23] J. Wang, H. Ma, S. Zhao, J. Huang, Y. Yang, B.E. Tabashnik, Y. Wu, Functional redundancy of two ABC transporter proteins in mediating toxicity of Bacillus thuringiensis to cotton bollworm, PLoS Pathog, 16 (2020) e1008427.
[24] C. Zou, P. Wang, Y. Xu, Bulked sample analysis in genetics, genomics and crop improvement, Plant Biotechnol J, 14 (2016) 1941-1955.
[25] N. Wybouw, A.H. Kurlovs, R. Greenhalgh, A. Bryon, O. Kosterlitz, Y. Manabe, M. Osakabe, J. Vontas, R.M. Clark, T. Van Leeuwen, Convergent evolution of cytochrome P450s underlies independent origins of keto-carotenoid pigmentation in animals, Proc Biol Sci, 286 (2019) 20191039.
[26] A.H. Kurlovs, S. Snoeck, O. Kosterlitz, T. Van Leeuwen, R.M. Clark, Trait mapping in diverse arthropods by bulked segregant analysis, Current Opinion in Insect Science, (2019).
[27] P. Demaeght, E.J. Osborne, J. Odman-Naresh, M. Grbi, R. Nauen, H. Merzendorfer, R.M. Clark, T. Van Leeuwen, High resolution genetic mapping uncovers chitin synthase-1 as the target-site of the structurally diverse mite growth inhibitors clofentezine, hexythiazox and etoxazole in *Tetranychus urticae*, Insect Biochem. Mol. Biol., 51 (2014) 52-61.
[28] H. Bastide, J.D. Lange, J.B. Lack, A. Yassin, J.E. Pool, A variable genetic architecture of melanic evolution in *Drosophila melanogaster*, Genetics, 204 (2016) 1307-1319.
[29] K.M. Benowitz, J.M. Coleman, L.M. Matzkin, Assessing the Architecture of Drosophila mojavensis Locomotor Evolution with Bulk Segregant Analysis, G3 (Bethesda), 9 (2019) 1767-1775.
[30] R.L. Hamm, J.-R. Gao, G.G.-H. Lin, J.G. Scott, Selective advantage for III^M males over Y^M males in cage competition, mating competition and pupal emergence in *Musca domestica* L. (Diptera: Muscidae), Environ. Entomol., 38 (2009) 499-504.
[31] Y. Takada, T. Hiroyoshi, M. Hirano, Linkage group analysis of permethrin resistance in the Miyakonojo colony of the housefly, *Musca domestica* L. (Diptera: Muscidae), Appl. Entomol. Zool., 23 (1988) 122-126.
[32] D.J. Finney, Probit Analysis, 3rd ed., Cambridge University Press, Cambridge, UK, 1971.

[33] W.S. Abbott, A method of computing the effectiveness of an insecticide, J. Econ. Entomol., 18 (1925) 265-267.
[34] M. Tsukamoto, Methods for linkage-group determination of insecticide-resistance factors in the housefly, Botyu-Kagaku, 29 (1964) 51-59.
[35] R. Hamm, T. Shono, J.G. Scott, A cline in frequency of autosomal males is not associated with insecticide resistance in house fly (Diptera: Muscidae), J. Econ. Entomol., 98 (2005) 171-176.
[36] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation, Genome Res, 27 (2017) 722-736.
[37] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat Methods, 10 (2013) 563-569.
[38] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One, 9 (2014) e112963.
[39] M.J. Roach, S.A. Schmidt, A.R. Borneman, Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies, BMC Bioinformatics, 19 (2018) 460.
[40] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics, 29 (2013) 1072-1075.
[41] M. Seppey, M. Manni, E.M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness, Methods Mol Biol, 1962 (2019) 227-245.
[42] K.M. Kapheim, H. Pan, C. Li, S.L. Salzberg, D. Puiu, T. Magoc, H.M. Robertson, M.E. Hudson, A. Venkat, B.J. Fischman, A. Hernandez, M. Yandell, D. Ence, C. Holt, G.D. Yocum, W.P. Kemp, J. Bosch, R.M. Waterhouse, E.M. Zdobnov, E. Stolle, F.B. Kraus, S. Helbing, R.F.A. Moritz, K.M. Glastad, B.G. Hunt, M.A.D. Goodisman, F. Hauser, C.J.P. Grimmelikhuijzen, D.G. Pinheiro, F.M.F. Nunes, M.P.M. Soares, E.D. Tanaka, Z.L.P. Simoes, K. Hartfelder, J.D. Evans, S.M. Barribeau, R.M. Johnson, J.H. Massey, B.R. Southey, M. Hasselmann, D. Hamacher, M. Biewer, C.F. Kent, A. Zayed, C. Blatti, S. Sinha, J.S. Johnston, S.J. Hanrahan, S.D. Kocher, J. Wang, G.E. Robinson, G. Zhang, Genomic signatures of evolutionary transitions from solitary to group living, Science, 348 (2015) 1139-1143.
[43] R.P. Meisel, J.G. Scott, Using genomic data to study insecticide resistance in the house fly, *Musca domestica*, Pestic. Biochem. Physiol., 151 (2018) 76-81.
[44] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinformatics, 12 (2011) 491.
[45] M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, Bioinformatics, 24 (2008) 637-644.
[46] A. Smit, R. Hubley, RepeatModeler Open-1.0, 2008.
[47] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat Biotechnol, 29 (2011) 644-652.
[48] L.B. Smith, R. Tyagi, K. Kasai, J.G. Scott, CYP-mediated permethrin resistance in *Aedes aegypti* and evidence for *trans*-regulation, PLoS Negl Trop Dis, 12 (2018) e0006933.

[49] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics, 30 (2014) 2114-2120.
[50] H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 2013.
[51] Broad Institute, Picard toolkit, Broad Institute, 2019.
[52] R. Poplin, V. Ruano-Rubio, M.A. DePristo, T.J. Fennell, M.O. Carneiro, G.A. Van der Auwera, D.E. Kling, L.D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M.J. Daly, B. Neale, D.G. MacArthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples, BioRxiv, (2018).
[53] P. Cingolani, A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, D.M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3, Fly (Austin), 6 (2012) 80-92.
[54] B.N. Mansfeld, R. Grumet, QTLseqr: An R package for bulk segregant analysis with next-generation sequencing, Plant Genome, 11 (2018).
[55] P.M. Magwene, J.H. Willis, J.K. Kelly, The statistics of bulk segregant analysis using next generation sequencing, PLOS Comput. Biol., 7 (2011) e1002255.
[56] J. Koster, S. Rahmann, Snakemake--a scalable bioinformatics workflow engine, Bioinformatics, 28 (2012) 2520-2522.
[57] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data [Online], 2010.
[58] B. Bushnell, BBMap, 2014.
[59] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics, 29 (2013) 15-21.
[60] P. Ewels, M. Magnusson, S. Lundin, M. Kaller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics, 32 (2016) 3047-3048.
[61] A. Shumate, S.L. Salzberg, Liftoff: an accurate gene annotation mapping tool, bioRxiv, (2020).
[62] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2020.
[63] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol, 15 (2014) 550.
[64] M. Stephens, False discovery rates: a new deal, Biostatistics, 18 (2017) 275-294.
[65] R. Kolde, M.R. Kolde, Package 'pheatmap', R Package, 1 (2015) 790.
[66] W. De Coster, S. D'Hert, D.T. Schultz, M. Cruts, C. Van Broeckhoven, NanoPack: visualizing and processing long-read sequencing data, Bioinformatics, 34 (2018) 2666-2669.
[67] W. De Coster, P. De Rijk, A. De Roeck, T. De Pooter, S. D'Hert, M. Strazisar, K. Sleegers, C. Van Broeckhoven, Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome, Genome Res, 29 (2019) 1178-1187.
[68] F.J. Sedlazeck, P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, M.C. Schatz, Accurate detection of complex structural variations using single-molecule sequencing, Nat Methods, 15 (2018) 461-468.
[69] H. Sun, S. Kasai, J.G. Scott, Two novel house fly *Vssc* mutations, D600N and T929I, give rise to new insecticide resistance alleles, Pestic. Biochem. Physiol. , 143 (2017) 116-121.
[70] N. Liu, J.G. Scott, Increased transcription of *CYP6D1* causes cytochrome P450-mediated insecticide resistance in house fly, Insect Biochem. Molec. Biol., 28 (1998) 531-535.

[71] S. Kasai, O. Komagata, K. Itokawa, T. Shono, L.C. Ng, M. Kobayashi, T. Tomita, Mechanisms of pyrethroid resistance in the dengue mosquito vector, *Aedes aegypti*: target site insensitivity, penetration, and metabolism, PLoS Negl. Trop. Dis., 8 (2014) e2948.
[72] K. Itokawa, O. Komagata, S. Kasai, M. Masada, T. Tomita, *Cis*-acting mutation and duplication: history of molecular evolution in a P450 haplotype responsible for insecticide resistance in *Culex quinquefasciatus*, Insect Biochem. Mol. Bio., 41 (2011) 503-512.
[73] J.G. Scott, N. Liu, Z. Wen, F.F. Smith, S. Kasai, C.E. Horak, House fly cytochrome P450 *CYP6D1*: 5 prime flanking sequences and comparison of alleles., Gene, 226 (1999) 347-353.
[74] Q. Ma, Role of nrf2 in oxidative stress and toxicity, Annu Rev Pharmacol Toxicol, 53 (2013) 401-426.
[75] C.S. Wilding, Regulating resistance: CncC:Maf, antioxidant response elements and the overexpression of detoxification genes in insecticide resistance, Current Opinion in Insect Science, 27 (2018) 89-96.
[76] M. Ge, T. Zhang, M. Zhang, L. Cheng, Ran participates in deltamethrin stress through regulating the nuclear import of Nrf2, Gene, (2020) 145213.

2.5 Supplemental information

1. Supplementary materials and methods

1.1. *De novo* transciptome assembly with Trinity for gene annotation

RNA-seq data was chosen to cover a diversity of genotypes, life stages, tissues, and both sexes. A total of 3,771,164,664 reads were used in total (see Table S4). Raw reads were trimmed using Trimmommatic v0.39 using the parameters recommended in the manual [49]. Trinity v2.8.4 was run with *in silico* normalization on single and paired-end data separately.

1.2. High molecular weight DNA extraction for Nanopore sequencing

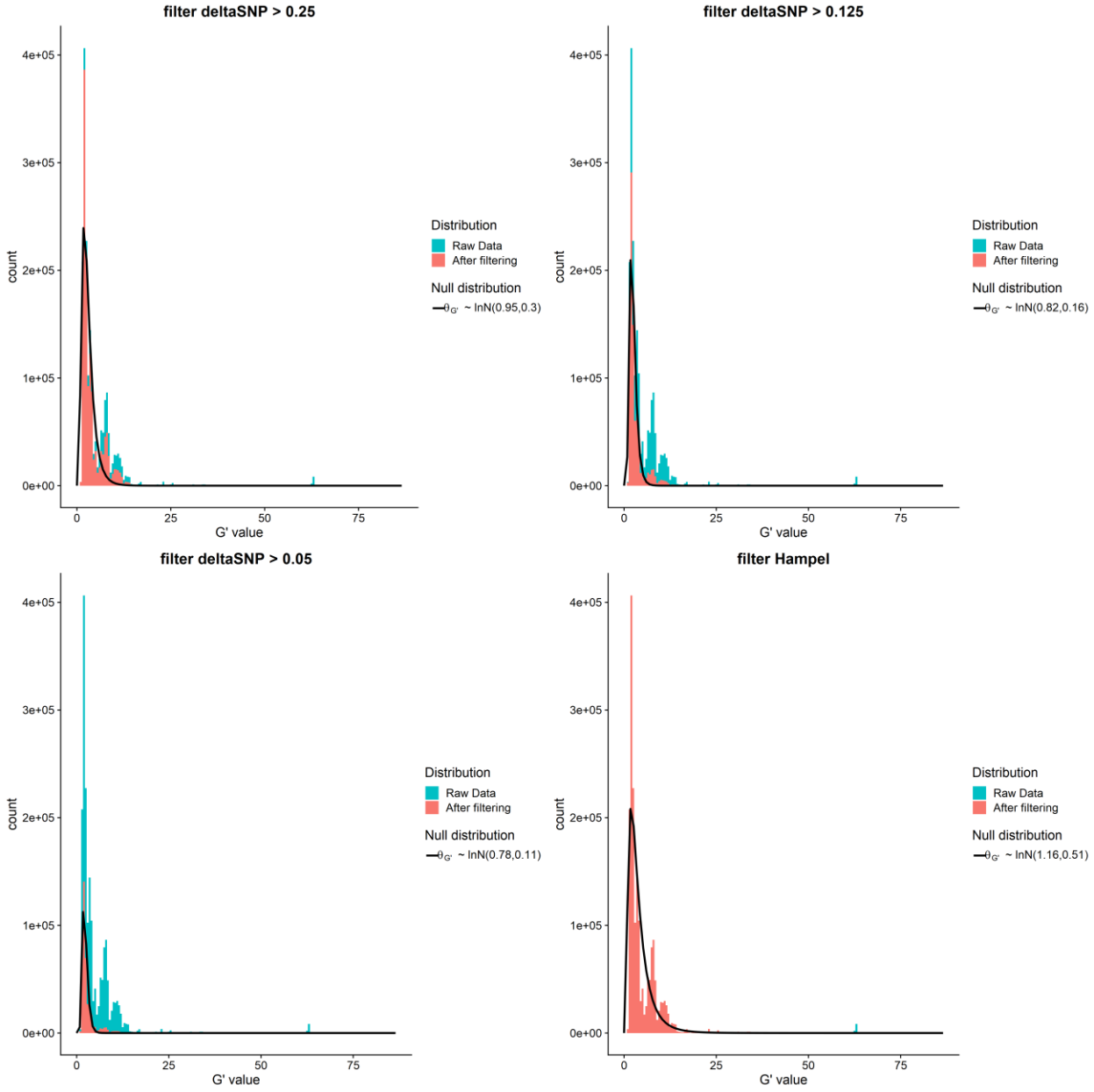
High molecular weight DNA was extracted using a protocol adapted for insect tissue from the Agilent RecoverEase DNA Isolation Kit (Agilent Technologies, Inc., Santa Clara, CA, USA) which includes nuclei enrichment, digestion with RNAse and proteinase K, and sample cleanup through dialysis. Major changes included increasing the time for proteinase K digestion, increasing the concentration of proteinase K, and adding a phenol:chloroform separation between digestion and dialysis. Flies removed from the pupal case (pharate adults) before emergence were found to provide the best yield and purity of DNA. Pharate adults were removed from their pupal case approximately 2 d before emergence (adult structures are visible but there is little pigmentation outside of the eyes), and their abdomens were removed in a petri dish with 95% ethanol. Individuals were ground in 0.8 mL of lysis buffer (140 mM NaCl, 3 mM KCl, 1 mM EDTA, 350 mM sucrose, with pH 8.3) on ice with a plastic pestle, with 6 individuals pooled into one extraction. Homogenate was strained through a 100-µm cell strainer (VWR, Radnor, PA, USA) into a 50-mL conical tube (VWR), and lysis buffer was added to a final volume of 8 mL. Conical tubes were centrifuged at 4°C for 12 min at 1100 x g. The supernatant was removed from the resulting pellet, and the tube was inverted to dry 1 min. One hundred µL of digestion

buffer (12 mM Na_2HPO_4, 136 mM NaCl, 2.68 mM KCl, 1.5 mM KH_2PO_4, and 10 mM EDTA) was combined with 2 µL of Invitrogen RNAse Cocktail (ThermoFisher Scientific) and 100 µL of this mixture was added to the cell pellet. Proteinase K solution (20 mg/mL proteinase K (Promega Corp., Madison, WI, USA), 2% w/v sodium dodecyl sulfate, 100 mM EDTA pH 7.5) was pre-warmed to 55°C for 4 min and 100 µL were added to the pellet, which was swirled gently to dislodge it from the bottom of the tube.

The pellet was incubated at 55°C for 2.5 h swirling gently to mix about every half hour. Add 200 µL TE pH 8.0, swirl gently to mix, and allow to cool to room temperature before proceeding with phenol:chloroform extraction.

Transfer cell lysate to 1.7 mL microcentrifuge tube, and add an equal volume of phenol:chloroform:isoamyl alcohol (Ambion Inc., Austin, TX, USA). Gently invert to mix (it should take ~30 min to emulsify). Centrifuge 10 min at 4.5 x g, and transfer aqueous phase gently to a 2 mL screw-cap tube. Add 1/20 volume of 4 M NaCl and 2 volumes of absolute ethanol and swirl gently to mix. Rolling the tube on a hot dog roller (Gold Medal Products Co., Cincinnati, OH) for 30 min improved clumping of DNA strands. Allow sample to precipitate overnight at 4°C without mixing. DNA should be visible as a tangled clump. Remove ethanol, avoiding the clump of DNA, and wash twice with 70 % ethanol. Centrifuge 5 sec at 14.1 x g, and remove as much of 70% ethanol as possible, dry tube with cap open until traces of ethanol have evaporated. Add 100 µL TE and allow pellet to resuspend at 4°C without mixing for at least two days before proceeding to dialysis. Dialyze samples in a Genomic Tube-O-Dialyzer tube (G Biosciences, St. Louis, MO, USA) according to manufacturer's instructions.

Figure S1: The statistical significance of BSA peaks was assessed by as described in [55] . G` values are compared against a null distribution of G` fit using the mean and variance of the data minus outliers (which are assumed to be QTL). Multiple thresholds for outlier determination were tested are shown below to illustrate the effect on the null distribution fitting, but $\Delta_{SNP} > 0.25$ was used for the analysis to balance the fit of the distribution against overfiltering of the data.



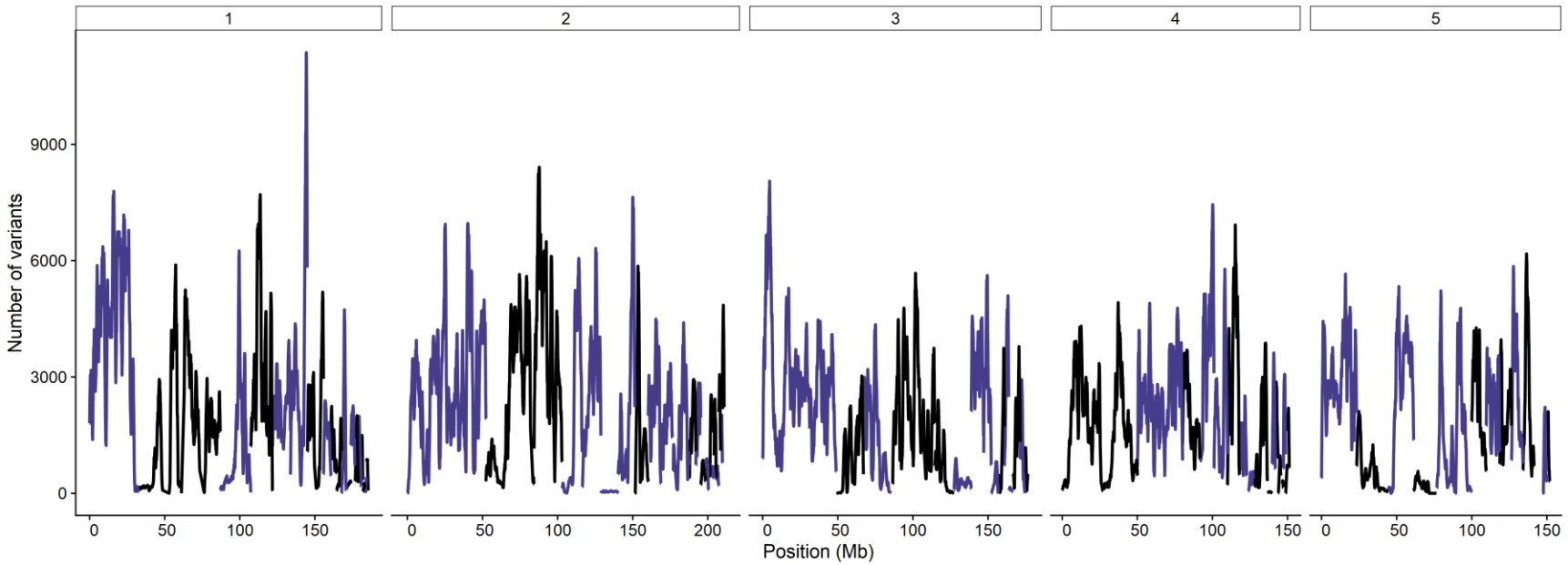


Figure S2: Number of variants included in the windows used to calculate G` are not evenly distribution throughout the genome.

Alternation of colors represents breaks between scaffolds.

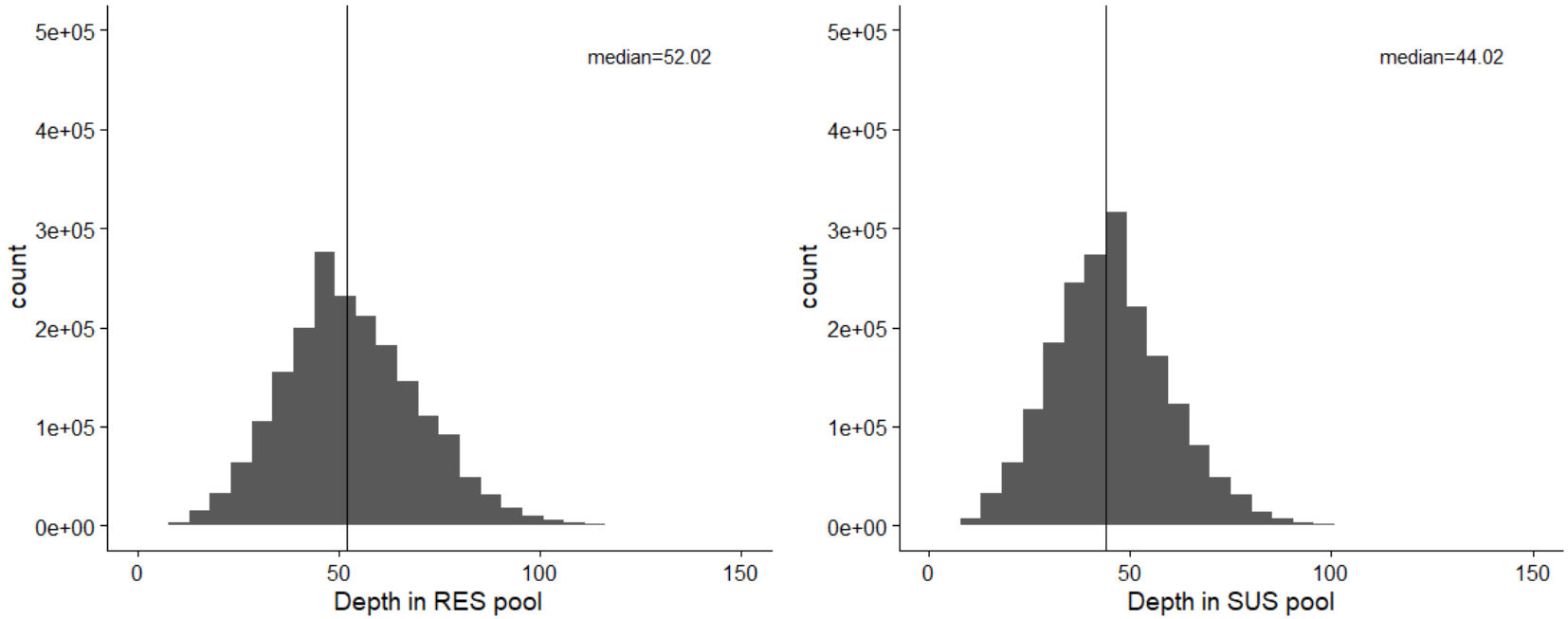


Figure S3: Coverage distribution of variants called in the BSA RES1-3 and SUS1-3 pools. Variants with coverage <20 reads over the 3 pools were filtered for calculating G` and Δ_{SNP}.

Figure S4: A ~50 kb region near the peak G` value (~2.344 Mb) for chromosome 5 shows a reduction in coverage for the IsoCS male parent, visualized here as a lack of variants called for that library. Top track shows SNPs and bottom track indels. Color represents genotype called (green is REF/REF, blue REF/ALT, red ALT/ALT), which is based on a diploid model and not accurate for the pooled samples.

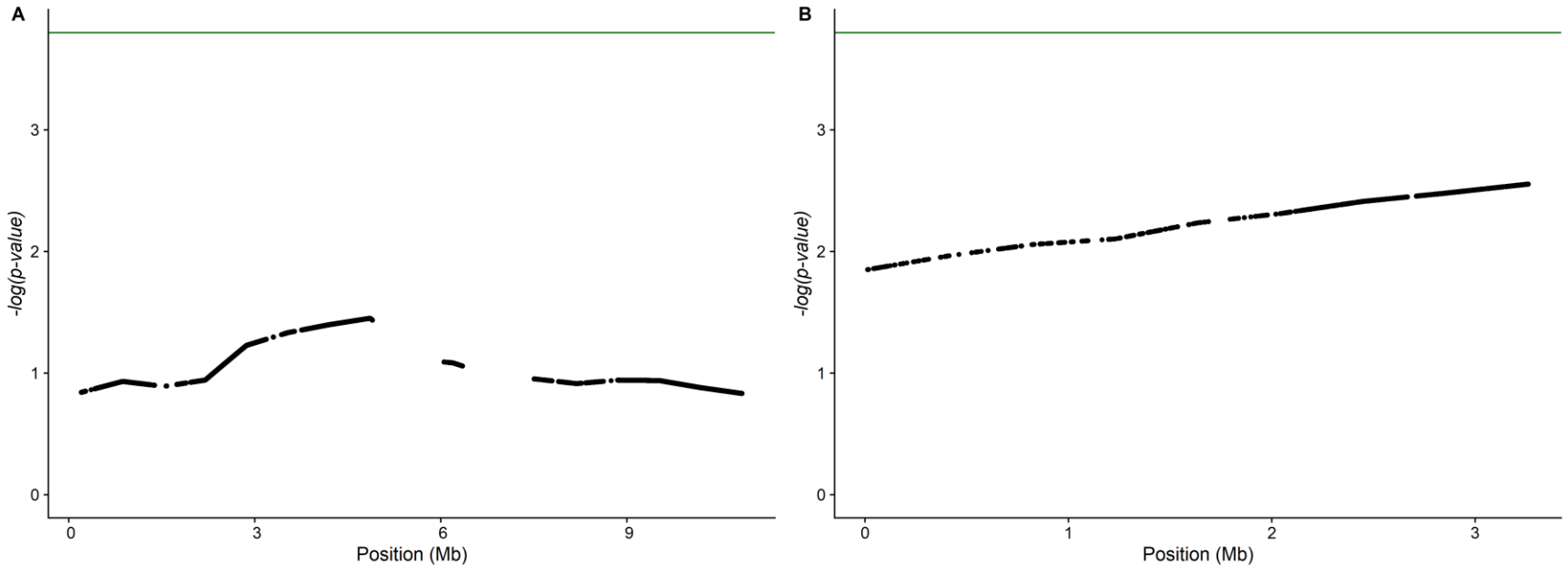


FIGURE S5: Scaffolds on chromosome 1 with signal of G` above the baseline of the rest of the chromosome are difficult to interpret. Scaffold 34 (A) has a gap in the middle, so it's unclear whether the G` value should hit a maximum within the gap. The signal on scaffold 78 (B) appears to peak on the connecting scaffold, though which scaffold that it is unclear.

Figure S6: Length distribution of Nanopore reads for the IsoCS and KS17-R libraries.

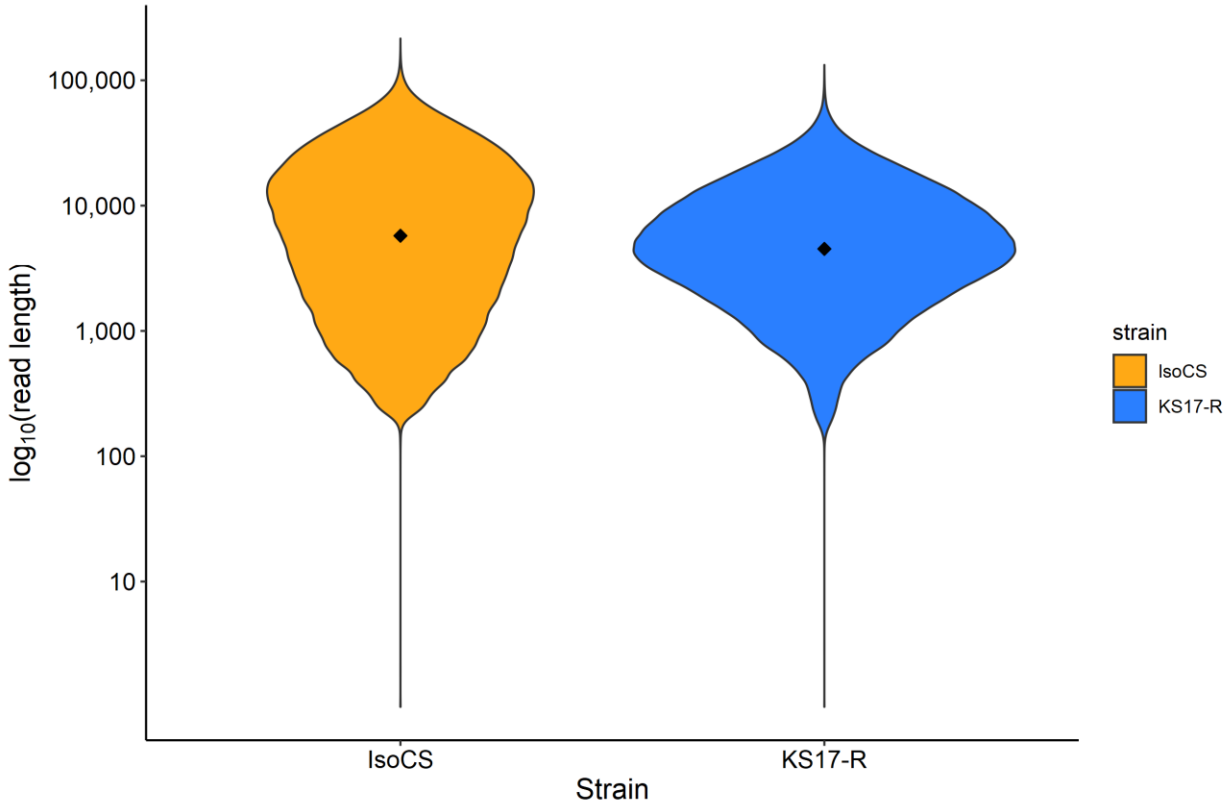


Table S1: Validation of strain resistance phenotypes. All strains used have the expected resistance phenotype over the course of all experiments. When LD_{50} values were determined, the slope estimate and standard error of the slope are included, along with the 95% CI for the LD_{50} estimate. As LD_{50} values could not be determined for the KS17-R strain, the mortality observed at 400 µg (the highest dose tested in all assays) is provided to give a better sense of the variation observed. Generation information post selection is included for the KS17-R strain.

Strain	Experiment	LD_{50} (µg/fly)	95% CI	n	Mean mass (mg)	Slope	SE Slope	Mortality (%) at 400 µg/fly	n at 400 µg/fly
KS17S2G4	BSA start	>800	-	640	17.4	-	-	0.07	100
KS17S2G8	BSA start	>800	-	700	18.8	-	-	0.18	140
KS17S3G10	synergist bioassays	>800	-	640	18.9	-	-	0.1	140
KS17S3G33	RNA-seq	>400†	-	480	15.2	-	-	0.15	100
IsoCS	BSA start	0.008	0.007-0.008	700	14.4	3.93	0.26	-	-
IsoCS	synergist bioassays	0.0094	0.009-0.986	660	12.2	4.32	0.18	-	-
IsoCS	RNA-seq	0.0107	0.0094-0.012	760	13.6	3.2	0.26	-	-
aabys	linkage	0.004	0.003-0.004	700	10.5	5.14	0.75	-	-
aabys	RNA-seq	0.00478	0.0047-0.0049	680	13.2	3.92	0.08	-	-
BSAS4G10	RNA-seq	27.9	18.4-42.5	460	14.4	3.2	1.06	-	-

† 400 µg/fly is the highest dose tested in this assay, this does not represent a difference in resistance in the KS17-R strain.

Table S2: Selections of the KS17-R strain.

Selection	Generation	Sex	Dose (µg/fly)	Number of flies treated	Mortality (%)
#1	KS17G4	Male	400	1,710	97.8
		Female	800	1,255	96.6
#2	KS17S1G2	Male	800	600	74.3
		Female	800	390	70.3
#3	KS17S2G10	Male	800	202	74.3
		Female	800	140	56.4

Table S3: Factorial analysis of the linkage of permethrin resistance in the KS17-R house fly strain following the method of [34]. A negative effect size indicates that the presence of the resistant parent chromosome is decreases the probability of mortality.

Autosome(s)	Effect	Mean Square	*F* Value
3	-624.24	24354.95	179.65
5	-434.72	11811.37	87.12
2	-323.21	6528.96	48.16
1	-232.53	3379.52	24.93
2+3+5	129.92	1055.00	7.78
4	-102.69	659.02	4.86
2+5	-78.44	384.55	2.84
3+5	-74.86	350.28	2.58
1+5	-60.05	225.36	1.66
1+2+3+5	58.82	216.23	1.59
4+5	-49.45	152.85	1.13
1+2+5	48.87	149.30	1.10
2+4	-43.30	117.20	0.86
1+2+4+5	37.54	88.09	0.65
1+3+4+5	37.30	86.95	0.64
3+4+5	34.96	76.37	0.56
2+3	30.31	57.41	0.42
1+2+3	29.29	53.63	0.40
2+4+5	-17.92	20.07	0.15
3+4	15.71	15.43	0.11
1+3	-14.65	13.41	0.10
1+3+5	92.02	529.26	0.10
1+2	-13.53	11.45	0.08
1+2+4	13.29	11.04	0.08
1+4	-8.44	4.45	0.03
2+3+4	4.30	1.16	0.01
2+3+3+5	1.17	0.09	0.00
1+4+5	1.33	0.11	0.00
1+3+4	0.67	0.03	0.00
1+2+3+4	0.00	0.00	0.00
1+2+3+4+5	-1.17	0.09	0.00
Error		135.57	

Table S4: RNA-seq data from NCBI SRA used for *de novo* transciptome assembly with Trinity.

SRA Identifier	Description	Number of Reads
SAMN01087791	whole adult, strain aabys (*Musca domestica*)	91,122,606
SAMN01087793	whole adult, strain ALHF2 (*Musca domestica*)	52,302,608
SAMN01087794	whole adult, strain ALHF (*Musca domestica*)	69,272,482
SAMN01087795	whole adult, strain CS (*Musca domestica*)	79,055,636
SAMN01823488	larva (*Musca domestica*)	63,603,468
SAMN01823489	whole fly, female (*Musca domestica*)	32,221,870
SAMN01823490	whole fly, male (*Musca domestica*)	26,681,158
SAMN01823491	whole fly, female (*Musca domestica*)	38,445,292
SAMN01823492	whole fly, male (*Musca domestica*)	30,525,250
SAMN02213969	embryo (*Musca domestica*)	49,931,558
SAMN02401116	larva (*Musca domestica*)	104,524,102
SAMN02427181	mixed eggs, larva, pupae, adult (*Musca domestica*)	66,049,270
SAMN03219953	eggs (*Musca domestica*)	78,501,900
SAMN03219954	larvae (*Musca domestica*)	69,111,592
SAMN03219955	adults (*Musca domestica*)	67,716,714
SAMN03342160	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	19,211,524
SAMN03342161	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	14,465,507
SAMN03342162	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	17,476,071
SAMN03342163	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	24,220,584
SAMN03342164	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	16,401,149
SAMN03342165	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	14,512,000
SAMN03342166	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	19,131,086
SAMN03342167	whole fly (*Musca domestica*, 5-7 days, pooled male and female)	15,939,031
SAMN03433663	head (*Musca domestica*, male)	12,314,852
SAMN03433664	head (*Musca domestica*, female)	19,327,087
SAMN03433665	head (*Musca domestica*, male)	21,496,438
SAMN03433666	head (*Musca domestica*, male)	15,216,505

SRA Identifier	Description	Number of Reads
SAMN03433667	head (*Musca domestica*, male)	18,814,975
SAMN03433669	head (*Musca domestica*, male)	17,673,926
SAMN03433670	ovary (*Musca domestica*, female)	14,842,298
SAMN03433671	testis (*Musca domestica*, male)	13,696,932
SAMN03433672	testis (*Musca domestica*, male)	17,716,731
SAMN03433673	ovary (*Musca domestica*, female)	15,140,306
SAMN03433674	testis (*Musca domestica*, male)	16,341,750
SAMN03433675	testis (*Musca domestica*, male)	16,134,200
SAMN03433676	head (*Musca domestica*, male)	18,403,310
SAMN03433677	head (*Musca domestica*, female)	16,167,867
SAMN03433678	testis (*Musca domestica*, male)	17,264,995
SAMN03433679	testis (*Musca domestica*, male)	13,304,623
SAMN03438122	whole (*Musca domestica*, pooled male and female)	52,033,902
SAMN05509747	Dorsal closure rep1 (*Musca domestica*)	230,876,120
SAMN05509748	Germ band extension rep2 (*Musca domestica*)	178,446,658
SAMN05509749	Germ band extension rep1 (*Musca domestica*)	209,607,682
SAMN05509750	Gastrula rep2 (*Musca domestica*)	150,358,362
SAMN05509751	Gastrula rep1 (*Musca domestica*)	209,747,658
SAMN05509752	Cellular blastoderm rep2 (*Musca domestica*)	151,106,316
SAMN05509753	Cellular blastoderm rep1 (*Musca domestica*)	212,293,508
SAMN05509754	Synsytial blastoderm rep2 (*Musca domestica*)	125,496,978
SAMN05509755	Synsytial blastoderm rep1 (*Musca domestica*)	51,395,004
SAMN05509756	Dorsal closure rep2 (*Musca domestica*)	153,197,776
SAMN05905666	whole fly, sterile wound (*Musca domestica*, 4 days, female)	57,241,274
SAMN05905667	whole fly, bacterial infection (*Musca domestica*, 4 days, female)	49,826,137
SAMN06198094	biological replicate1_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	37,041,692
SAMN06198095	biological replicate1_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	36,623,766

SRA Identifier	Description	Number of Reads
SAMN06198096	biological replicate2_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	35,704,182
SAMN06198097	biological replicate2_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	35,317,360
SAMN06198098	biological replicate3_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	33,486,194
SAMN06198099	biological replicate3_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	33,162,546
SAMN06198100	biological replicate1_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	36,219,036
SAMN06198101	biological replicate1_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	35,789,416
SAMN06198102	biological replicate2_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	31,361,216
SAMN06198103	biological replicate2_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	31,042,294
SAMN06198104	biological replicate3_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	34,047,078
SAMN06198105	biological replicate3_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	33,708,260
SAMN06198106	biological replicate1_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	33,645,414
SAMN06198107	biological replicate1_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	33,282,872
SAMN06198108	biological replicate2_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	32,164,782
SAMN06198109	biological replicate2_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	31,798,106
SAMN06198110	biological replicate3_Whole body_R1 read (*Musca domestica*, 5-7 day old, female)	35,627,544
SAMN06198111	biological replicate3_Whole body_R2 read (*Musca domestica*, 5-7 day old, female)	35,236,278

Table S5: Selection of the BSAS4 line used in the RNA-seq experiment. The first two selections are those for the BSA experiment (shown in Figure 1). The BSAS2 line was started from a subset mixing flies from all three of the BSA resistant bulks (~200 females).

Selection	Generation	Sex	Dose (μg/fly)	Number of flies treated	Mortality (%)
#1	BSAF6	Male	0.39	3,792	84.3
		Female	0.78	4,046	84.8
#2	BSAS1	Female	100	11,688	85.2
#3	BSAS2G23	Male	5	1,605	85.3
		Female	10	1,842	85.8
#4	BSAS3G2	Male	75	3020	56.4
		Female	100	2,175	93

Table S3 Percent of mapped reads (strand specific) that are assigned to the original NCBI annotation (+0 bp), as well as two versions of the annotation with a buffer region (+200 bp and + 400 bp) added to the 3' end of transcripts intended to compensate for incomplete gene annotation models.

Annotation	Exonic	Intronic	Intergenic
+0 bp	69.1%	8.2%	22.7%
+200 bp	73.5%	7.8%	18.7%
+400 bp	74.5%	7.7%	17.8%