The quantitative genetics of insecticide resistance in *Drosophila melanogaster*

Paul Battlay

https://orcid.org/0000-0001-6050-1868

Submitted in total fulfillment of the requirements of the degree of Doctor of Philosophy

October 2019

School of Biosciences

The University of Melbourne

For Pete.

Abstract

While understanding insecticide resistance in *Drosophila melanogaster* is informative for controlling pest insects that threaten agricultural yields and vector deadly diseases, it also serves as a powerful model of microevolution which can be interrogated with an exceptionally powerful genetic toolkit. The *Drosophila* Genetic Reference Panel (DGRP) provides the opportunity to study population-genetic signatures of natural selection in individuals that can be repeatedly measured for a range of phenotypes. In this work, genomic and transcriptomic data from the DGRP are compared with phenotypes from nine insecticidal compounds against the background of genome-wide signals of selection. The two most prominent signatures of selection in the population are attributable to insecticides from a single, widely-used chemical class, the organophosphates. Evidence suggests that insecticide-based selection is limited to these two loci, however the genetic bases of insecticide phenotypes appear to be complex. Insecticide-associated variation includes both structural effects through amino acid substitution and chimeric gene formation, and regulatory effects on transcript abundance by *cis*- and *trans*-acting factors. Resistance mechanisms exhibiting pleiotropic effects on insecticides from different chemical classes is found to be rare; one such case is correlated with constitutive, modular regulation of oxidative stress-related transcripts, the genetic basis of which is mapped to multiple *trans*-acting factors. Comparisons of the results from the DGRP with diverse population genomics data suggests that the outcomes of these analyses are applicable to populations of *D. melanogaster* worldwide.

Declaration

This is to certify that:

i. this thesis comprises only my original work towards the PhD except where indicated in the preface;

ii. due acknowledgement has been made in the text to all other material used;

iii. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Paul Battlay

December 2019

Preface

In the course of completing my PhD, I have contributed to seven journal articles. Five of these, "Genomic and transcriptomic associations identify a new insecticide resistance phenotype for the selective sweep at the *Cyp6g1* locus of *Drosophila melanogaster*" (Battlay *et al.* 2016, *G3: Genes, Genomes, Genetics*, *6*[8]), "Insights into DDT resistance from the *Drosophila melanogaster* genetic reference panel" (Schmidt *et al.* 2017, *Genetics*, *207*[3]), "Multiple P450s and variation in neuronal genes underpins the response to the insecticide imidacloprid in a population of *Drosophila melanogaster*" (Denecke *et al.* 2017, *Scientific reports*, *7*[1]), "Structural variants and selective sweep foci contribute to insecticide resistance in the Drosophila genetic reference panel" (Battlay *et al.* 2018, *G3: Genes, Genomes, Genetics*, *8*[11]), and "*Cis*-and *trans*-acting variants contribute to survivorship in a naïve *Drosophila melanogaster* population exposed to ryanoid insecticides" (Green *et al.* 2019, *Proceedings of the National Academy of Sciences*, *116*[21]), form the basis of the research chapters in this thesis. I have attempted to limit the results described in this thesis to my own contributions to these projects, and to acknowledge others' work from these projects through citation of the articles. Beyond this, major contributions to this work are as follows:

Charles Robin, Philip Batterham, and Alex Fournier-Level provided advice on experimental design and analyses. Pontus Leblanc and I collected the malathion and permethrin DGRP data as a team in which we contributed equally; Pontus submitted a separate analysis of this data as part of his MSc thesis in 2017. Philip Batterham, Shane Denecke, Francis Jiggins and Chuan Cao provided fly stocks for transgenic experiments. Work with Robert Good published in Good *et al.* (2014, *Genome biology and evolution*, *6*[5]) provided the basis for much of the structural variation annotation described in this work. Joshua Schmidt provided *i*HS and nS_L data for the DGRP, and he, along with Nandita Garud, advised on the use and interpretation of H12 scans. Alex Fournier-Level assisted with scripts for quantitative genetics analyses. Charles Robin, Joshua Schmidt, Llewellyn Green, Alex Fournier-Level and Shane Denecke contributed to the writing of the five manuscripts which included work in this thesis.

Acknowledgements

Throughout this PhD I have never wanted for inspiration, and this entirely attributable to my primary supervisor, Charles Robin. Of the innumerable ways in which Charlie contributed to this work, his infectious enthusiasm for science has been the most valuable. Thanks Charlie.

The Robin Lab has constantly been populated with brilliant and interesting people, and it has been a pleasure to work with them. In particular, I would like to thank Rob Good and Josh Schmidt, who were patient and generous with their expertise from when I joined the group as a MSc student, and have continued to make time to support me long after leaving the lab.

I carried out the work described in this thesis as part of a cohort of students studying insecticide resistance traits using the Drosophila Genetic Reference Panel; I thank Josh Schmidt, Llewellyn Green, Shane Denecke and Pontus Leblanc for sharing their data and insights with me. This work would not have been possible without their collaboration.

My committee members Phil Batterham, Owain Edwards and Alex Fournier-Level have been a great source of advice and guidance throughout this project. But additionally, they have worked to provide me with opportunities beyond my PhD for which I am truly grateful.

My girlfriend, Arianne, has provided me with unwavering support since the beginning of my PhD; as a motivator, an advisor and a refuge. I could not have done this without you, darling.

Table of Contents

1. General introduction **1**

1.1. Resistance to insecticides in insect populations 2

1.2. Insecticide resistance as a microevolutionary paradigm 5

1.3. This thesis 8

2. Genome-wide associations studies of insecticide resistance **9**

2.1. Introduction 10

2.1.1. Genome-wide association studies 10

2.1.2. The Drosophila Genetic Reference Panel 11

2.1.3. Resistance to organophosphate and pyrethroid insecticides 13

2.2. Results 15

2.2.1. Drosophila Genetic Reference Panel genome-wide association studies 15

2.2.2. Transgenic validation of genome-wide association study candidates 29

2.3. Discussion 34

2.3.1. Variation in associations among nested phenotypes 34

2.3.2. Organophosphate associations and selective sweeps 36

2.3.3. Permethrin phenotype associations 39

2.3.4. Conclusions 40

2.4. Materials and methods 41

2.4.1. Fly lines 41

2.4.2. Phenotype selection 41

2.4.3. Insect Bioassays 42

2.4.4. Calculation of LC_{50} 42

2.4.5. Estimation of broad-sense heritability and sex effect 43

2.4.6. Genome-wide association studies 43

2.4.7. *in silico* genotyping of structural variation 43

3. Transcriptome-wide association studies of insecticide resistance **44**

3.1. Introduction 45

3.1.1. Insecticide resistance and the transcriptome 45

3.1.2. Quantitative transcriptomics 47

3.2. Results 50

3.2.1. Transcriptome to phenotype associations 50

3.2.2. Mapping eQTL of male GCTM 79 64

3.2.3. Validation of *Cyp12d1* in chlorantraniliprole and malathion resistance 66

3.3. Discussion 67

3.3.1. The quantitative transcriptomics of insecticide phenotypes 67

3.3.2. Modular transcription linked to cncC regulation 69

3.3.3. Structural-variant eQTL 72

3.3.4. Conclusions 76

3.4. Materials and methods 77

3.4.1. Fly lines 77

3.4.2. Phenotype data 77
3.4.3. Transcriptome data 77
3.4.4. Transcriptome-wide association studies 77
3.4.5. Mapping eQTL of male GCTM 79 78
3.4.6. Structural variation in TWAS candidates 78
3.4.7. Transgenic overexpression of *Cyp12d1* 79

4. A meta-analysis of insecticide resistance in the Drosophila Genetic Reference Panel 80
4.1. Introduction 81
4.1.1. The DGRP as a population genomics resource 81
4.1.2. Signatures of selection in the DGRP 83
4.1.3. Phenomics 84
4.1.4. Beyond the DGRP 85
4.2. Results 86
4.2.1. A comparison of DGRP insecticide phenotypes 86
4.2.2. Population genomics of insecticide resistance 92
4.3. Discussion 99
4.3.1. The quantitative nature of phenotypes from multiple insecticides 99
4.3.2. Between-phenotype correlations 101
4.3.3. Evidence for insecticide-driven selection in the DGRP 102
4.3.4. Inferences from populations outside the DGRP 118
4.3.5. Conclusions 120
4.4. Materials and methods 121
4.4.1. Genome-wide association studies 121
4.4.2. Associations with multiallelic variants 121
4.4.3. DGRP phenome assembly 121
4.4.4. Selective sweep statistics 121
4.4.5. Analysis of *91-R* and *91-C* genomes 122
4.4.6. Worldwide population data 122

5. General discussion 123
5.1. Geometric adaptation and selective sweeps 124
5.2. *Acetylcholinesterase* 126
5.3. *Cyp6g1* 128
5.4. *Cyp6a17* and *Cyp6a23* 132
5.5. *Cyp6w1* 134
5.6. *CHKov1* 135
5.7. Conclusions 137

References 139

Appendices 158

Figures

Figure 2.2.1. DGRP genomic variant associations with single-concentration azinphos-methyl phenotypes 17
Figure 2.2.2. DGRP genomic variant associations with the azinphos-methyl LC_{50} phenotype 18
Figure 2.2.3. DGRP genomic variant associations with male malathion phenotypes 19
Figure 2.2.4. DGRP genomic variant associations with female malathion phenotypes 20
Figure 2.2.5. DGRP genomic variant associations with permethrin phenotypes 22
Figure 2.2.6. DGRP structural variation involving *Cyp6g1* and *Cyp6g2* 24
Figure 2.2.7. Amino acid substitutions conferring insecticide resistance in *D. melanogaster Ace* 25
Figure 2.2.8. Influence of *Ace* substitutions on malathion phenotypes 26
Figure 2.2.9. DGRP structural variation involving *Cyp6a17* and *Cyp6a23* 27
Figure 2.2.10. Influence of structural variation at *Cyp6a17* and *Cyp6a23* on permethrin phenotypes 28
Figure 2.2.11. Functional validation of *Cyp6g1* in organophosphate resistance 30
Figure 2.2.12. Functional validation of *Cyp6a17* in permethrin resistance 31
Figure 2.2.13. Influence of *Doc* insertion into *CHKov1* on azinphos-methyl phenotypes 32
Figure 2.2.14. Toxicology of *CHKov1* mutants 33
Figure 2.3.1. The relationship between phenotypic distribution and strength of association 35
Figure 2.3.2. Ace inhibitory constants (data from Menozzi *et al.* 2004) 37
Figure 3.1.1. Association study approaches utilizing genomic, transcriptomic and phenotypic data 48
Figure 3.2.1. Examples of TWAS associations 51
Figure 3.2.2. Relationships between insecticide association study candidates and eQTL 55
Figure 3.2.3. Repeat units involving *Cyp6g1* and *Cyp6g2* are *cis*-eQTL of both transcripts 58
Figure 3.2.4. *Cyp6g1* structural variation associations with organophosphate phenotypes 60
Figure 3.2.5. *Cyp12d1* structural variation association with chlorantraniliprole LC_{50} 60
Figure 3.2.6. *Cyp6a17* and *Cyp6a23* structural variation associations with pyrethroid phenotypes 61
Figure 3.2.7. Intersection of transcripts associated with chlorantraniliprole and malathion phenotypes, and cncC-regulated transcripts identified by Misra *et al.* (2011) 62
Figure 3.2.8. DGRP transcription module associations with two insecticide phenotypes 63
Figure 3.2.9. A phenotype for mapping eQTL of modular cncC-regulated transcripts 64
Figure 3.2.10. DGRP genomic variant associations with PC1 male GCTM 79 | cncC-regulated transcripts 65
Figure 3.2.11. Functional validation of *Cyp12d1* 66
Figure 3.3.1. Oxidative stress response regulation by cncC and Keap1 (after Wilding 2018) 69

Figure 3.3.2. Influence of structural variation at *Cyp6a17* and *Cyp6a23* on deltamethrin and permethrin phenotypes 75
Figure 4.1.1. Models of haplotype structure in selective sweeps 82
Figure 4.2.1. DGRP insecticide phenotype line sharing 86
Figure 4.2.2. Insecticide phenotype correlations 89
Figure 4.2.3. Presence of insecticide-associated DGRP variants in 15 DGN populations 96
Figure 4.2.4. Allele frequencies of top DGRP insecticide candidates in 15 DGN populations 97
Figure 4.2.5. H12 across five populations 98
Figure 4.3.1. Ace inhibitory constants (data from Menozzi *et al.* 2004) 103
Figure 4.3.2. Organophosphate associations with *Ace* and *timeout* 104
Figure 4.3.3. The *Ace* selective sweep 105
Figure 4.3.4. Resistance allele structure at the *Ace* selective sweep 106
Figure 4.3.5. The *Cyp6g1* selective sweep 107
Figure 4.3.6. Resistance allele structure at the *Cyp6g1* selective sweep 108
Figure 4.3.7. Comparison of pyrethroid phenotypes and *Cyp6a17/Cyp6a23* alleles 111
Figure 4.3.8. *Cyp6a17* triallele associations in context 112
Figure 4.3.9. Linkage disequilibrium with *Cyp6a17/Cyp6a23* alleles 113
Figure 4.3.10. Evidence for selection at *Cyp6w1* 114
Figure 4.3.11. Resistance allele structure at the selective sweep including *Cyp6w1* 115
Figure 4.3.12. Lack of evidence for selection at other insecticide candidates 116
Figure 4.3.13. Lack of evidence for selection at other insecticide candidates 117
Figure 5.1.1. Adaptive walk models of haplotype structure in selective sweeps 125
Figure 5.2.1. Duplications involving *Ace* 127
Figure 5.3.1. Structure of derived *Cyp6g1* alleles in the DSPR (after Chakraborty *et al.* 2019) 129
Figure 5.6.1. Structurally variant alleles of *CHKov1* (after Magwire *et al.* 2011) 136

Tables

Table 2.2.1. Insecticide phenotypes 16
Table 3.2.1. Details and sources of insecticide phenotypes used in this chapter 50
Table 3.2.2. Transcripts associated with insecticide phenotypes from more than one insecticide 52
Table 3.2.3. Transcripts identified by GWAS-associated eQTL and veQTL 54
Table 3.2.4. Structural variants in TWAS-identified genes and their eQTL targets 57
Table 3.2.5. Structural variants in candidate genes associated with insecticide phenotypes 59
Table 3.3.1. eQTL-transcript networks associated with insecticide phenotypes 67
Table 4.2.1. Significantly associated variants from GWAS of 35 insecticide phenotypes (results pooled for each insecticide) 87
Table 4.2.2. DGRP variants associated with phenotypes from more than one insecticide 87
Table 4.2.3. Triallelic DGRP variants associated with insecticide phenotypes 88
Table 4.2.4. Sources of DGRP phenotypes 91
Table 4.2.5. Insecticide-associated variants which overlap H12 selective sweep peaks, and show an enrichment for resistance alleles in DGRP lines carrying the two most common haplotypes 92
Table 4.2.6. Insecticide-associated variants showing evidence of selection from two or more selective sweep statistics 93
Table 4.2.7. DGRP variants associated with DDT or PC1 GCTM 79|cncC-activated transcripts, and also differentiated between *91-R* and *91-C* in the expected direction 94
Table 4.2.8. Number of SNPs associated with each insecticide, and percentage that are observed at least once in any of 15 DGN populations 95
Table 4.3.1. Transgenically validated insecticide GWAS candidates 99

1

General introduction

The defining factor in the success of humans as a species has been our ability to modify our environment. We have levelled mountains, rerouted rivers and constructed cities to suit our needs. Our understanding of physiology and medicine have increased the human lifespan while reducing the burden of debilitating and deadly diseases. But perhaps the most dramatic and important transformations have come from our mastery of plant and animal populations. Through selective breeding we have been able to bend evolutionary forces to our will, producing strains of organisms fine-tuned to the requirements of our species. However, while deliberate manipulation of natural selection has allowed us to refine beneficial organisms, this same force is unintentionally triggered in opposition to our attempted control of pests and pathogens. A principal example of this scenario is the swift evolution of insect populations to survive increasing levels of insecticides used to control them.

1.1. Resistance to insecticides in insect populations

Heritable change in the resistance of pest populations to insecticides has been rapid and repeatable. Since the introduction of synthetic insecticides in the 1940s, insecticide resistance has been reported to all major classifications of insecticidal compounds (Georghiou and Taylor 1977; Mota-Sanchez, Bills and Whalon 2002). This has led to a plateau in our ability to control vectors of deadly diseases (e.g. malaria; Hemingway *et al.* 2016), and a breakdown in the control of agricultural pests (Sparks and Nauen 2015). Combatting insecticide resistance by increasing the amount of toxin applied has also amplified collateral damage of insecticide usage to non-pest species, in environmental sentinel species such as *Drosophila melanogaster* (Wilson 2005), and notably in the western honey bee (*Apis mellifera*), where insecticide exposure has been implicated in colony collapse disorder (Steinhauer *et al.* 2018).

Insecticides are typically thought to exert their lethal effect by binding a specific protein, most commonly in the nervous system (Roush 1993; Sparks and Nauen 2015). It is in genes encoding these 'molecular targets' that mutations with the largest effect on insecticide resistance are observed. Alterations in protein structure reduce the affinity to insecticides while preserving a level of the essential endogenous function, resulting in alleles conferring huge levels of resistance. However, in these alleles enzyme function is often impeded and they therefore carry a fitness cost in the absence of selection by insecticides (ffrench-Constant *et al.* 1998; Kliot and Ghanim 2012). Targets are generally highly conserved, producing a small mutational target for insecticide resistance variation. As a result, evolution of target site resistance is often observed to be parallel across insect species, even at the level of individual amino acid substitutions (Heckel 2012).

Alleles which increase the ability of an insect to detoxify an insecticide are also associated with large effects on resistance, surpassed in magnitude only by those observed at the target site. Detoxification via metabolism of insecticides by members of extensive multigene families such as glutathione S-transferases, carboxylesterases and cytochrome P450s has been widely described as a large contributing factor to resistance

in insect species. In the cases where causative mutations have been characterized, mechanisms have included polymorphisms increasing enzyme specificity to the insecticide (e.g. Newcomb *et al.* 1997), although more commonly transcript abundance is increased through *cis*-acting local regulatory polymorphism or gene amplification events. Additionally, master-regulatory genes that control, in *trans*, detoxification enzymes have been inferred and linked to resistance in multiple arthropod species (Wilding 2018), however the genetic bases of such phenotypes are poorly understood.

Reduced penetration of insecticides through the cuticle has long been known to contribute to resistance (Sawicki and Farnham 1968). Insects exhibiting reduced-penetration phenotypes have been shown to have significantly thickened cuticles and a higher abundance cuticular hydrocarbons (Yadav, Singh and Shrivastava 2010; Strycharz *et al.* 2013; Balabanidou *et al.* 2016). Transcriptomic evidence suggests such phenotypes are due to increased expression of genes encoding either cuticular constituents (Vontas *et al.* 2007; Puinean *et al.* 2010; Gregory *et al.* 2011; Qiu *et al.* 2013) or enzymes putatively involved in the biosynthesis of cuticular hydrocarbons (Toé *et al.* 2015; Balabanidou *et al.* 2016; Chen *et al.* 2020). More elusive are the genetic bases of heritable changes in pest behavior which decrease exposure to insecticides in the field. Such mechanisms have been described in a range of pest species (Sparks *et al.* 1989), however the study of these phenotypes is impeded by the difficulty of their replication in a laboratory setting.

While these canonical models of insecticide resistance are useful in categorizing and understanding common evolutionary responses to insecticides, they are neither exhaustive nor universally precise. For example, the demarcation between target site resistance and metabolic resistance becomes challenging when considering overproduced carboxylesterases which sequester acetylcholinesterase-targeting organophosphate insecticides (Devonshire and Field 1991). The idea that insecticides exert their toxic effects through a single molecular target is also contentious: DDT (which 'targets' the voltage-gated sodium channel) is also capable of toxicity by inhibiting mitochondrial ATP-synthase in *A. mellifera* (Younis *et al.* 2002). Moreover, mounting evidence suggests that a range of insecticidal compounds, in addition to their perturbation of molecular targets, act to increase oxidative stress (reviewed in Wilding 2018).

Increased metabolism phenotypes have been the central focus of detoxification gene identification; however, it appears that multiple additional layers contribute to effective detoxification of insecticides. The insecticide metabolites of resistance-associated detoxification enzymes may be no less toxic than their precursor (Fusetto *et al.* 2017), instead being more amenable to transport and excretion, or themselves substrates of additional metabolic processes. Transport of insecticides or their metabolites has also been increasingly shown to contribute to resistance, with variation in the ABC family of transporters emerging as an important contributor to insecticide resistance (Dermauw and Leeuwen 2014).

Understanding the mechanisms by which insecticide resistance can evolve is vital to the effective use of insecticides to control pests. While resistance may be inevitable, it can be deferred by insecticide rotation (Coyne 1951), insecticide mixing (Comins 1986) and mosaic application designs (Rousch 1989). These strategies, however, share the requirement that the compounds used must impose selection at different loci (i.e. 'cross resistance' between compounds must be avoided). Cross resistance at target site alleles is generally well understood, as insecticides within the same class share a molecular target and hence resistance alleles (e.g. Menozzi *et al.* 2004). However, cross resistance associated with other insecticide resistance mechanisms is far less predictable (e.g. Daborn *et al.* 2007). Negative cross-resistance between insecticides, although rare, has also been described, for example between spinosad and neonicotinoid-class insecticides (Perry *et al.* 2012; Azhar Ali Khan, Akram and Shad 2014) that appear to target different subunits of the same receptor (Perry *et al.* 2012).

Insights into the genetic bases of insecticide resistance have extended the utility of insecticidal compounds and increased the efficiency of insect control (Sparks and Nauen 2015). However, the bulk of insecticide resistance studies have focused on the characterization of genetic variants of large effect, and thus the contribution of quantitative factors to variation in, and evolution of, insecticide resistance in field populations remains relatively unknown.

1.2. Insecticide resistance as a microevolutionary paradigm

In addition to applied uses for the study of insecticide resistance, the adaptation of pests in response to insecticide-based control strategies is also of significant use as a model of microevolution, as selective agents are known, and their usage is well-documented temporally as well as spatially. Moreover, studies of insecticide resistance have yielded a result which challenges our understanding of evolutionary processes: that the adaptive response to insecticides involves alleles of extremely large effect.

For more than a century biologists have sought to unite Charles Darwin's (1859) theory of evolution by natural selection with Gregor Mendel's (1866) laws of inheritance into a single mathematical framework, which over time has come to be known as the 'Modern Synthesis'. Early attempts to combine natural selection and Mendelian genetics were confounded by the discrepancy between the continuous variation observed in phenotypes at the population level and the discrete units of inheritance described by Mendel. Initially thought to represent two different sources of variation, fierce debate ensued as to whether evolution acted upon continuous variation or discrete Mendelian traits. This debate was largely resolved in the 1920s and 30s by founding population geneticists R. A. Fisher, J. B. S. Haldane and Sewall Wright, who demonstrated that the continuous variation observed in populations could be explained by a large number of alleles inherited in a Mendelian manner (e.g. Fisher 1918), and natural selection would likely favor such quantitative variation over likely deleterious variants of large effect (Fisher 1930). Due to the mathematical density of these findings by early population geneticists, they remained unappreciated by mainstream biologists until the publication of Theodosius Dobzhansky's 'Genetics and the Origin of Species' (1937), which applied population genetics theory to biological data and became the standard-bearer for the modern synthesis.

As technologies for mapping phenotypes to genetic variation have improved, the models of the early population geneticists for the quantitative architecture of traits have been increasingly supported with empirical data. For many traits in complex organisms, the number of discrete contributors seems limited only by our power to detect them; the

'omnigenic' model arising from human association study data (Boyle, Li and Pritchard 2017) is highly congruent with Fisher's (1918) infinitesimal model. The genetic architecture of most traits in higher organisms seems demonstrably complex, however the complexity of adaptation is less clear. Artificial selection in agriculture has yielded linear and continuous increases in trait values (Walsh and Lynch 2018). These patterns reject the involvement of only a small number of loci (which would quickly reach fixation, causing response to plateau) and alleles of large effect (which would result in departures from linearity in the response). Instead they suggest adaptation acting via small fluctuations in population frequencies of a large number of alleles.

A contrast to the model of weak, highly polygenic adaptation is provided by multiple studies describing selection of sufficient strength to produce molecular signatures discernable at the genomic level. These 'selective sweeps' (Smith and Haigh 1974; Kaplan, Hudson and Langley 1989), occur around alleles that have risen in frequency with such rapidity that they exhibit significantly extended linkage disequilibrium with surrounding variants. For example, such signatures are observed around variants involved in lactose tolerance (Bersaglieri *et al.* 2004) and malaria resistance (Hamblin and Di Rienzo 2000; Tishkoff *et al.* 2001) in humans. However, the overall contribution of large effect loci to evolutionary processes remains undetermined, as their current prevalence in the zeitgeist may be an artefact of the relative ease with which they can be detected (Pritchard, Pickrell and Coop 2010).

Insights into the architecture of positively selected traits provided by studies of insecticide resistance have been varied. Despite the involvement of large-effect monogenes in resistance to insecticides, polygenetic responses to insecticide-based selection in laboratory populations have been described since as early as Crow (1954). This discrepancy between laboratory and field-derived insecticide resistance has been attributed to two major factors: For monogenic resistance to arise, the size of the selected population must be large enough to include rare alleles of major effect, and the selective pressure imposed by the insecticide must be outside of the normal phenotypic range (Roush and McKenzie 1987). These theoretical predictions were subsequently demonstrated experimentally by Mackenzie *et al.* (1992): By increasing genetic diversity in a laboratory population of the Australian sheep blowfly *Lucilia cuprina* with mutagenesis, alleles at a single, large-effect locus were selected for, but only by

using insecticide doses beyond the phenotypic range. However, as technologies for elucidating the genetic bases of field resistance have improved, the number of loci implicated in field-resistant insects has begun to increase (e.g. Bass *et al.* 2011; The *Anopheles gambiae* 1000 Genomes Consortium, 2017), suggesting that predictions of field-resistance architecture have been subdued by our power to detect resistance-associated loci.

1.3. This thesis

This thesis attempts to dissect the genetic bases of insecticide resistance in the model insect, *D. melanogaster* using a quantitative genetics approach in a single population sample, the Drosophila Genetic Reference Panel (DGRP; Mackay *et al.* 2012). In the first research chapter, phenotypes for three insecticides representing putative selective agents for the population are tested for associations with genetic variation across the genomes of the mapping panel. The second research chapter leverages powerful datasets arising from the characterization of the DGRP's transcriptomes to probe insecticide associations at the transcript level. Neglected and novel approaches for incorporating transcriptome data into DGRP studies are applied to phenotypes arising from nine distinct insecticidal compounds. The third and final research chapter continues the characterization of the phenotypes from the nine insecticides, this time assessing correlations with 113 DGRP phenotypes and returning to genetic variation to quantify resistance-associated variants under strong recent selection, as well as the applicability of DGRP findings to other populations of *D. melanogaster*.

Utilizing the DGRP as both a population sample as well as a quantitative trait mapping panel, these analyses seek to address two central questions: Firstly, what genetic features underlie variation in insecticide phenotypes in a wild-derived population sample of *D. melanogaster*? Studies of insecticide resistance have tended to focus on single loci of large effect; what insights can be gained through the analysis of insecticide phenotypes using quantitative genetics approaches? Secondly, have insecticides shaped the recent evolutionary history of this non-pest species? The DGRP offers an unparalleled opportunity to assess genome-wide signatures of selection and multiple phenotypes in a single set of clonal lines. With this increased power, can specific alleles and selective agents be implicated?

2

Genome-wide associations studies of insecticide resistance

The genetic basis of insecticide resistance has classically been limited to the characterization of genes of large effect. A greater resolution for mapping genetic variation at a range of effect sizes can be achieved by utilizing quantitative genetics resources such as the Drosophila Genetic Reference Panel (DGRP). Understanding the genetic architecture of insecticide resistance in the DGRP is of particular interest as evidence suggests that insecticides have imposed strong selective pressures on the progenitor population of this mapping panel, and while hundreds of phenotypes have already been mapped in the DGRP, correspondence between associated variants and signatures of selection have been extremely rare. Three insecticidal compounds were selected for study utilizing the DGRP, representing two of the most commonly used classes of insecticides, the organophosphates and the pyrethroids. This chapter describes genome-wide association studies (GWAS) of mortality-related phenotypes from these three insecticides, allowing a direct comparison between putatively selective phenotypes and molecular signatures of selection in a single population sample.

2.1. Introduction

2.1.1. Genome-wide association studies

Since the demonstration by Fisher (1918) that the normal distribution of phenotypes observed in populations can be explained by the contribution of multiple alleles, a major goal of genetics has been to understand the way in which heritable variation manifests in these 'quantitative traits'. By the mid-2000's, high-throughput genotyping methods had progressed to the point that it was feasible to gather population-level data for genetic variants across the genome, and calculate the association of each variant with the value of a quantitative trait measured in the population. These 'genome-wide association studies' (GWAS) were embraced in human genetics where many other genetic mapping approaches have limited utility, and have since been widely applied in attempts to establish the genetic basis of human phenotypes, in particular, disease risk.

Although broadly successful in identifying variants contributing to phenotypes, the results of early GWAS presented researchers with two main surprises: the vast majority of associated variants were noncoding, and together associated variants were only able to explain a fraction of the heritability of most traits (Manolio *et al.* 2009). These findings suggested that, in contrast to monogenic Mendelian traits, most phenotypes are the result of multiple small-effect variants, with the corollary that increased sample size would identify further loci. As GWAS were able to grow in resolution (through further development in genotyping capabilities) and power (through the accumulation of population datasets), this prediction was realized (Visscher *et al.* 2017). For example, heritability estimates in schizophrenia, one of the most well-studied human phenotypes, suggest that between 71 to 100 percent of one-kilobase windows in the genome contain variants contributing to the trait (Loh *et al.* 2015). Such findings led to the development of the 'omnigenic' model (Boyle, Li and Pritchard 2017), which posits that regulatory networks are sufficiently interconnected such that most functional variants in the genome which affect genes expressed in a relevant cell type will have non-zero effects on a given phenotype, and therefore GWAS' missing heritability lies spread among an enormous number of genetic variants below statistical significance.

2.1.2. The Drosophila Genetic Reference Panel

In 2012, the *Drosophila melanogaster* community was introduced to the Drosophila Genetic Reference Panel (DGRP), a resource comprised of inbred lines sampled from a single North American population of *D. melanogaster* in 2003, designed for quantitative genetic analysis of the model organism through GWAS (Mackay *et al.* 2012). The DGRP consists of 205 inbred lines (in freeze 2; Huang *et al.* 2014) resequenced at high coverage with Illumina paired end reads, which enabled the identification of most simple variants (single nucleotide polymorphisms and small insertion/deletion variants) across the population (4,565,215 in freeze 2; Huang *et al.* 2014).

Due to the design of the DGRP, GWAS in this mapping panel have many benefits over those performed in humans. The permanence of the effectively clonal lines coupled with the rapid developmental time of *D. melanogaster* allows repeated phenotyping of each genotype, facilitating more accurate estimates of phenotypic values. Prevalent homozygosity within lines also facilitates mapping by producing more extreme phenotypes than those observed in largely heterozygous outbred populations (Mackay *et al.* 2012). Moreover, flies can be raised under controlled conditions, reducing the contribution of environmental factors to phenotypes. DGRP GWAS also benefit from *D. melanogaster*'s experimental tractability and the rich genetic toolkit available for the organism; candidate loci can be validated transgenically in a controlled genetic background using existing RNAi and mutagenesis libraries (Dietzl *et al.* 2007; Bellen *et al.* 2011), and CRISPR-meditated introduction of candidate alleles (e.g. Highfill *et al.* 2017).

Aside from the advantages of working with a model organism, the mapping panel design offers additional advantages over human GWAS: Repeated use of the resource incentivizes investment in its further characterization, for example in the refinement of genome sequences (Zichner *et al.* 2013; Huang *et al.* 2014) and addition of transcriptome data (Huang *et al.* 2015). Furthermore, phenotypes measured on a panel can be directly compared between one another and correlations can be assessed, implicating links in genetic determinants between traits. To date, over 50 studies (including hundreds of nested phenotypes) have been phenotypically characterized in DGRP lines (reviewed in Anholt and Mackay 2018), and these span categories such as

physiology, morphology, stress tolerance, life history traits, molecular processes, and even the genetic basis of phenotypic variability and plasticity (Ayroles *et al.* 2015).

Although the full-genome sequencing approach employed by the DGRP does not discriminate against rare alleles (in contrast to SNP arrays used until recently by the majority of human studies), GWAS in the DGRP are still limited in their power to detect associations with rare variants due to the population's relatively small sample size; the resolution required to identify candidates is lost at minor allele frequencies below approximately five percent (Huang *et al.* 2014). Moreover, DGRP GWAS must still rely on linkage disequilibrium with simple variants as markers for complex structural variation such as large gene duplications and transposable element insertions, which are often not called in DGRP line genotypes. The relatively low level of linkage disequilibrium in *D. melanogaster* (decaying over a few hundred base pairs on average [Mackay *et al.* 2012]) is a double-edged sword; while it allows more accurate identification of associations with genotyped variants, it may obscure the inference of those variants excluded from genotyping.

In contrast to other mapping panels (Bennett *et al.* 2010; Long, Macdonald and King 2014), the DGRP represents a frozen sample of natural genetic variation from a single population, in both segregating alleles and their frequencies (albeit inbred to almost complete homozygosity). This allows for inferences to be drawn from patterns of variation in the panel to forces shaping its wild progenitors (Robin, Battlay and Fournier-Level 2019). In 2015, Garud *et al.* identified regions of the *D. melanogaster* genome under strong, recent selection by interrogating the sequences of DGRP lines for signatures of selective sweeps. The top three regions identified in this screen, *Cyp6g1*, *Ace*, and *CHKov1*, had all been previously associated with resistance to insecticides (Daborn *et al.* 2001; Pralavorio and Fournier 1992; (Aminetzach, Macpherson and Petrov 2005); two of them to a particular class, the organophosphates (OPs).

2.1.3. Resistance to organophosphate and pyrethroid insecticides

Resistance to OPs is arguably the best understood of any resistance to an insecticide class. Widespread use of this class of insecticides for more than half a century on a range of pests has resulted in many well-studied cases of insecticide resistance (Siegfried and Scharf 2001). Acetylcholinesterase (Ace) is the molecular target of OPs. At the synapse, Ace hydrolyses the ester bond in acetylcholine following neurotransmission, ending the activation signal. OPs bind irreversibly to Ace, causing a build-up of acetylcholine in the synapse, and continuous stimulation of the postsynaptic neuron. This results in paralyzing seizures, and the eventual death of the insect. Four substitutions in *D. melanogaster* Ace are known to reduce sensitivity of the enzyme to OPs. These mutations occur together in some alleles, acting cooperatively to increase resistance, with differing combinations maximizing resistance to different insecticides by restricting insecticide access and affecting the position of key catalytic residues (Mutero *et al.* 1994; Menozzi *et al.* 2004).

Pyrethroids have also been extensively utilized both spatially and temporally in insect control, however, natural variation contributing to pyrethroid insecticide class resistance in *D. melanogaster* is less well understood. Analogously to Ace, insensitivity mutations in the molecular target of pyrethroids and organochlorides (a separate insecticide class which includes dichlorodiphenyltrichloroethane [DDT]), the voltage-gated sodium channel, are common in insect pest species (Dong *et al.* 2014). However, orthologous mutations have not been described as natural variants in *D. melanogaster*, although EMS mutagenesis has yielded mutations in *para* (the *D. melanogaster* voltage-gated sodium channel alpha subunit) that cause resistance to DDT and the pyrethroid deltamethrin (Pittendrigh *et al.* 1997). At least one other gene has been shown to be involved in pyrethroid biology in *D. melanogaster*. *Cyp4e3* (a member of the detoxification-associated cytochrome P450 gene family), is both induced in response to permethrin exposure, and capable of increasing resistance to the insecticide when overexpressed (Terhzaz *et al.* 2015), however, once again natural variation in this gene has not been described, and any contribution of this locus to pyrethroid resistance in wild populations is yet to be determined.

DGRP GWAS of insecticide resistance phenotypes are expected to contrast human disease studies in one key aspect: Where large-effect human disease alleles are by

definition deleterious and hence expected to be rare in populations, alleles in *D. melanogaster* conferring high levels of resistance to insecticides will range in frequency based on the interplay between pressure from the selective agent, the cost of the alleles in the absence of insecticide-driven selection, and the time over which these pressures have been maintained. Moreover, these patterns will vary from insecticide to insecticide, as it is unknown which compounds have shaped the evolutionary history of the DGRP's ancestors.

Organophosphates and pyrethroids are two of the oldest and most widely used insecticide classes in the world today. The experiments in this chapter seek to investigate the genetic basis of resistance in the DGRP to representatives of each of these classes: Malathion, the most commonly used organophosphate in the USA (Bonner *et al.* 2007), azinphos-methyl, another common organophosphate with putative links to two selective sweeps in the DGRP (Garud *et al.* 2015), and permethrin, the most commonly used pyrethroid insecticide in the world (Anadón, Martínez-Larrañaga and Martínez 2009).

2.2. Results

2.2.1. Drosophila Genetic Reference Panel genome-wide association studies

2.2.1.1. Azinphos-methyl larval phenotype

First instar survival to pupation on azinphos-methyl-laced media was measured for 178 DGRP lines at four concentrations. Broad sense heritability (H^2) of these phenotypes ranged from 0.84-0.94 (table 2.2.1). The median lethal concentration (LC_{50}) was calculated for each DGRP line using the four common concentrations and additional concentrations as required. Azinphos-methyl LC_{50}s ranged from 0.08 to 7.3μg/ml.

2.2.1.2. Malathion and permethrin adult phenotypes

Malathion and permethrin phenotypes were measured in three-to-five day old males and females separately from 170 DGRP lines in scintillation vials coated with 1 and 10μg of insecticide respectively. Malathion mortality phenotypes were measured at 3, 6, 12 and 24-hour timepoints, with H^2 ranging from 0.56-0.68 (table 2.2.1). As pyrethroid insecticides produce a field-relevant 'knockdown' phenotype, both mortality and knockdown were measured. Mortality was measured at 24 hours, and knockdown at 3 hours. H^2 of permethrin phenotypes ranged from 0.56-0.61 (table 2.2.1). Sex effects were measured for all phenotypes (table 2.2.1) and were larger for permethrin phenotypes (0.11-0.24) than malathion phenotypes (0.02-0.04).

2.2.1.3. Genome-wide association studies

Phenotypes were tested for associations with genomic variants using the DGRP2 pipeline, which corrects for the effects on phenotype of *Wolbachia pipientis* infection status and five common chromosomal inversion genotypes in each DGRP line (Huang *et al.* 2014). The magnitude of significant effects of these covariates are reported in table 2.2.1. Wolbachia infection significantly decreased mortality in response to malathion and permethrin at multiple phenotypes, but no significant correlation with Wolbachia infection was identified in any azinphos-methyl phenotype ($p<0.05$; table 2.2.1). *p*-values arising from these mixed linear models are reported as 'mixed *p*-values', and two thresholds were utilized in the assessment of significant associations: The 'genome-wide significance threshold' (1×10^{-5}), corrected for the number of

phenotypes tested for each insecticide, and the Bonferroni significance threshold (the conventional 0.05 corrected for the number of DGRP variants and phenotypes tested).

Table 2.2.1. Insecticide phenotypes

phenotype	broad-sense heritability	sex effect	significant covariate effects
azinphos-methyl 0.25µg/ml mortality	0.84	-	In_2R_NS (0.24)
azinphos-methyl 0.5µg/ml mortality	0.91	-	-
azinphos-methyl 1µg/ml mortality	0.94	-	-
azinphos-methyl 2µg/ml mortality	0.93	-	-
azinphos-methyl LC50	*	-	-
malathion male 3-hr mortality	0.56	0.02	-
malathion female 3-hr mortality	0.64	-	In_2R_NS (0.28); In_3R_P (0.34)
malathion male 6-hr mortality	0.65	0.04	Wolbachia (-0.16)
malathion female 6-hr mortality	0.68	-	Wolbachia (-0.14)
malathion male 12-hr mortality	0.64	0.04	Wolbachia (-0.16)
malathion female 12-hr mortality	0.68	-	Wolbachia (-0.15); In_3R_P (0.33)
malathion male 24-hr mortality	0.61	0.04	Wolbachia (-0.10)
malathion female 24-hr mortality	0.58	-	-
permethrin male 3-hr knockdown	0.60	0.11	Wolbachia (-0.12)
permethrin female 3-hr knockdown	0.56	-	-
permethrin male 24-hr mortality	0.61	0.24	Wolbachia (-0.15)
permethrin female 24-hr mortality	0.61	-	Wolbachia (-0.12)

*As the LC_{50} is calculated from all replicates across multiple concentration phenotypes, broad-sense heritability could not be calculated for this phenotype

2.2.1.4. Genomic variants associated with azinphos-methyl phenotypes

Across all five azinphos-methyl GWAS, 193 unique variants reached genome-wide significance (mixed $p<2\times10^{-6}$; fig. 2.2.1; fig. 2.2.2; appendix 1.1), 48 of which were also below the Bonferroni significance threshold for these GWAS (mixed $p<5.25\times10^{-9}$). The strongest association (mixed $p=1.27\times10^{-25}$) is from the 0.25µg/ml survival phenotype, and is located in an intron of *Cyp6g1*. A selective sweep containing *Cyp6g1* has previously been described in the DGRP (located at `2R:7722921-8486706`; Garud *et al.* 2015) and associations in this region account for 94 of the genome-wide significant variants, and 43 of the Bonferroni-significant variants.

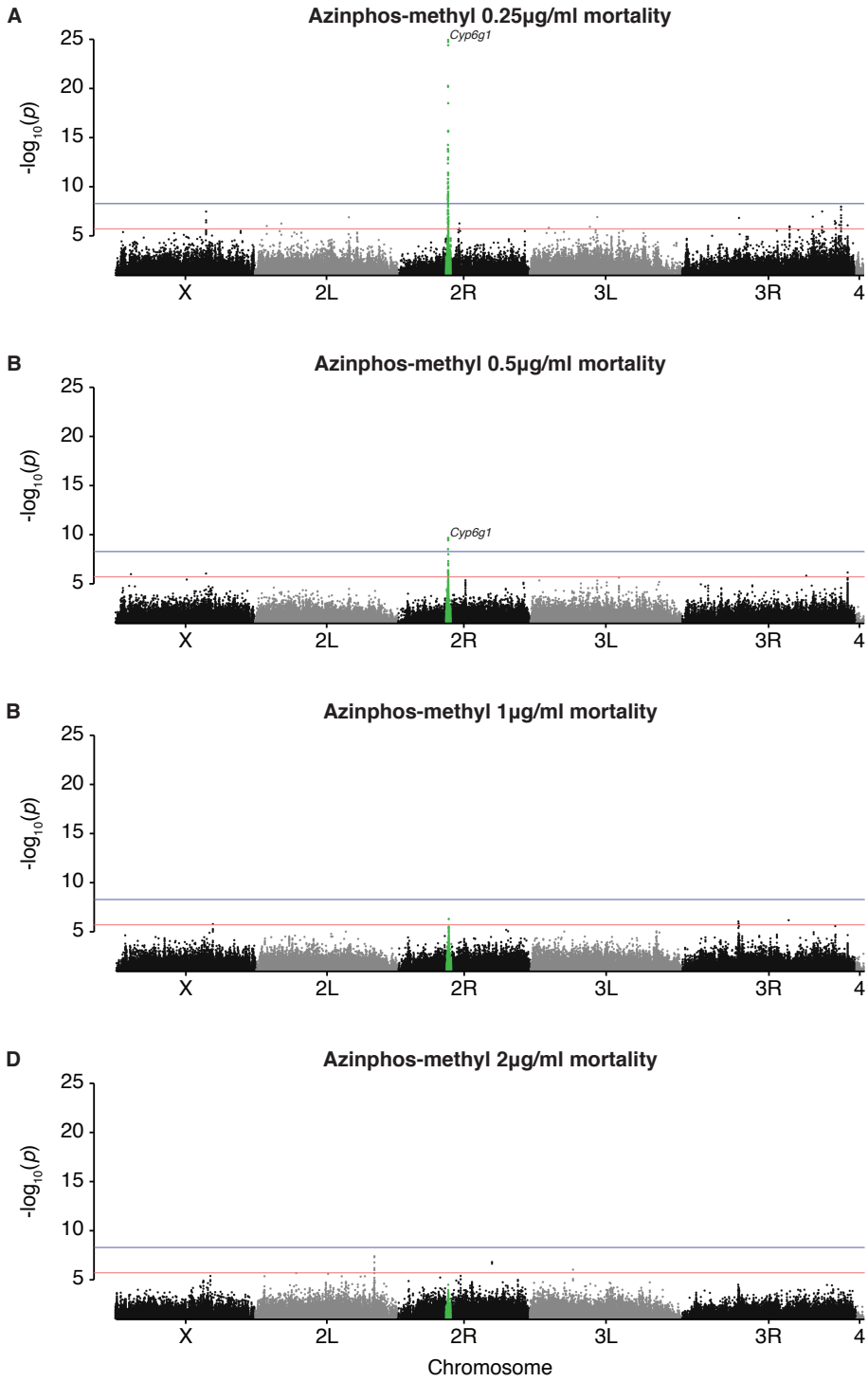


Figure 2.2.1. DGRP genomic variant associations with single-concentration azinphos-methyl phenotypes

Manhattan plots (mixed *p*-value against genomic location) for four single-concentration azinphos-methyl phenotypes (A-D). The genome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue. Green highlights show variants within the H12 selective sweep statistic peak identified around *Cyp6g1* (Garud *et al.* 2015).

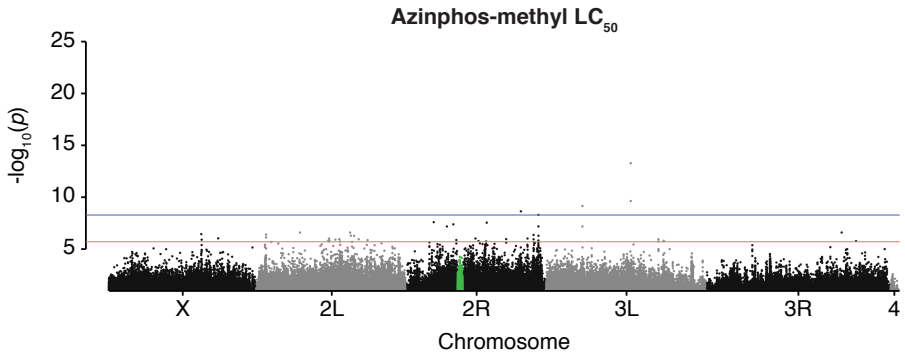


Figure 2.2.2. DGRP genomic variant associations with the azinphos-methyl LC_{50} phenotype
Manhattan plot (mixed *p*-value against genomic location) for azinphos-methyl LC_{50} phenotypes. The genome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue. Green highlights show variants within the H12 selective sweep statistic peak identified around *Cyp6g1* (Garud *et al.* 2015).

2.2.1.5. Genomic variants associated with malathion phenotypes

For the eight malathion phenotypes, 273 unique variants were identified with mixed *p*-values below the genome wide significance threshold (1.25×10^{-6}; fig 2.2.3; fig 2.2.4; appendix 1.7), more than half of which (176) were only associated with a single sex. Twelve nonsynonymous variants in eight genes were associated with malathion phenotypes, including three in *Ace,* the molecular target of organophosphate insecticides. Like *Cyp6g1*, *Ace* is also located in a selective sweep region identified by Garud *et al.* (2015; located at `3R:8563358-9432482`), and malathion associations in *Cyp6g1* and *Ace* selective sweep regions account for 18 and 128 of the genome-wide significant variants respectively. 62 malathion phenotype-associated variants achieved mixed *p*-values below the Bonferroni-corrected significance threshold (3.33×10^{-9}), of which only four were outside *Cyp6g1* and *Ace* selective sweep regions.

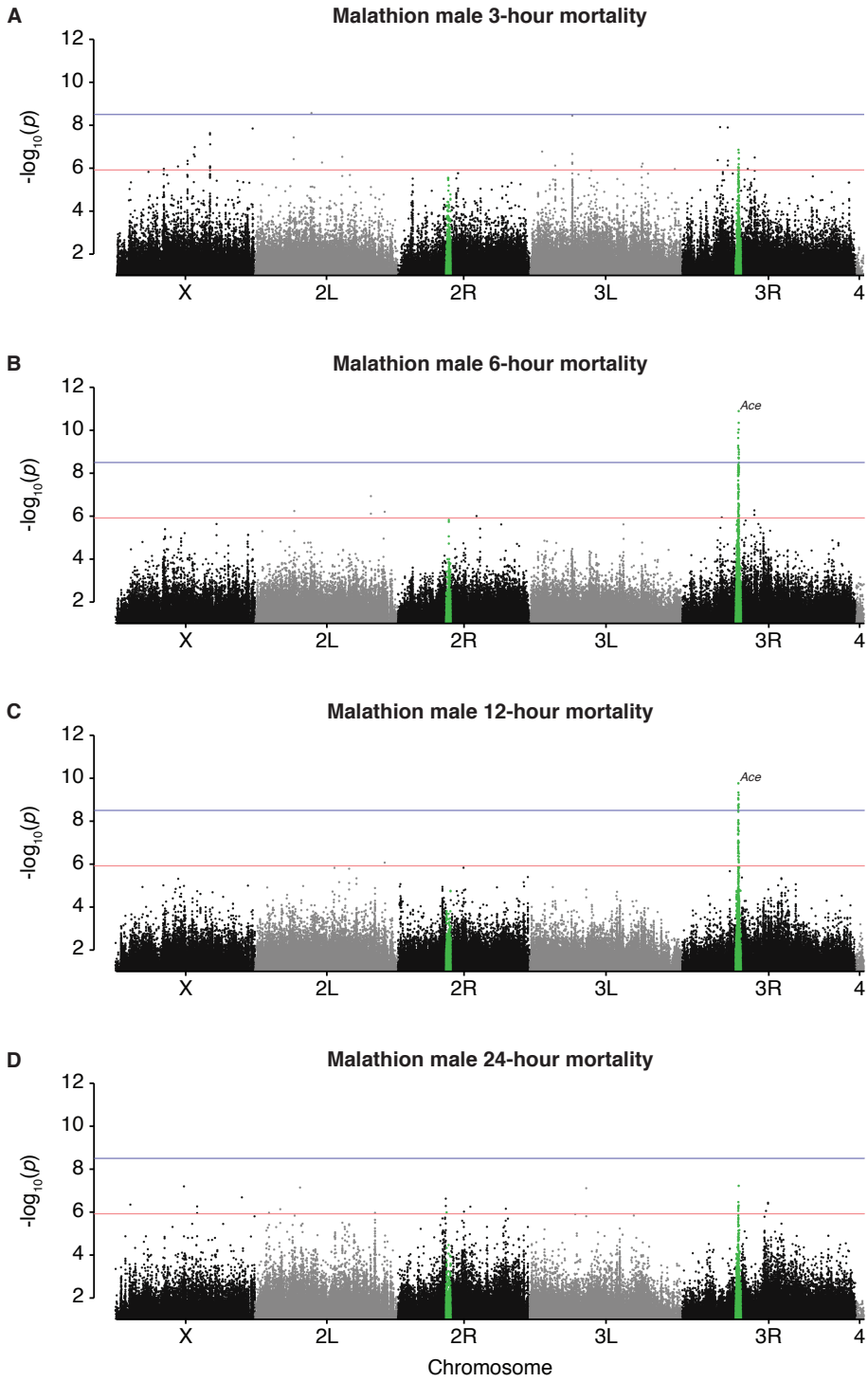


Figure 2.2.3. DGRP genomic variant associations with male malathion phenotypes

Manhattan plots (mixed *p*-value against genomic location) for male malathion phenotypes (A-D). The genome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue. Green highlights show variants within the H12 selective sweep statistic peaks identified around *Cyp6g1* and *Ace* (Garud *et al.* 2015).

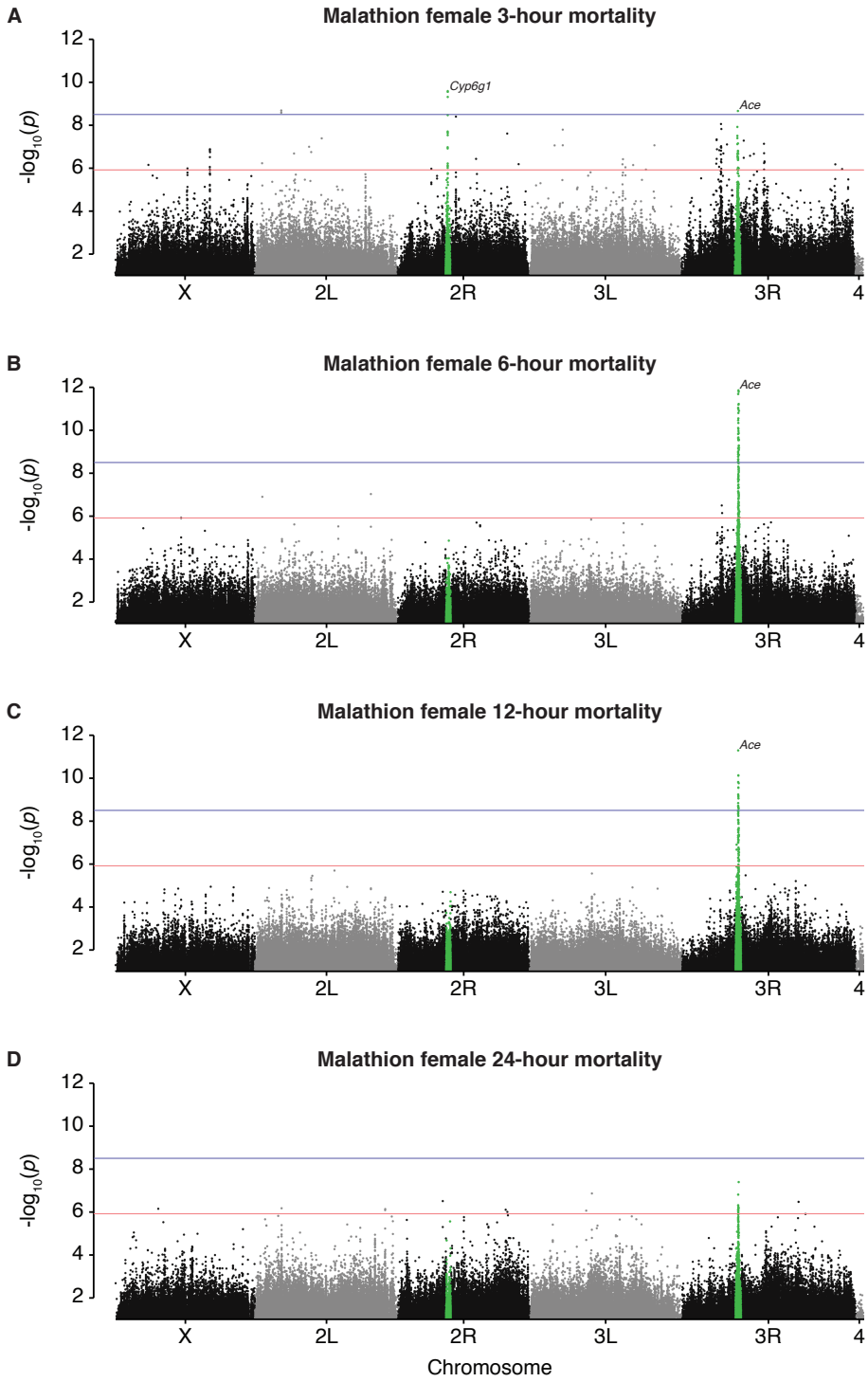


Figure 2.2.4. DGRP genomic variant associations with female malathion phenotypes

Manhattan plots (mixed *p*-value against genomic location) for female malathion phenotypes (A-D). The genome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue. Green highlights show variants within the H12 selective sweep statistic peaks identified around *Cyp6g1* and *Ace* (Garud *et al.* 2015).

2.2.1.6. Genomic variants associated with permethrin phenotypes

Among the four permethrin phenotypes, 39 unique variants were significantly associated at the genome-wide significance threshold (2.5×10^{-6}; fig. 2.2.5; appendix 1.9), two of which achieved Bonferroni-corrected significance ($p<6.66\times10^{-9}$). Eight of the nine variants common to both sexes, including the Bonferroni-significant associations, were identified in a cluster of cytochrome P450 genes on the right arm of chromosome 2.

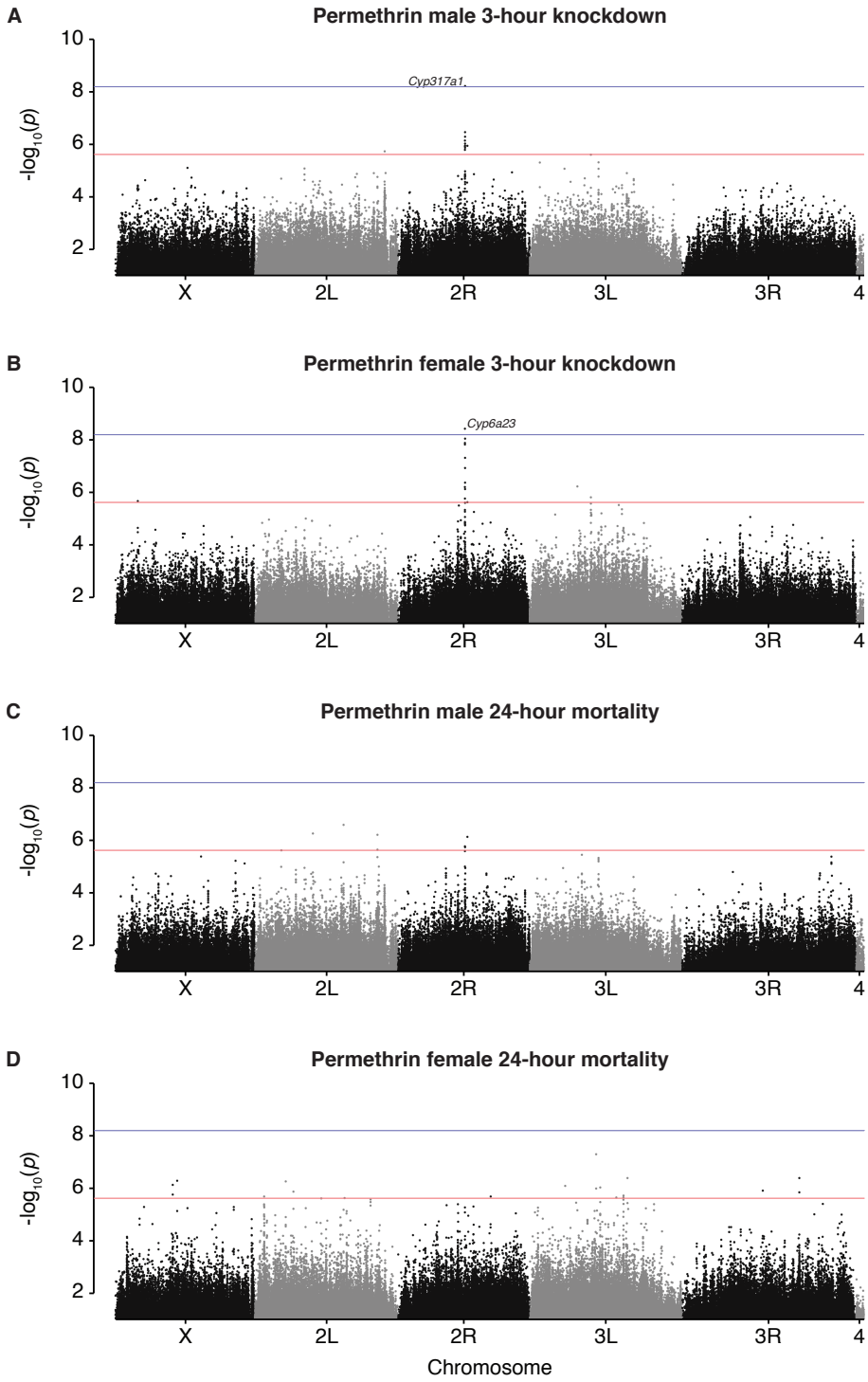


Figure 2.2.5. DGRP genomic variant associations with permethrin phenotypes

Manhattan plots (mixed *p*-value against genomic location) for four permethrin phenotypes (A-D). The genome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue.

2.2.1.7. Organophosphate resistance in the DGRP is associated with structural variation at *Cyp6g1*

Genome-wide association studies of larval azinphos-methyl phenotypes and adult malathion phenotypes identified multiple associated variants from within a selective sweep region on the right arm of chromosome 2 containing *Cyp6g1* (fig. 2.2.1A,B; fig. 2.2.4A). An allelic series of structural variants has previously been described at *Cyp6g1* (Schmidt *et al.* 2010; fig. 2.2.6). Derived alleles are characterized by a complete duplication of *Cyp6g1* coding sequence, insertion of the long terminal repeat (LTR) of an *Accord* retrotransposon upstream of the transcription start site of *Cyp6g1*, and a repeat unit containing partial copies of the coding region of *Cyp6g1* and its tandem paralog, *Cyp6g2*. Among derived *Cyp6g1* alleles, subsequent insertions of *Beagle* and *P-element* transposable elements into the *Accord* LTR differentiate $Cyp6g1^{BA}$ and $Cyp6g1^{BP}$ alleles from $Cyp6g1^{AA}$ alleles.

Inspection of 210 DGRP genome sequence alignments revealed only nine lines to be homozygous for the ancestral $Cyp6g1^{M}$ allele, with 189 lines homozygous for the duplication events indicative of derived *Cyp6g1* alleles. Moreover, read depth in the region suggests substantial additional variation in copy number of both duplications in the DGRP population. Among derived alleles, SNP markers described by Schmidt (2014) were used to impute 152 lines as $Cyp6g1^{AA}$, and 37 lines as $Cyp6g1^{BA}$ (fig. 2.2.6). There is no evidence of the $Cyp6g1^{BP}$ allele among DGRP lines, consistent with previous surveys which suggest this allele is restricted to populations in India and Australia (Catania *et al.* 2005; Schmidt *et al.* 2010).

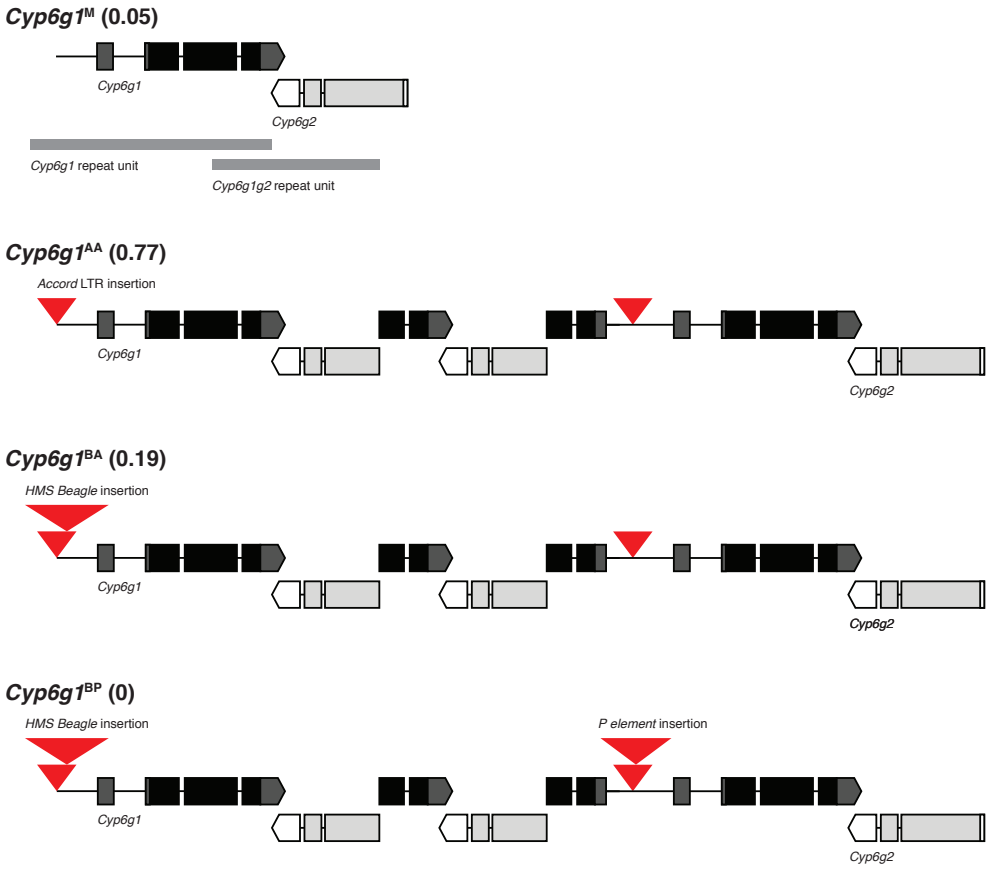


Figure 2.2.6. DGRP structural variation involving *Cyp6g1* and *Cyp6g2*

Known structurally variant alleles involving *Cyp6g1* and *Cyp6g2* (after Schmidt *et al.* 2010), and their frequencies among DGRP lines. Grey bars indicate repeat regions observed in derived alleles.

The most highly associated variants in the *Cyp6g1* sweep region, with both azinphos-methyl and malathion phenotypes, map around the 5' end of *Cyp6g1* and have a minor allele frequency of between eight and ten lines. One of these variants, a 509bp insertion polymorphism at `2R:8072884`, is the *Accord* LTR insertion. The minor allele at this variant correctly genotypes the nine DGRP lines which are homozygous for the ancestral *Cyp6g1*^{M} arrangement at this locus, which lacks any structural variation. Presence of the *Accord* LTR insertion is associated with increased resistance to both organophosphates, suggesting derived alleles of *Cyp6g1* confer organophosphate resistance in the DGRP.

2.2.1.8. *Ace* insensitivity mutations contribute to malathion resistance in the DGRP

Four common amino acid substitutions in *D. melanogaster Ace* are associated with insecticide resistance (Mutero *et al.* 1994). The ability of these mutations to increase the insensitivity of *Ace* to organophosphate and carbamate insecticides has been quantified (Menozzi *et al.* 2004), and variation at all four sites exists among DGRP lines (I199V [3R:9069721], G303A [3R:9069408], F368Y [3R:9069054] and G406A [3R:9063921]; fig. 2.2.7).

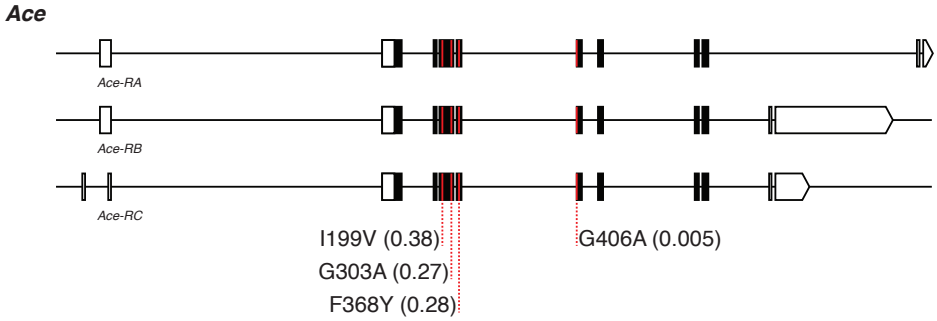


Figure 2.2.7. Amino acid substitutions conferring insecticide resistance in *D. melanogaster Ace*
Relative location (and DGRP allele frequencies) of the four common amino acid substitutions which confer insecticide resistance in *D. melanogaster Ace* (Mutero *et al.* 1994). Exons of the three annotated transcripts are represented in black.

The three most common of these nonsynonymous sites independently achieved Bonferroni-significant associations with male and female malathion mortality at 6 hours. G406A is only present in a single DGRP line (RAL-399) and was therefore excluded by the GWAS' low allele frequency threshold. Only three of the eight possible combinations of these *Ace* alleles are present in the DGRP at moderate frequencies: The ancestral, susceptible *Ace* haplotype (Ace^{IGFG}), and two resistant *Ace* substitution haplotypes: Ace^{VGFG} (one substitution) and Ace^{VAYG} (three substitutions). Ace^{VGFG} and Ace^{VAYG} enzymes have inhibitory constants of 6.4 and 32 to malaoxon (the activated form of malathion) respectively (Menozzi *et al.* 2004), consistent with DGRP population mean malathion mortalities for each haplotype (fig. 2.2.8). This suggests that the previously characterized role of *Ace* resistance substitutions explains the strong malathion associations detected within *Ace* and the surrounding haplotype.

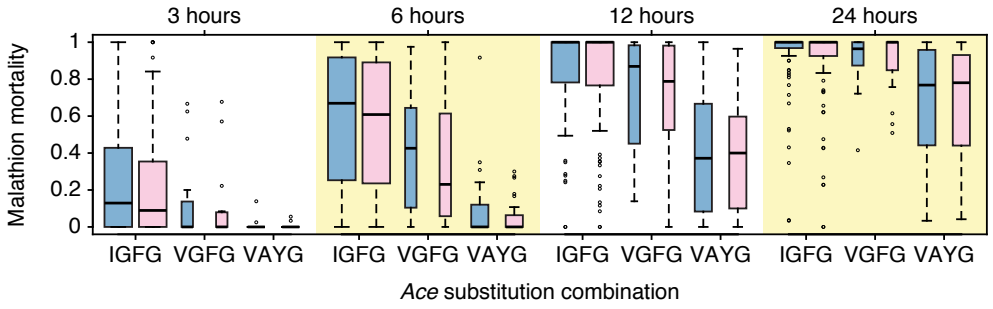


Figure 2.2.8. Influence of *Ace* substitutions on malathion phenotypes

Distributions of DGRP malathion phenotypes partitioned by combinations of *Ace* resistance substitutions. Blue bars represent males and pink bars represent females. Boxes indicate the 25 to 75 percentile range, whiskers indicate the nine to 91 percentile range.

2.2.1.9. Permethrin resistance is associated with structural variation in the cytochrome P450 cluster

The strongest genome-wide associations for permethrin phenotypes were from variants in *Cyp317a1* and *Cyp6a23* (fig. 2.2.5A,B), members of a nine-gene cluster of cytochrome P450s on the right arm of chromosome 2. Two deletions involving genes from this region (*Cyp6a17* and *Cyp6a23*) have previously been described in DGRP lines (Zichner *et al.* 2013; Good *et al.* 2014; fig. 2.2.9), one of which (*Cyp6a17*^{del}) is significantly correlated with all permethrin phenotypes ($6.93\times10^{-6}\leq p\leq2.89\times10^{-3}$, two-tailed t-test assuming unequal variances; fig. 2.2.10).

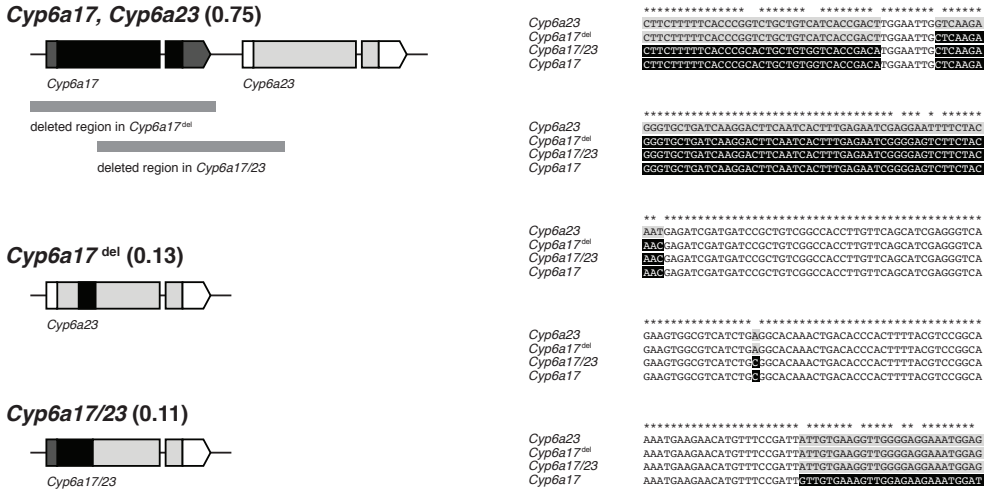


Figure 2.2.9. DGRP structural variation involving *Cyp6a17* and *Cyp6a23*

Structurally variant alleles involving *Cyp6a17* and *Cyp6a23*, and their frequencies among DGRP lines, inferred from sequence read mapping (reproduced from Good *et al.* 2014). Grey bars indicate deletion events inferred in the creation of structural variants.

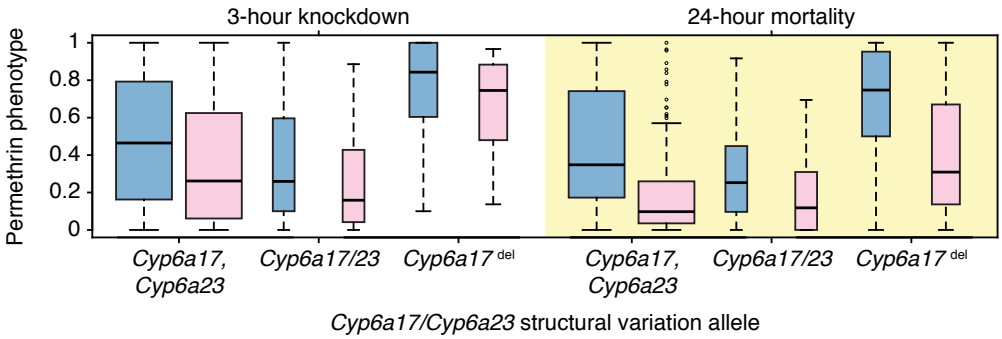


Figure 2.2.10. Influence of structural variation at *Cyp6a17* and *Cyp6a23* on permethrin phenotypes
Distributions of DGRP permethrin phenotypes partitioned by structural variation at *Cyp6a17* and *Cyp6a23*. Blue bars represent males and pink bars represent females. Boxes indicate the 25 to 75 percentile range, whiskers indicate the nine to 91 percentile range.

2.2.2. Transgenic validation of genome-wide association study candidates

2.2.2.1. *Cyp6g1* overexpression increases resistance to organophosphate insecticides

Azinphos-methyl and malathion phenotypes are strongly associated with the presence of derived *Cyp6g1* alleles, which have been demonstrated to increase expression of *Cyp6g1* in key metabolic tissues (Daborn *et al.* 2002; Chung *et al.* 2007; Schmidt *et al.* 2010). Previously, Pyke *et al.* (2004) mapped resistance to another organophosphate, diazinon, in an Australian natural population of *D. melanogaster* to a region containing *Cyp6g1*. The effect of *Cyp6g1* overexpression on azinphos-methyl resistance in larvae was tested using the GAL4-UAS system. Expression of a *Cyp6g1* transgene was driven in the midgut, Malpighian tubules and fat body using upstream elements from a derived *Cyp6g1* allele (Chung *et al.* 2007), resulting in a significant 6.5-fold increase in LC_{50} on azinphos-methyl-laced media compared to control crosses without the transgene (fig. 2.2.11A).

The involvement of *Cyp6g1* in azinphos-methyl resistance was further evaluated in adult flies using RAL-517-*Cyp6g1*^{KO}, a DGRP line in which the natural derived allele *Cyp6g1*^{BA} was removed (Denecke *et al.* 2017; fig. 2.2.11E). A significantly decreased azinphos-methyl LC_{50} in both male and female RAL-517-*Cyp6g1*^{KO} flies was observed when compared to unmodified RAL-517 flies (fig. 2.2.11B). RAL-517-*Cyp6g1*^{KO} was also used to test the involvement of *Cyp6g1* in adult resistance to other organophosphates diazinon and malathion. In both cases, removal of the natural *Cyp6g1*^{BA} allele from RAL-517 significantly reduced resistance to diazinon and malathion in male and female adults (fig. 2.2.11C,D).

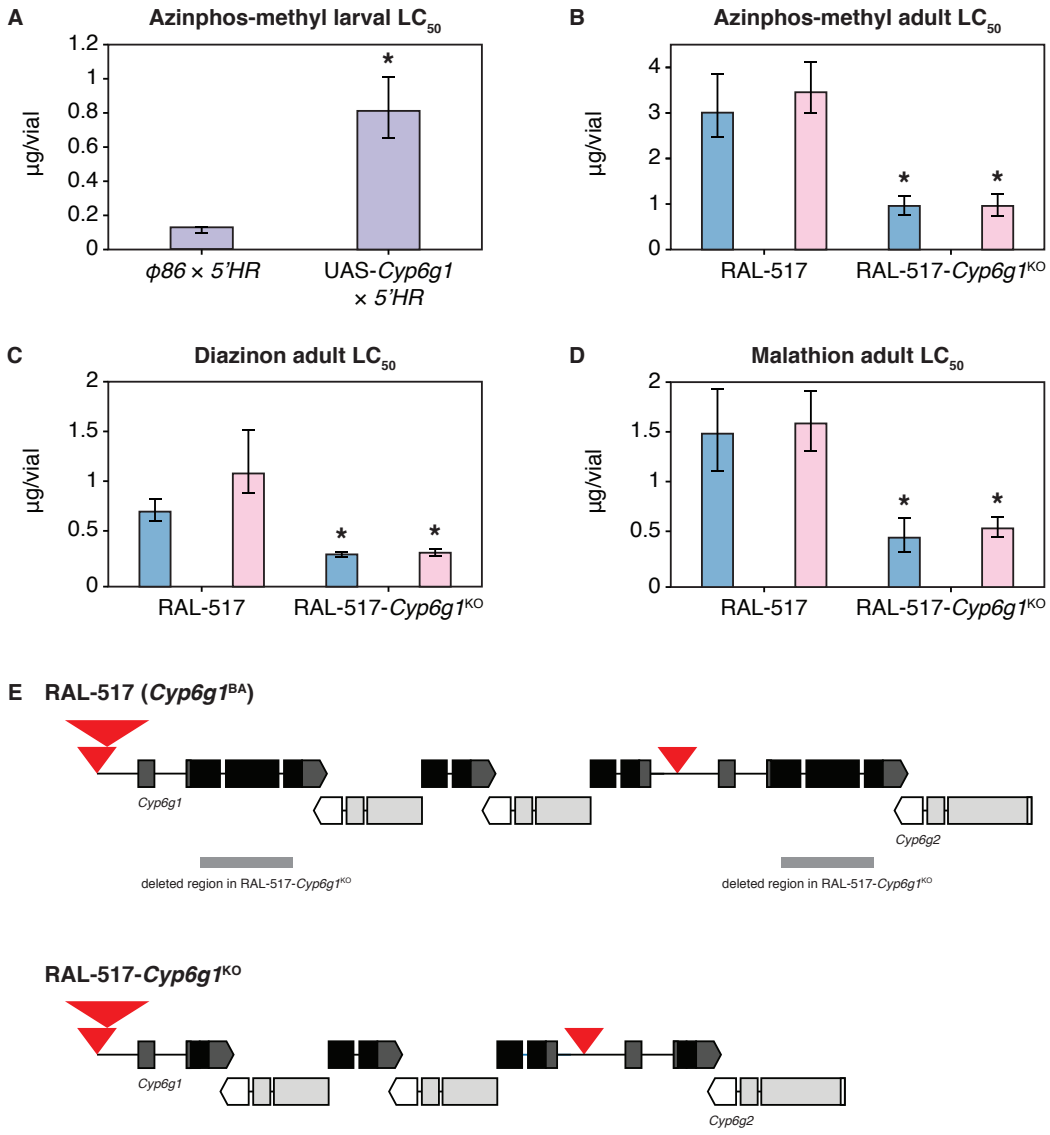


Figure 2.2.11. Functional validation of *Cyp6g1* in organophosphate resistance

Insecticide LC_{50}s for larvae (A) and both male and female adults (B-D). Error bars represent 95% confidence interval. Blue bars represent males and pink bars represent females. Transgenic overexpression of *Cyp6g1* in key metabolic tissues confers resistance to azinphos-methyl (A). CRISPR knockout of *Cyp6g1* in the DGRP line RAL-517 background (E) significantly increases susceptibility to organophosphate insecticides azinphos-methyl (B), diazinon (C) and malathion (D) for adults of both sexes.

2.2.2.2. Disruption of *Cyp6a17* reduces resistance to permethrin

A structural variant involving *Cyp6a17* and *Cyp6a23* is correlated with increased susceptibility to permethrin among DGRP lines. This chimeric deletion fuses coding regions of *Cyp6a17* and *Cyp6a23*, leaving a single gene composed almost entirely of *Cyp6a23* sequence, save for four nucleotides derived from *Cyp6a17* sequence. One of these represents a nonsynonymous (I105V) substitution in *Cyp6a23* sequence; the remaining three, synonymous substitutions. To validate this relationship, a Gene Disruption Project line (*Cyp6a17*^{KG04448}; Bellen *et al.* 2011) with a transposable element inserted into the first exon of *Cyp6a17* was phenotyped on permethrin, revealing it to be significantly more susceptible than genetic background line *y w* (fig. 2.2.12).

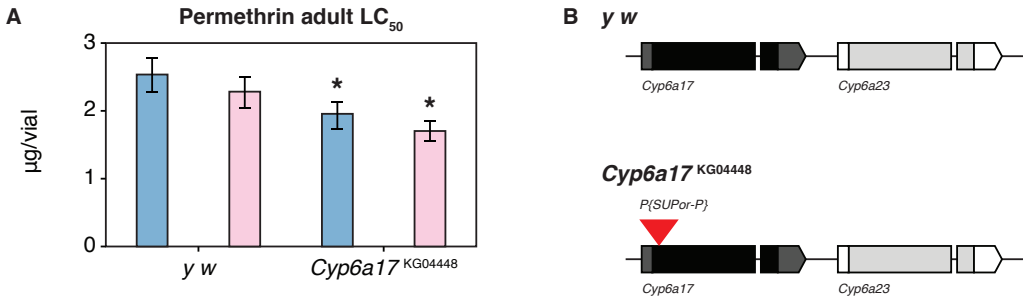


Figure 2.2.12. Functional validation of *Cyp6a17* in permethrin resistance

Gene Disruption Project line *Cyp6a17*^{KG04448} (B) shows significantly decreased permethrin LC_{50}s for male and female adults relative to genetic background control *y w* (A). Error bars represent 95% confidence interval. Blue bars represent males and pink bars represent females.

2.2.2.3. *CHKov1 Doc* insertion does not confer resistance to azinphos-methyl

It has previously been reported that insertion of a *Doc* transposable element into the coding region of *CHKov1* confers resistance to azinphos-methyl in *D. melanogaster* adults (Aminetzach, Macpherson and Petrov 2005). The $CHKov1^{Doc}$ allele is present in the DGRP and although this variant is not directly genotyped, it is detectable in GWAS by annotated variants in linkage disequilibrium with the insertion (Magwire *et al.* 2011; e.g. `3R:21,149,736`). However, none of the azinphos-methyl phenotypes from this work were associated with variants in the *CHKov1* region at the genome-wide significance threshold (mixed $p<2\times10^{-6}$). Correspondingly, there was no significant difference (at the LC_{50} or any concentration) between azinphos-methyl phenotype means partitioned by *Doc* insertion genotype at *CHKov1* (fig. 2.2.13).

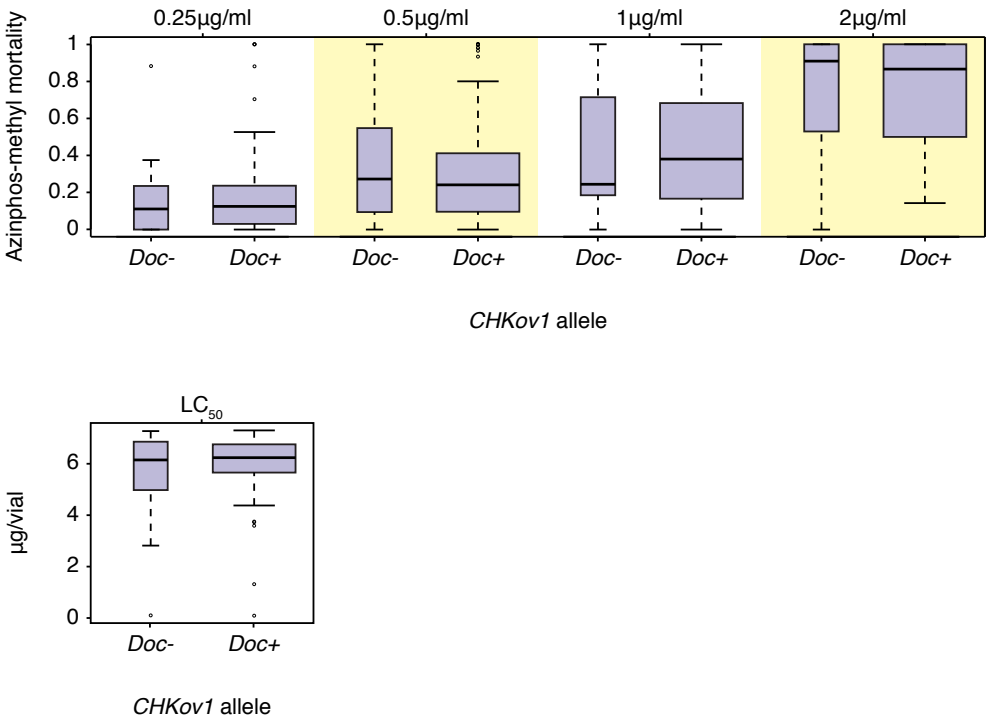


Figure 2.2.13. Influence of *Doc* insertion into *CHKov1* on azinphos-methyl phenotypes
Distributions of DGRP azinphos-methyl phenotypes partitioned by presence of *Doc* insertion into *CHKov1* coding region. Boxes indicate the 25 to 75 percentile range, whiskers indicate the nine to 91 percentile range.

To further investigate this result, CRISPR mutants of *CHKov1* were obtained from Chuan Cao (pers. comm.) and phenotyped for azinphos-methyl mortality at the adult life stage. Although there was a significant difference in LC_{50}s between the two genetic backgrounds of these mutants, there was no significant difference between lines in the three relevant comparisons (fig. 2.2.14 i and ii vs iii, i and ii vs iv, v vs vi).

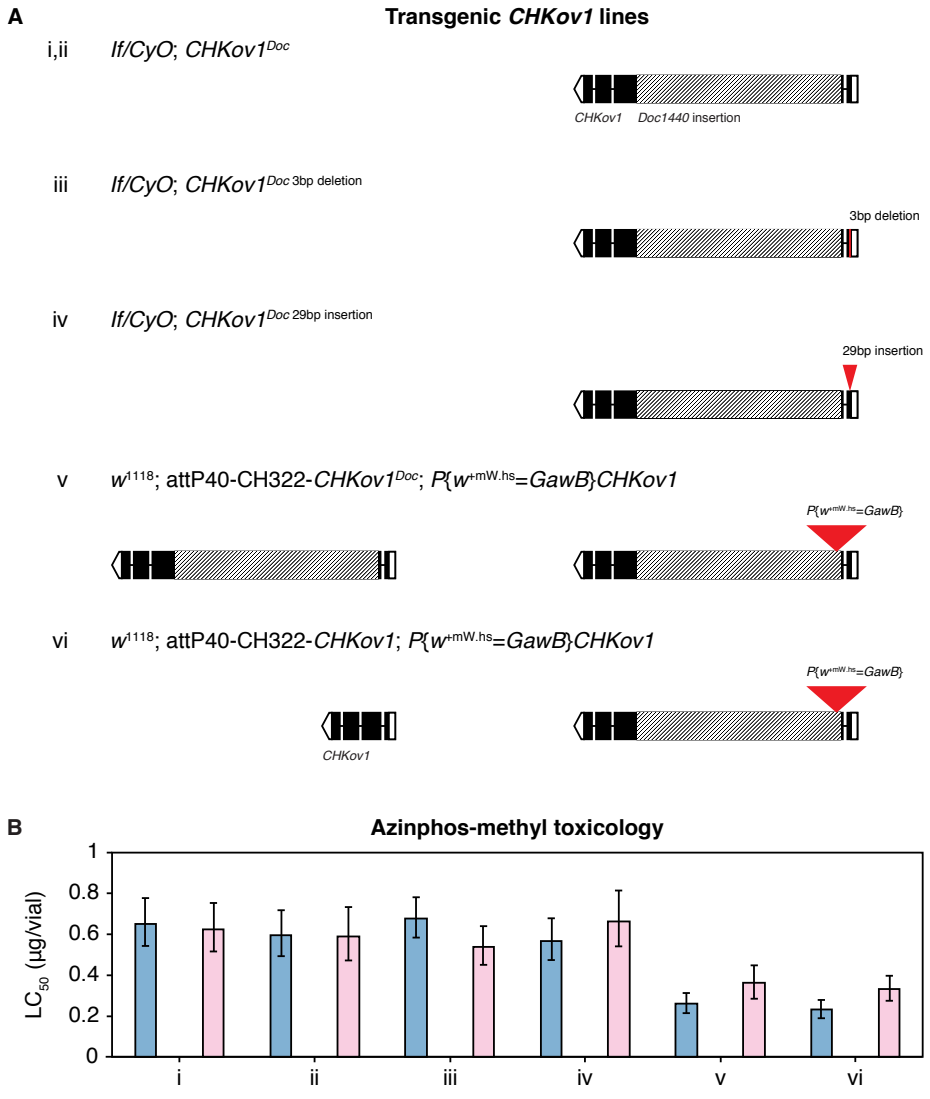


Figure 2.2.14. Toxicology of *CHKov1* mutants

Homozygous transgenic *CHKov1* lines (A) include CRISPR-mediated 3bp deletion (iii) and 29bp insertion (iv) into a $CHKov1^{Doc}$ genetic background. Controls for CRISPR lines are two non-mutants from CRISPR injections (i,ii). Both $CHKov1^{Doc}$ (v) and *CHKov1* ancestral (vi) alleles were also inserted at the attP40 site on chromosome two in a genetic background with a disrupted $CHKov1^{Doc}$ allele. Azinphos-methyl LC_{50}s (B) for both male (blue) and female (pink) adults showed no significant difference in relevant comparisons. Error bars represent 95% confidence interval.

2.3. Discussion

2.3.1. Variation in associations among nested phenotypes

In these experiments, the DGRP was assayed for resistance to azinphos-methyl, malathion and permethrin, representatives of two of the most widely used insecticide classes; organophosphates and pyrethroids. Multiple 'nested' phenotypes were collected for each insecticide, increasing the resolution with which variants contributing to resistance could be identified. For azinphos-methyl, larvae were screened at multiple concentrations, and for malathion male and female adults were screened separately at multiple timepoints. In the case of permethrin, knockdown and mortality phenotypes were measured for male and female adults.

Associated candidates were often restricted to phenotypes from one insecticide, and even within an insecticide the majority of associations significant at the genome-wide association threshold were restricted to a specific sex and/or phenotype. For example, variants in and around a major azinphos-methyl and malathion candidate, *Cyp6g1* were only significant at the lowest two azinphos-methyl concentration phenotypes (0.25 and 0.5μg/ml; fig. 2.2.1A,B), and for females at the three-hour malathion mortality phenotype (fig. 2.2.4A). Different phenotypes may result in different but genuine associated candidates through two avenues: biological effects, like sexual dimorphism or a threshold exposure time required to elicit a response, (i.e. gene induction; Willoughby *et al.* 2006), and statistical effects that are consequences of skewed phenotypic distributions, whereby genes with low minor allele frequencies can be associated at more extreme phenotypes. *Cyp6g1* appears to fall into the latter category, as the susceptible allele at this locus is rare (only present in nine DGRP lines) and the transgenic results demonstrate that *Cyp6g1* confers resistance to azinphos-methyl and malathion at a range of concentrations in both sexes (fig. 2.2.11A,B,D). While this result appears superficially technical, it highlights the context-dependence of resistance alleles: In the DGRP's progenitor population, the frequency of derived *Cyp6g1* alleles had been driven so high at the time of sampling that its effect is detectable only at extreme thresholds (fig. 2.3.1).

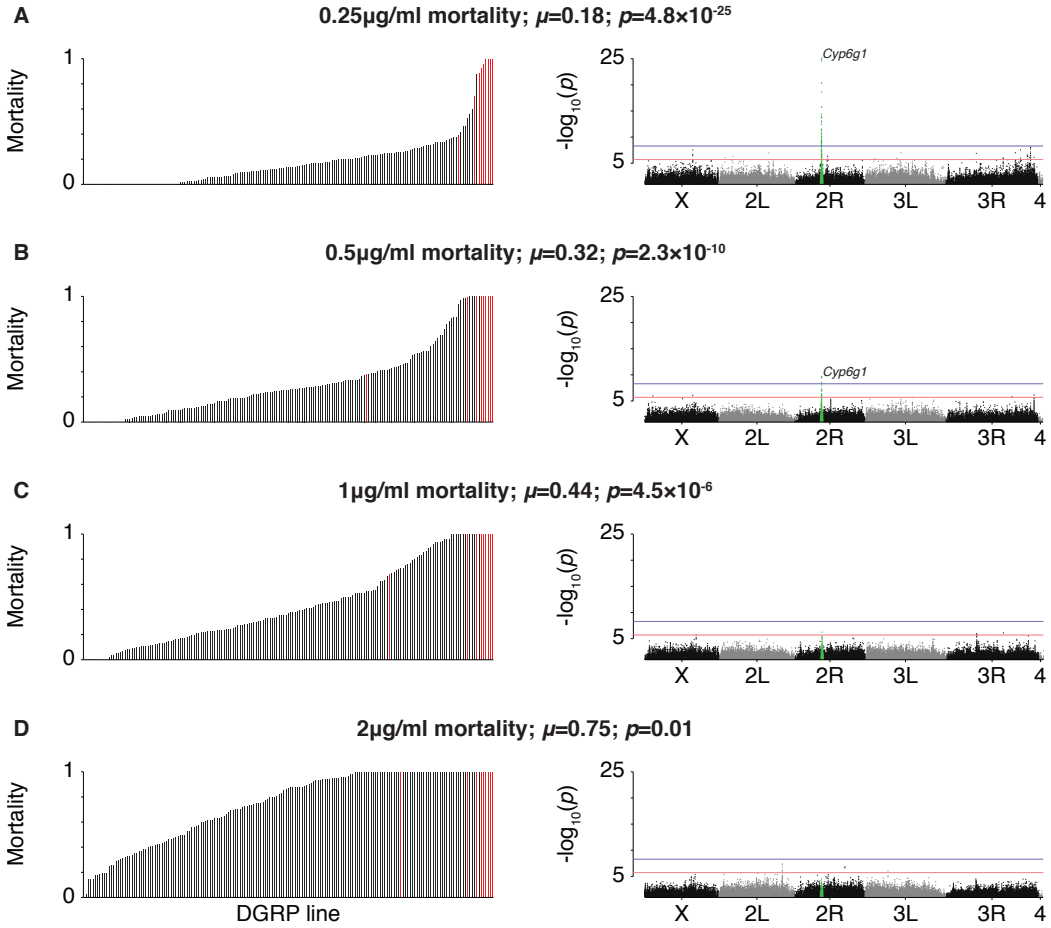


Figure 2.3.1. The relationship between phenotypic distribution and strength of association
Individual DGRP line mortality means for four azinphos-methyl single-concentration phenotypes ordered separately for each phenotype (left), and corresponding Manhattan plots reproduced from fig. 2.2.1 (right). Lines carrying ancestral *Cyp6g1*^{M} alleles are denoted by red bars. While the relative susceptibility of *Cyp6g1*^{M} lines does not change drastically with increasing concentrations, the strength of the association between the phenotype and the *Cyp6g1*^{M} allele (p) is increased as the skew of the distribution shifts (measured by the population mean, μ).

2.3.2. Organophosphate associations and selective sweeps

A major motivation in the selection of compounds for these experiments was the experimental validation of DGRP selective sweep peaks speculated to be driven by insecticide-based selection (Garud *et al.* 2015). Of particular interest, azinphos-methyl resistance has been previously linked to two sweep peaks, *Ace* and *CHKov1* (Menozzi *et al.* 2004; Aminetzach, Macpherson and Petrov 2005).

Insertion of a *Doc* transposable element into the coding region of *CHKov1* has previously been associated with resistance to azinphos-methyl in a single, introgressed *D. melanogaster* line (Aminetzach, Macpherson and Petrov 2005). More recently, Magwire *et al.* (2011) found, through linkage mapping and a subsequent DGRP GWAS, that the $CHKov1^{Doc}$ allele was associated with resistance to the Sigma virus. Magwire *et al.* (2011) were able to detect $CHKov1^{Doc}$ in their GWAS from a haplotype of SNPs in linkage disequilibrium with the insertion, however a similar result was not observed for any azinphos-methyl phenotype. This negative result was reinforced through *in silico* genotyping DGRP lines for $CHKov1^{Doc}$, producing patterns which once again showed no association with any azinphos-methyl phenotype measured (fig. 2.2.13).

There are at least two explanations for the discrepancy between the work of Aminetzach, Macpherson and Petrov (2005) and the work described here; differences in genetic background between experiments, and the difference in life stage of flies used in the experiments (Aminetzach, Macpherson and Petrov [2005] performed their toxicology on adult flies, this DGRP work used larvae). To account for these factors, adult male and female flies from three $CHKov1^{Doc}$ CRISPR mutant lines and their controls were phenotyped on azinphos-methyl. However, no significant effect on azinphos-methyl resistance attributable to $CHKov1^{Doc}$ variation was observed (fig. 2.2.14). Given that the azinphos-methyl phenotype ascribed to $CHKov1^{Doc}$ by Aminetzach, Macpherson and Petrov (2005) was unable to be replicated either through population-level screening of the DGRP or phenotyping of transgenic lines, Sigma virus resistance (Magwire *et al.* 2011) seems a more likely explanation than insecticide resistance for the signatures of selection observed around the $CHKov1^{Doc}$ allele.

Another DGRP selective sweep peak putatively driven by insecticide resistance surrounds Acetylcholineesterase (Ace), the molecular target of organophosphate and

carbamate insecticides. Four non-synonymous substitutions in the enzyme's active groove have been demonstrated to affect the binding capacity of 17 insecticides from the organophosphate and carbamate classes (Menozzi *et al.* 2004; organophosphate insecticides require activation *in vivo* to achieve full potency [Siegfried and Scharf 2001], and the activated 'oxon' compounds were used in this work). In most cases the substitutions reduced the affinity of Ace to insecticides, hence increasing resistance at the molecular level. All sixteen combinations of the four Ace substitutions were characterized by Menozzi *et al.* (2004), however only three substitutions, existing in three combinations are present among DGRP lines at moderate frequencies: Ace^{IGFG} (no substitutions), Ace^{VGFG} (one substitution), and Ace^{VAYG} (three substitutions). For both Ace^{VGFG} and Ace^{VAYG}, Menozzi *et al.* (2004) report a much larger reduction in binding affinity to malaoxon (the activated form of malathion) than azinphos-methyl oxon (the activated form of azinphos-methyl) relative to the ancestral Ace^{IGFG} state (fig. 2.3.2). This provides an explanation for why strong associations with *Ace* variants were detected in malathion (fig. 2.2.3; fig 2.2.4) but not azinphos-methyl (fig. 2.2.1; fig. 2.2.2) GWAS, and also indicates a limit at which variation in the molecular insensitivity phenotype will be detectible at the population scale. These results suggest that selection for insecticide resistance from the use of malathion and other organophosphate insecticides could explain the selective sweep signals detected around the *Ace* locus among DGRP lines.

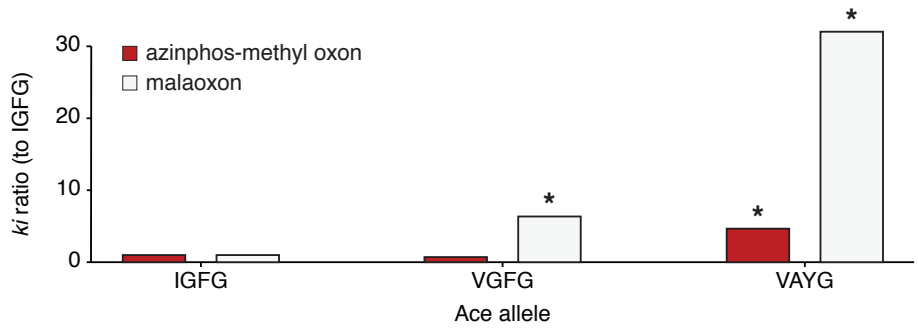


Figure 2.3.2. Ace inhibitory constants (data from Menozzi *et al.* 2004)

ki ratios (relative to the ancestral IGFG allele) of activated forms of azinphos-methyl (methyl-azinphos oxon) and malathion (malaoxon) for insecticide-resistant Ace alleles present in the DGRP. Asterisks represent statistically significant differences from the IGFG allele.

The third of Garud *et al.*'s selective sweeps ostensibly driven by insecticide resistance includes near its peak the cytochrome P450 gene *Cyp6g1*. *Cyp6g1* was first directly linked to the insecticide DDT by Daborn *et al.* (2001), and it was subsequently shown that DDT-resistant lines of *D. melanogaster* contain the long terminal repeat (LTR) region of an *Accord* transposable element inserted upstream of the gene, which correlates with increased *Cyp6g1* expression (Daborn *et al.* 2002). *Cyp6g1*-linked cross-resistance was described to the neonicotinoids imidacloprid (Daborn *et al.* 2001) and nitenpyram (Daborn *et al.* 2007), and the capacity of the enzyme to metabolize insecticides was demonstrated in *Nicotiana tabacum* cell culture (Joußen *et al.* 2008) and in *D. melanogaster* larval tissue (Hoi *et al.* 2014). Although the link between natural alleles which overexpress *Cyp6g1* and resistance to DDT has been demonstrated in a worldwide sample (Catania *et al.* 2004) and Australian populations (Schmidt *et al.* 2010), a similar result was not observed in a DGRP GWAS of DDT resistance (Schmidt *et al.* 2017).

In contrast, GWAS of resistance to both azinphos-methyl and malathion resulted in associations with multiple variants in and around *Cyp6g1* (fig. 2.2.1A,B; fig. 2.2.4A), and transgenic experiments demonstrated that overexpression of the enzyme confers resistance to the organophosphates azinphos-methyl, diazinon and malathion (fig. 2.2.11A-D). Previous investigations have mapped organophosphate resistance to a region including *Cyp6g1*. In 1961 Kikkawa mapped parathion resistance to a region on chromosome two and reported cross-resistance to malathion. In 2004, Pyke *et al.* mapped diazinon resistance to a similar region. Ogita (1958) reported that the DDT resistance mapping region on chromosome two conferred cross resistance to malathion. Le Goff *et al.* (2003) also reported malathion cross-resistance in DDT-resistant lab lines (*Hikone-R* and *Wisconsin*), both of which showed heightened levels of *Cyp6g1* transcript. Likewise, DDT-resistant *91-R* (which carries a resistance allele at the *Cyp6g1* locus [Schmidt *et al.* 2017] and overexpresses the enzyme [Pedra *et al.* 2004]), shows cross resistance to malathion (Misra *et al.* 2013). However, the experiments described in this chapter demonstrate for the first time that naturally-occuring *Cyp6g1* alleles are strongly associated with resistance to a range of organophosphate insecticides, and suggest that organophosphate selection may be a more likely explanation than DDT for the sweep observed at the *Cyp6g1* locus in the DGRP's progenitor population (Schmidt *et al.* 2017; §4.3.3.2).

2.3.3. Permethrin phenotype associations

DGRP variants most strongly associated with permethrin phenotypes map to a region on the right arm of chromosome two containing nine P450 genes, with peaks over *Cyp6a23* and *Cyp317a1*. Structural variation in the DGRP has previously been described involving *Cyp6a23* and its tandem paralog *Cyp6a17* (Zichner *et al.* 2013; Good *et al.* 2014). The DGRP harbors two deletion alleles in this region, one resulting in a single chimeric gene comprised of *Cyp6a17* and *Cyp6a23* sequence. In the other, *Cyp6a17* is deleted, save for a small section which exists as a putative gene conversion in the otherwise intact *Cyp6a23*. Due to the homology between these genes, this putative gene conversion introduces only four nucleotide changes and a single non-synonymous substitution in *Cyp6a23* (Good *et al.* 2014). Among DGRP lines, it is this deletion of *Cyp6a17* that is associated with increased susceptibility to permethrin (fig. 2.2.10). This is congruent with the finding that when *Cyp6a17* is disrupted in $Cyp6a17^{KG04448}$, it leads to relative susceptibility (fig. 2.2.12).

An additional intriguing permethrin candidate is the gain of an additional start site in the phospholipase-encoding *iPLA2-VIA*. This common variant (MAF=0.41) is associated with increased three-hour knockdown in females (appendix 1.9). Knockout of *iPLA2-VIA* in *D. melanogaster* has been shown to increase sensitivity to oxidative stress (Kinghorn *et al.* 2015), which is known to be caused by insecticide exposure and has been specifically demonstrated to occur in *D. melanogaster* following permethrin ingestion (Terhzaz *et al.* 2015). Natural variation in oxidative stress pathways has been observed in insecticide-resistant insects (Misra *et al.* 2013; Wilding 2018), however the genetic basis of this variation is undetermined.

2.3.4. Conclusions

In this investigation of insecticide resistance in the DGRP using two organophosphate and one pyrethroid insecticides, stark differences were observed in the genes involved as well as the evidence for a selective response to the compounds. The top candidates from azinphos-methyl and malathion GWAS correlate with peaks of selection in the DGRP population, making organophosphate resistance a highly credible selective pressure on the population ancestral to the DGRP. Conversely, *Cyp6a17,* the top permethrin candidate, does not lie within a selective sweep peak, nor would the allele described be expected to be the target of positive insecticide-based selection, given that it increases susceptibility to permethrin. However, *Cyp6a17* ranks fourth among *D. melanogaster* P450s (after *Cyp6a13, Cyp6a2,* and *Cyp6a14*) in similarity to malaria vector *Anopheles funestus CYP6P9a* and *CYP6P9b*. Naturally occurring duplications of each of these genes are associated with pyrethroid resistance in *A. funestus* (Wondji *et al.* 2009), and selective sweeps have been described at *CYP6P9a* in response to pyrethroid-based malaria interventions (Barnes *et al.* 2017).

This suggests that although the DGRP harbors variation in resistance to both organophosphate and pyrethroid insecticides, selection by organophosphate compounds has played a far more important role in the evolutionary history of the DGRP's progenitors than pyrethroids. While there are many possible explanations for this pattern (such as a fitness cost limiting resistance mutations at the pyrethroid target site, *para* [Pittendrigh *et al.* 1997]), the focus of selective sweeps on organophosphate resistance loci is most likely due to the widespread use of insecticides from this class: Organophosphates dominated the insecticide market in the United States in the years leading up to the DGRP's collection (Atwood and Paisley-Jones 2017).

Furthermore, although generally supporting the hypothesis that top DGRP sweeps are attributable to insecticide resistance (Garud *et al.* 2015), analyses described in this chapter have identified several departures from expectations at overlaps between selective sweeps and insecticide phenotypes. This highlights the importance of such validation experiments for putatively adaptive alleles, and also underscores the utility of the DGRP for such analyses.

2.4. Materials and methods

2.4.1. Fly lines

DGRP lines were generated by Mackay *et al.* (2012) and obtained from the Bloomington Drosophila stock center in Indiana. The *6g1*HR-GAL4 driver and UAS-*Cyp6g1* lines were generated by Chung *et al.* (2007). The RAL-517-*Cyp6g1*^{KO} line was generated by Denecke *et al.* (2017). *CHKov1* mutants were generated by, and obtained from Chuan Cao (pers. comm.). The *Cyp6a17*^{KG04448} line was generated by Bellen *et al.* (2011); this and its control, *y w* and were obtained from the Bloomington Drosophila Stock Center (ID 25741 and 6599 respectively). All fly stocks were maintained at 25°C on rich medium containing, maltose (46g/l), dextrose (75g/l), yeast (35g/l), soy flour (20g/l), maize meal (73g/l), agar (6g/l), acid mix (14ml/l), and tegosept (16ml/l). The acid mix solution was made up of orthophosphoric acid (42ml/l), and propionic acid (412ml/l), while the tegosept solution was 50g tegosept dissolved in 950ml of 95% EtOH.

2.4.2. Phenotype selection

Calculation of the median lethal concentration (LC_{50}) is the standard method for assessing insecticide resistance, and was used as the phenotype for azinphos-methyl larval screens along with four individual concentration phenotypes measured for its calculation. For the more elaborate adult screening used for malathion and permethrin, a single discriminant concentration (1µg/vial for malathion, 10µg/vial for permethrin) was identified by screening 20 randomly selected DGRP lines on a range of concentrations. Twenty-four-hour mortality in adult insects was selected as a phenotype as this is also a standard assay in *D. melanogaster* insecticide resistance literature (Daborn *et al.* 2001; Schmidt *et al.* 2010; Schmidt *et al.* 2017), and DGRP lines show negligible control mortality in scintillation vial bioassays at this timepoint (Schmidt *et al.* 2017). 'Knockdown', a phenotype in which the fly lies paralyzed and twitching, or exhibits uncontrolled flight, is a field-relevant effect of pyrethroid insecticides. Therefore, incidence of this phenotype was also measured at three hours in permethrin-treated flies. As no pronounced knockdown effect was expected (or observed) in malathion-treated flies, mortality was instead scored at additional timepoints (three, six and 12 hours). A minimum of three biological replicates were

performed for each sex of each DGRP line, and for each sex at each concentration for transgenic lines.

2.4.3. Insect Bioassays

For azinphos-methyl larval screens, first-instar larvae (<24 hr old) were collected from laying plates and transferred onto insecticide media at a density of 20 larvae per vial. Controls were performed using media containing no insecticide. The number of fully formed pupae were scored after 7 days. DGRP lines were screened on four common concentrations (0.25, 0.5, 1 and 2µg/ml) with additional concentrations (ranging from 0.0625 to 8µg/ml) used to calculate the LC_{50} of lines with extreme phenotypes. A minimum of three biological replicates were performed for each concentration, and all concentration data were corrected for control mortality using Abbott's correction.

For malathion and permethrin adult screens, flies were anaesthetised with CO_2 at 0-24 hours after eclosion and sorted by sex into holding vials containing rich media, where they were kept for 3-4 days, resulting in 3-5 day old adults for use in bioassays. Assays commenced between 11am and 12pm. 20ml glass scintillation vials were treated with 500µl of acetone/insecticide solution at the required concentration and rotated using the rollers of a hot dog grilling machine (with the heat off) until the acetone had evaporated. ~7 flies were transferred to each vial without the use of anesthesia, and cotton wool moistened with 10% sucrose solution was used to stopper the scintillation vials. DGRP lines were screened at a single concentration, while transgenic lines were screened at multiple concentrations and scored for mortality at 24 hours to enable the calculation of the LC_{50}.

2.4.4. Calculation of LC_{50}

For each DGRP line in the larval azinphos-methyl screen, and all transgenic lines, linear models were fitted to concentration-mortality data on a log-probit scale using `glm` in R (R Core Team 2018) and scripts from Johnson *et al.* (2013). LC_{50} values and 95% confidence intervals were calculated using Fieller's method from fitted linear models (Finney 1971).

2.4.5. Estimation of broad-sense heritability and sex effect

The broad-sense heritability of each DGRP phenotype (excluding azinphos-methyl LC_{50}) was estimated as $\sigma^2_G / (\sigma^2_G + \sigma^2_E)$, using the variance components of a linear model of the form *phenotype ~ 1 + (1 | line)* using the 'lme4' package in R (R Core Team 2018). For malathion and permethrin, heritability of each phenotype for each sex was estimated separately. Male sex effect was estimated from the male sex intercept of the model *phenotype ~ sex + (0 + sex | line)* using the 'lme4' package in R (R Core Team 2018).

2.4.6. Genome-wide association studies

Phenotype files for 170 DGRP lines, consisting of mean mortality data for both males and females, were generated for all phenotypes and were submitted to the Mackay lab DGRP2 GWAS pipeline (`http://dgrp2.gnets.ncsu.edu/`; Huang *et al.* 2014). The genome wide significance threshold (1×10^{-5}) was corrected for the number of phenotypes tested for each insecticide (5, 8 and 4 for azinphos-methyl, malathion and permethrin respectively) and applied to the Mackay lab pipeline 'mixed *p*-value' (association after correction for the effects of Wolbachia and major chromosomal inversions). Bonferroni significance thresholds were calculated as 0.05 divided by the product of the number of phenotypes tested and the number of genomic variants tested (1,904,240, 1,877,810 and 1,876,330 for malathion and permethrin respectively) for each insecticide.

2.4.7. *in silico* genotyping of structural variation

BAM files containing alignments of DGRP line sequences from Illumina platforms to the *y; cn bw sp;* reference genome were recovered from the Baylor College of Medicine website (`https://www.hgsc.bcm.edu/content/dgrp-lines`; Mackay *et al.* 2012). Local alignments at *Cyp6g1, Cyp6a17/23* and *CHKov1* were visualized with IGV software (Thorvaldsdóttir, Robinson, and Mesirov 2013) to manually score structural variation.

3

Transcriptome-wide association studies of insecticide resistance

After more than a decade of GWAS in humans, the prevailing view of quantitative trait architecture is that a huge number of small-effect polymorphisms perturbing gene expression explain the bulk of trait heritability. GWAS approaches are often underpowered to detect such variants on their own; the use of transcriptome measurements as an 'intermediate phenotype' can increase the efficacy of GWAS as well as ascribing function to associated variants. This chapter describes the incorporation of previously generated DGRP transcriptome data into association studies of insecticide phenotypes, which is achieved using two distinct approaches: identification of insecticide-associated genomic variants which are also significantly associated with levels of a particular transcript (expression quantitative trait loci), and association studies directly between phenotypes and transcript levels (transcriptome-wide association studies). The insecticide phenotypes from chapter 2 are supplemented with phenotypes measured by other researchers on different insecticides to expand these analyses to a total of 35 phenotypes from nine insecticides. The reexamination of insecticide GWAS incorporating transcriptome measurements is predicted to identify novel candidates, refine hypothesized roles for those previously identified, and provide an insight into the regulatory architecture of insecticide resistance.

3.1. Introduction

3.1.1. Insecticide resistance and the transcriptome

Since early in the characterization of insecticide resistance mechanisms, transcriptomic variation, specifically the upregulation of particular genes, has been known to play a large role. Perhaps the earliest example came from Devonshire (1977), who demonstrated that overproduction of a single esterase was the mechanism of organophosphate resistance in a strain of *Myzus persicae*, a phenomenon later ascribed to the amplification within the genome of the gene encoding the enzyme (Field, Devonshire and Forde 1988). The study of insecticide resistance in the housefly *Musca domestica* identified a cytochrome P450 transcript overexpressed in resistant strains (*CYP6A1*; Carino *et al.* 1992) the product of which was subsequently demonstrated to be capable of detoxifying the organophosphate diazinon when transgenically produced in *Escherichia coli* cells (Feyereisen *et al.* 1999). In contrast to the *cis*-acting gene amplification event observed in *M. persicae*, mapping of *CYP6A1*-linked resistance in *M. domestica* suggested upregulation by an unidentified *trans*-acting factor on a different chromosome (Plapp Jr 1984; Cohen *et al.* 1994). In 2002, Daborn *et al.* identified heightened transcription of cytochrome P450 *Cyp6g1* as the major DDT resistance factor in *D. melanogaster*, at least partially attributable to an insertion of an *Accord* LTR into the 5'UTR of the gene (Chung *et al.* 2007). This transposable element insertion was subsequently shown to be just one of multiple structural variants present in insecticide-resistant *Cyp6g1* alleles, although the *Accord* LTR insertion remains the common feature in almost all structurally variant alleles described at this locus to date (Catania *et al.* 2004; Schmidt *et al.* 2010; fig. 2.2.6).

With the advent of microarrays came the opportunity for concordantly measuring the expression levels of all known genes in an organism. One result of these 'transcriptome profiling' experiments was that insecticide-resistant lines of *D. melanogaster* harbour a battery of upregulated transcripts enriched with metabolic gene families previously associated with resistance to insecticides (e.g. Pedra *et al.* 2004). In microarray experiments targeting these gene families, many members demonstrated dramatic induction in response to the exposure of *D. melanogaster* to natural xenobiotic compounds caffeine and phenobarbital, although this phenomenon did not extend to insecticides DDT, diazinon, dicyclanil, lufenuron, nitenpyram or spinosad (Willoughby

et al. 2006). The similar sets of transcripts induced by caffeine and phenobarbital would later be demonstrated (this time using microarrays profiling the whole transcriptome) to overlap those differentially expressed following ectopic activation of *cap-n-collar* isoform-C (*cncC*; the *D. melanogaster* homolog of mammalian *Nrf2*) which regulates response to oxidative stress (Misra *et al.* 2011). Misra *et al.* (2011) also demonstrated that ectopic activation of *cncC* was sufficient to confer resistance to the insecticide malathion. Subsequently Guio *et al.* (2014) linked induction of the *cncC* oxidative stress pathway to the insecticide malathion, although these experiments are not directly comparable to those performed by Willoughby *et al.* (2006) as they were carried out on a different insecticide and over much longer time periods (up to 214 hours).

In 2013, Misra *et al.* demonstrated that the *cncC* oxidative stress pathway is constitutively active in two DDT-resistant lines of *D. melanogaster* (including *91-R*, one of the lines profiled by Pedra *et al.* [2004], in which *cncC* activation explains ~20% of differentially expressed transcripts). This suggests that regardless of the induction response to insecticides, insecticide resistance can occur through constitutive activation of the *cncC* oxidative stress pathway, a phenomenon that has subsequently been described in pest insect species (although not explicitly linked with insecticide resistance; Wilding 2018). While individual transcripts identified in *D. melanogaster* transcriptomic investigations were enriched for gene families canonically involved in insecticide metabolism, the experiments were underpowered (due to their sample size of one to three strains) to link any transcript individually with resistance. Also due to these power restrictions, the genetic basis of the constitutive upregulation of the *cncC* pathway in *91-R* or any other resistant insect remains unresolved.

3.1.2. Quantitative transcriptomics

In humans, transcriptomic experiments powered to detect individual transcripts are now routinely described, through the incorporation of transcriptome data into genome-wide association studies. The value of this approach was quantified by Pavlides *et al.* (2016), who leveraged transcriptome data to identify 71 novel susceptibility genes from 28 human disease phenotypes. The incorporation of transcriptome data into GWAS is achieved in one of two ways (fig. 3.1.1): Transcript abundance can be mapped to genomic variation, known as expression quantitative trait loci (eQTL) mapping (Schadt *et al.* 2003). Mapped eQTL can then be used to describe the links between GWAS-identified variants and the transcriptome. Alternatively, associations can be directly tested between a phenotype and the abundance of each transcript in the transcriptome, known as transcriptome-wide association studies (TWAS). The latter design allows researchers to overcome several limitations of GWAS: In TWAS, multiple variants of small effect in complex regulatory networks are simplified into their transcriptomic consequence. Similarly, rare variants with common transcriptomic effects are grouped together under a single transcript level, bypassing minor allele thresholds imposed by GWAS. Moreover, TWAS are not limited by variant discovery, allowing for the association of phenotypes with the consequence of complex genomic variation excluded from GWAS genotyping.

In 2015, the utility of the DGRP was expanded by the incorporation of microarray transcriptomes from adult whole-body samples of 185 DGRP lines for each sex (Huang *et al.* 2015), allowing for the incorporation of transcriptome data into DGRP GWAS. This population-level transcriptomic investigation yielded the major findings that sexual dimorphism is highly prevalent (42% of transcripts are genetically variable between sexes) and that 32 and 24 percent of male and female transcripts respectively can be arranged into genetically correlated modules. Huang *et al.* (2015) performed eQTL mapping for each transcript, identifying at least one significant (FDR=0.2) eQTL for 1339 and 941 male and female transcripts respectively, and also mapped significant veQTL (quantitative trait loci associated with differences in *variance* in abundance of a particular transcript) for 107 and 248 male and female transcripts respectively.

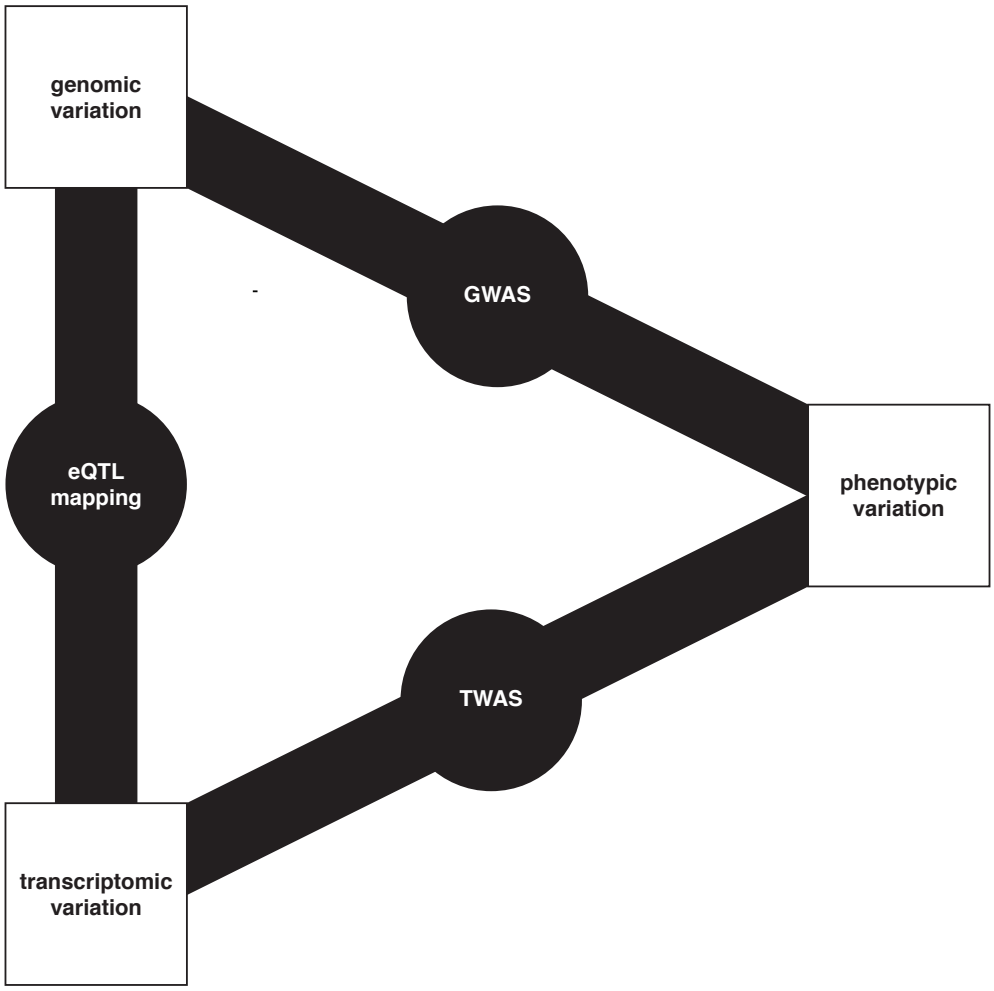


Figure 3.1.1. Association study approaches utilizing genomic, transcriptomic and phenotypic data
Transcriptomic data can be related to phenotypes either by comparing associations from GWAS and eQTL mapping approaches, or by measuring associations between the transcriptome and phenotype directly.

Despite their potential for enriching DGRP GWAS, the Huang *et al.* (2015) datasets have been largely neglected in DGRP association studies. To date, Huang *et al.*'s (2015) eQTL data has not been utilized in any published DGRP GWAS, and only a single publication (studying virgin egg retention; Akhund-Zade *et al.* 2017) describes a TWAS approach using the Huang *et al.*'s (2015) transcriptome data. This excludes publication arising from work described in this chapter (i.e. Battlay *et al.* 2016, Denecke *et al.* 2017, Battlay *et al.* 2018 and Green *et al.* 2019) and also excludes publications utilizing Ayroles *et al.*'s (2009) 40-transcriptome DGRP dataset (e.g. Teets and Hahn 2017).

Here, 35 DGRP resistance phenotypes from nine insecticides are tested for associations with transcriptomic variation utilizing TWAS along with eQTL and veQTL mapping data generated by Huang *et al.* (2015). Insecticide resistance phenotypes are expected to yield strong associations with the abundance of certain transcripts based on previous studies, as well as results from DGRP GWAS described in chapter two. Structural variation at *Cyp6g1* and *Cyp6a17/23*, associated with azinphos-methyl, malathion and permethrin phenotypes likely affects transcriptional output of these genes: *Cyp6g1* structural variation has previously been demonstrated to confer insecticide resistance through overexpression (Daborn *et al.* 2001; Chung *et al.* 2007; Schmidt *et al.* 2010), and the allele associated with permethrin susceptibility involves a complete deletion of *Cyp6a17* coding region. Therefore, we expect these structural variants to be insecticide-associated *cis*-eQTL. In this chapter, this hypothesis is tested and generalized, by mapping eQTL and insecticide associations to structural variants, genome-wide.

3.2. Results

3.2.1. Transcriptome to phenotype associations

3.2.1.1. Phenotypes

Thirty-five DGRP insecticide phenotypes from nine insecticides were compiled. Azinphos-methyl, malathion and permethrin phenotyping is described in §2.2.1. Boric acid, chlorantraniliprole, DDT, deltamethrin, imidacloprid and parathion phenotypes were generated in Najarro *et al.* 2017, Green *et al.* (2019), Schmidt *et al.* (2017), Duneau *et al.* (2018) and Denecke *et al.* (2017) respectively (table 3.2.1).

Table 3.2.1. Details and sources of insecticide phenotypes used in this chapter

insecticide	phenotypes	reference
azinphos-methyl	5; larval survival to pupation (4 concentrations and LC50)	§2.2.1.1
boric acid	1; adult female lifespan	Najarro *et al.* 2017
chlorantraniliprole	7; larval survival to adulthood (6 concentrations and LC50)	Green *et al.* 2019
DDT	2; adult female 4hr knockdown and 24hr mortality	Schmidt *et al.* 2017
deltamethrin	1; adult male 48hr mortality	Duneau *et al.* 2018
imidacloprid	2; larval activity at 2 doses	Denecke *et al.* 2017
malathion	8; adult male and female mortality at 4 timepoints	§2.2.1.2
parathion	5; adult male mortality at 5 timepoints	Duneau *et al.* 2018
permethrin	4; adult male and female 3hr knockdown and 24hr mortality	§2.2.1.2

3.2.1.2. Transcriptome-wide association studies (TWAS)

Associations between each of the 35 insecticide phenotypes and the abundance of each transcript from the DGRP transcriptome (Huang *et al.* 2015) were calculated by fitting linear models (e.g. fig. 3.2.1), and two thresholds were used to assess significant associations: Firstly, the 'transcriptome-wide significance threshold' (1×10^{-3}, which imposes approximately the same level of multiple-testing correction to transcriptome data as the 1×10^{-5} threshold imposes to DGRP genotype data) corrected for the number of phenotypes tested for each insecticide. Secondly, the Bonferroni significance threshold (0.05 corrected for the number of transcripts and phenotypes tested).

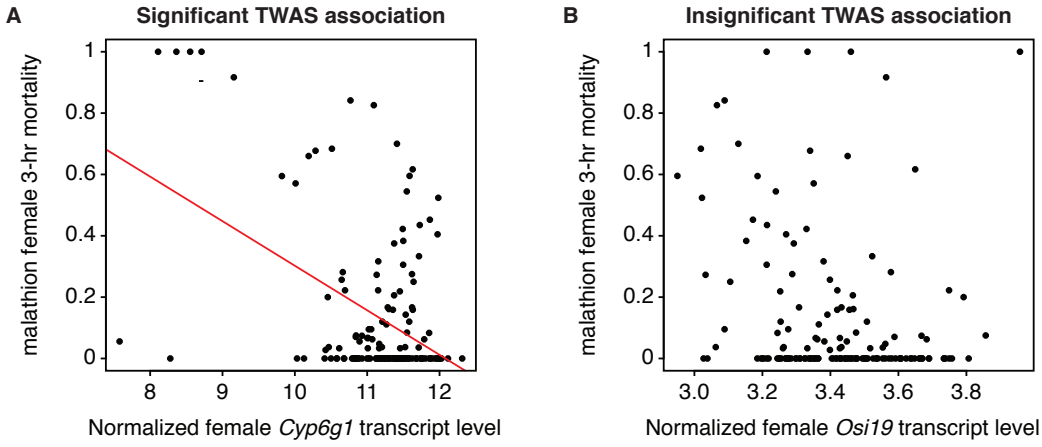


Figure 3.2.1. Examples of TWAS associations

Examples of a significant (A; r^2=0.20; p=2.98×10^{-9}) and an insignificant (B; r^2=0.05; p=3×10^{-3}) association between a phenotype and transcript abundance.

Phenotypes from all nine insecticides returned transcript associations with p-values below their transcriptome-wide significance threshold, with a total of 186 significant associations across all phenotypes (appendix 2). 13 transcripts were associated with phenotypes from two or more insecticides, and two transcripts, *Cyp6g1*, and *CG13759* were associated with phenotypes from three insecticides (table 3.2.2). Additionally, eight associations from three insecticides (azinphos-methyl, chlorantraniliprole, and malathion) achieved Bonferroni-corrected significance (appendix 2). Despite representing less than 0.5% of the transcriptome (86 out of 18,139 transcripts), cytochrome P450 genes dominated these analyses, representing 19 (10.3%) of the 186 associations below the transcriptome-wide significance threshold, five (62.5%) of the eight Bonferroni-significant associations, and five (38.5%) of the 13 transcripts

associated with multiple insecticides (table 3.2.2). In contrast, no significant associations were observed between transcript abundance of genes from the other main detoxification-associated families (carboxylesterases, glutathione S-transferases and UDP glucuronosyltransferases) and any insecticide phenotype.

Table 3.2.2. Transcripts associated with insecticide phenotypes from more than one insecticide

transcript	molecular function gene ontology	male GCTM	female GCTM	insecticides
Cyp6g1	heme binding; iron ion binding; oxidoreductase activity	175	23	azinphos methyl; imidacloprid; malathion
CG13759	N-acetyltransferase activity	128	25	deltamethrin; malathion; parathion
Cyp6g2	heme binding; iron ion binding; oxidoreductase activity	175	23	azinphos methyl; malathion
CG18343	-	-	16	azinphos methyl; imidacloprid
Cyp4d2	heme binding; iron ion binding; oxidoreductase activity	62	22	azinphos methyl; chlorantraniliprole
Uhg1	-	97	-	boric acid; malathion
Cyp12d1-p	heme binding; iron ion binding; oxidoreductase activity	79	23	chlorantraniliprole; malathion
Jheh1	juvenile hormone epoxide hydrolase activity	79	21	chlorantraniliprole; malathion
CG2065	NADP-retinol dehydrogenase activity	79	21	chlorantraniliprole; malathion
Cyp4e2	heme binding; iron ion binding; oxidoreductase activity	79	21	chlorantraniliprole; malathion
CG32264	actin binding	-	-	malathion; parathion
CG34129	-	180	-	malathion; parathion
DNApol-epsilon255	3'-5'-exodeoxyribonuclease activity; DNA-directed DNA polymerase activity	-	-	malathion; parathion

3.2.1.3. Insecticide-associated eQTL and veQTL

eQTL and veQTL are genomic variants which are significantly correlated with mean or variance respectively of a particular transcript among DGRP lines. Significant GWAS associations from 35 insecticide phenotypes (GWAS described in §2.2.1.3 and §4.2.1.1 and listed in appendix 1) were cross-referenced to eQTL and veQTL datasets generated by Huang *et al.* (2015) to produce lists of eQTL and veQTL associated with each insecticide (appendix 1). Of the 1693 unique insecticide-associated variants, 281 (16.6%) were eQTL of at least one transcript (fig. 3.2.2A), and 119 (6.6%) were veQTL of at least one transcript. This represents an enrichment of both classes, as eQTL and veQTL represent roughly 2 and 2.5% of DGRP variants respectively.

eQTL and veQTL can be categorized as acting in *cis* or *trans*; Huang *et al.* (2015) defined *cis*-eQTL as eQTL within ±1kb of the gene encoding their target transcript. Using this definition, it appears the vast majority of insecticide-associated eQTL and veQTL act in *trans*, with only 6% and 1.8% of insecticide-associated eQTL and veQTL respectively annotated as *cis*-acting (fig. 3.2.2B).

Given that these insights into *cis*- and *trans*-eQTL architecture hinge on the conservative definition of *cis*-eQTLs as occurring within 1kb of the target's coding region, these results were reinterpreted using a more permissive threshold of 10kb for *cis*-eQTLs (fig. 3.2.2C). Although this threshold increased the number of *cis*-eQTL relationships, the *trans*-eQTL relationships remained more abundant.

TWAS and eQTL mapping represent two distinct approaches for identifying transcript abundance differences associated with a phenotype (fig. 3.1.1), thus eQTL can be assessed for their ability to explain TWAS associations (fig. 3.2.2A). eQTL annotations identified 35 target transcripts, of which only nine were among the 170 unique insecticide-associated TWAS associations (table 3.2.3).

Table 3.2.3. Transcripts identified by GWAS-associated eQTL and veQTL

insecticide	transcripts identified by eQTL*	transcripts identified by veQTL*
azinphos-methyl	*CG34232; CG4734; CG7966;* ***Cyp6g1****;* ***Cyp6g2****;* ***Cyp6t3***	*CG11334; CG17707; CG7966; Dhc16F; Hsc70-2; rdgBbeta*
boric acid	-	-
chlorantraniliprole	***CG15083****; CG30098;* ***Cyp12d1-d****;* ***Cyp6a22****; LysP; p24-2; XLOC_004064; XLOC_004604*	*ade3; Apc7; CG11334; CG13743; CG17118; CG17691; CG32791; CG3921; CG5556; CG7768; Cnot4; Dhc16F; KCNQ; Klp98A; rdgBbeta; XLOC_000896*
DDT	-	*Dhc16F*
deltamethrin	*XLOC_000192;* ***Cyp6a23****; Cyp6a9;* ***rdgBbeta***	*CG7849;* ***Cyp6a23***
imidacloprid	-	-
malathion	*CG14372; CG14480; CG17691; CG31157; CG44142; CG7966; CR17025;* ***Cyp6g1****;* ***Cyp6g2****; Hsc70-2; Madm; XLOC_003905; XLOC_004064*	*CG7966; dik; Hsc70-2; XLOC_004064*
parathion	*CG10508; CG14372;* ***CG31157****; CG3699; CG44142; CG7966; CG8476; CG9799; CR17025; GstD1; Pgd; XLOC_003905; XLOC_006133*	*Ace*
permethrin	*Cyp6a17; iPLA2-VIA*	*CG11334; Cyp6a17*

*Transcripts in bold are also associated with the same insecticide through TWAS

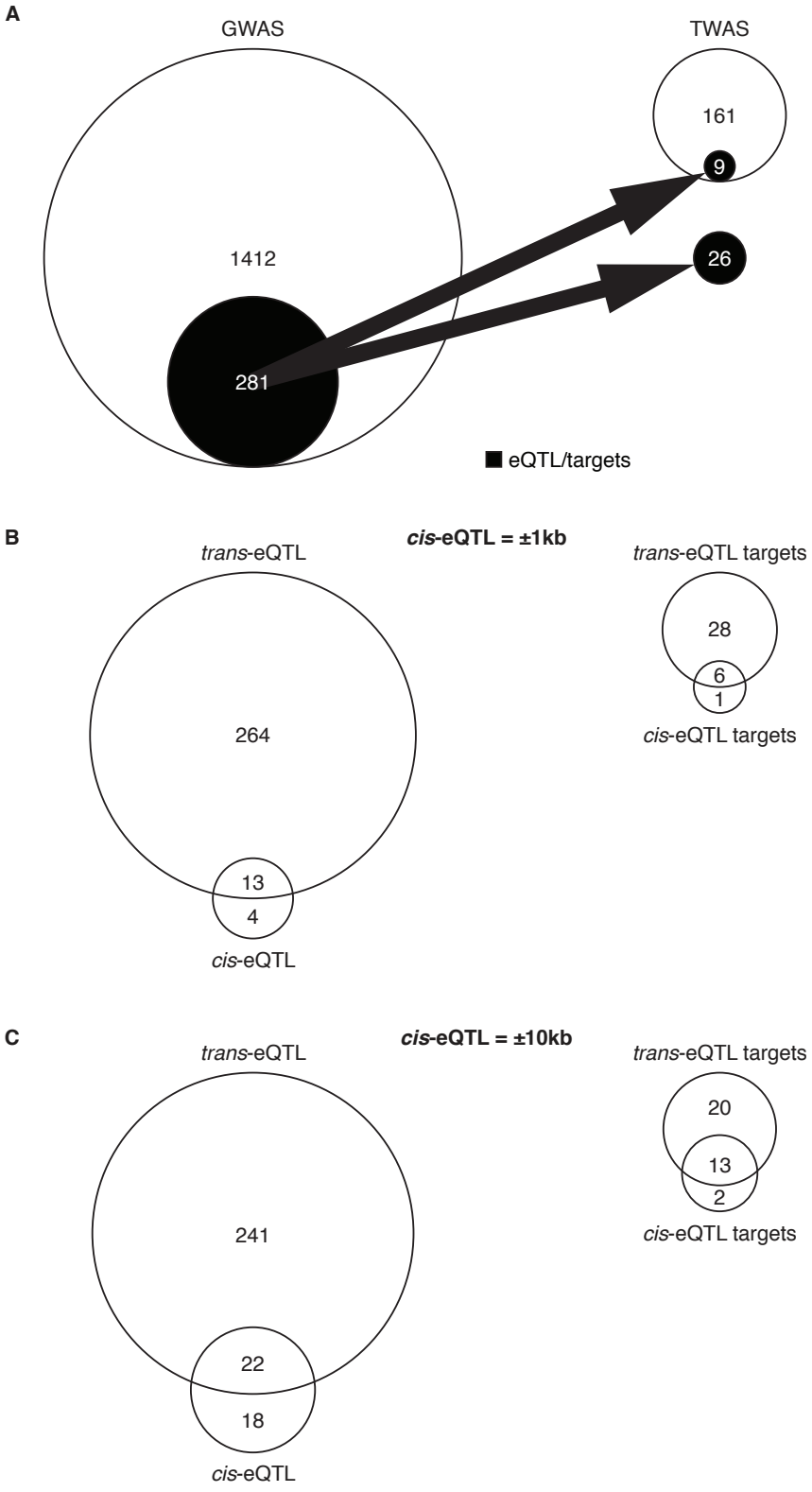


Figure 3.2.2. Relationships between insecticide association study candidates and eQTL

281 GWAS candidates are eQTL, which target nine TWAS candidates and an additional 26 transcripts not identified by TWAS (A). Both eQTL and their targets are predominantly *trans* relationships (B), even when the definition of *cis*-eQTL is expanded to targets 10kb from the eQTL (C).

3.2.1.4. Structural variation in genes encoding candidate transcripts

Structural variation, including gene duplications, deletions and transposable element insertions, has repeatedly been demonstrated to contribute to insecticide resistance by increasing transcriptional output, most commonly through the upregulation of genes involved in insecticide metabolism (e.g. Mouches *et al.* 1986; Field, Devonshire and Forde 1988; Daborn *et al.* 2001; Joußen *et al.* 2012). Complex structural variation is largely uncalled in DGRP genotypes, and thus its effect on transcript levels and phenotypes is underappreciated, including the large-effect *cis*-eQTL expected at major structural variants encompassing gene coding regions.

Previously, Zichner *et al.* (2013) quantified genome-wide structural variation in 38 DGRP lines by analysing discordant read mapping and aberrations in read depth among DGRP genome alignments. Additionally, a subset of calls were validated in PCR and microarray experiments. Structural variants identified by Zichner *et al.* (2013) that overlapped coding regions of TWAS candidates were scored in 205 DGRP genomes by manual analysis of paired-end read mapping. After filtering for a minimum minor allele frequency of 0.04 (reduced from the standard 0.05 to avoid excluding structural variation at *Cyp6g1*), there remained 13 unique structural variants in TWAS candidates (table 3.2.4). Additional variation in copy number was observed at the two repeat units present in derived *Cyp6g1* alleles (fig. 2.2.6). Copy number of each repeat unit in each DGRP line was inferred from read depth, ranging from one to five for the *Cyp6g1* repeat unit, and one to ten for the *Cyp6g1g2* repeat unit (fig. 3.2.3).

Structural variation is largely absent from DGRP genotypes, and hence not considered in Huang *et al.*'s (2015) eQTL dataset. To identify both *cis* and *trans* effects of these structural variants, eQTL mapping was performed by fitting linear models between each structural variant and each transcript level from the DGRP transcriptome. Utilizing a Bonferroni-corrected significance threshold ($p<2.76\times10^{-6}$), eleven of the 13 structural variants in TWAS candidates were eQTL of at least one transcript in at least one sex, and eight of these were *cis*-acting, affecting the TWAS candidate transcript used to identify them (table 3.2.4).

Table 3.2.4. Structural variants in TWAS-identified genes and their eQTL targets

structural variant type	TWAS association	gene CDS involved	DGRP frequency	male eQTL target	female eQTL target
tandem duplication (normalized read depth)	azinphos methyl; imidacloprid; malathion	*CG33964; Cyp6g1*	-	***Cyp6g1****(+)*	-
tandem duplication (normalized read depth)	azinphos methyl; imidacloprid; malathion	*Cyp6g1; Cyp6g2*	-	***Cyp6g1****(+);* ***Cyp6g2****(+); Cyp6t3(+)*	***Cyp6g1****(+);* ***Cyp6g2****(+)*
Accord LTR TE insertion	azinphos methyl; imidacloprid; malathion	none (*Cyp6g1* 5'UTR)	0.95	***Cyp6g1****(+); Cyp6g2(+); mir-966(+)*	***Cyp6g1****(+); Cyp6g2(+)*
HMS-Beagle TE insertion	azinphos methyl; imidacloprid; malathion	none (*Cyp6g1* 5'UTR)	0.25	*CG34232(+); tobi(+); CG6484(+)*	*CG34232(+); CG31104(-); tobi(+)*
tandem duplication	boric acid	*CG11872; CG6293*	0.24	***CG6293****(+); sle(+)*	***CG6293****(+); CG1129(+); Glycogenin(-)*
tandem duplication	chlorantraniliprole; malathion	*Cyp12d1-p*	0.24	***Cyp12d1-d****(+);* ***Cyp12d1-p****(+)*	***Cyp12d1-d****(+);* ***Cyp12d1-p****(+)*
deletion	deltamethrin	*Cyp6a17*	0.20	***Cyp6a17****(-);* ***Cyp6a23****(+); Cyp6a22(-)*	***Cyp6a17****(-);* ***Cyp6a23****(+)*
deletion	deltamethrin	*Cyp6a17; Cyp6a23*	0.13	***Cyp6a23****(+);* ***Cyp6a17****(-); rdgBbeta(+)*	***Cyp6a17****(-);* ***Cyp6a23****(+); Cyp6a22(-)*
tandem duplication	malathion	*CG12950; Pnn*	0.04	-	*CG30383(+); XLOC_006120(+)*
deletion	malathion	*CG6324*	0.77	-	-
Bari1 TE insertion	malathion	none (between *Jheh2* and *Jheh3*)	0.38	*MetRS(-)*	-
deletion	malathion	*SP1029; CG31427; CG31445*	0.19	-	-
tandem duplication	parathion	*CG7966; pic; CG31157; Hsc70-2*	0.17	***CG7966****(+);* ***CG31157****(+);* ***Hsc70-2****(+)*	***Hsc70-2****(+);* ***CG7966****(+);* ***CG31157****(+)*

**cis*-eQTL targets are indicated in bold

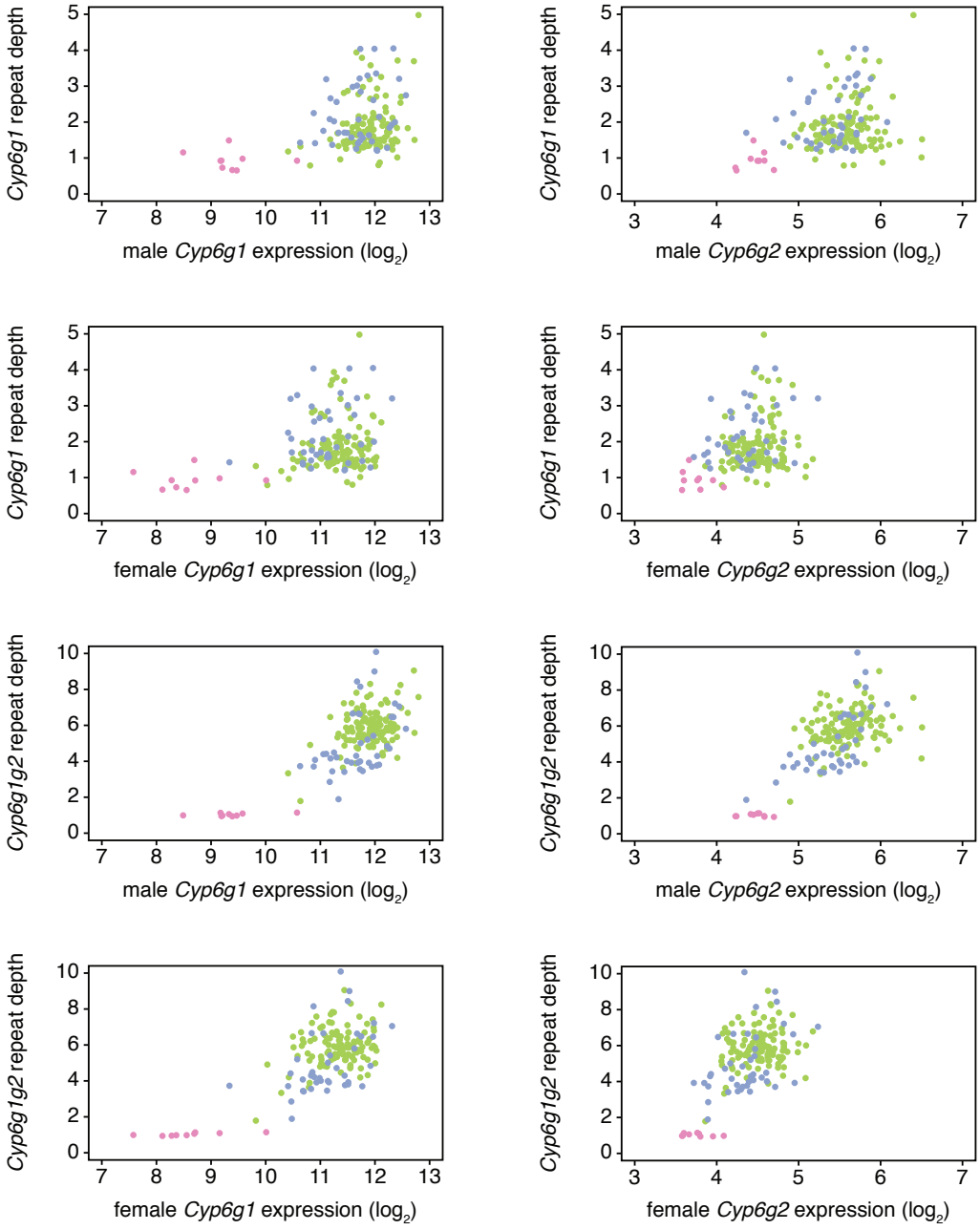


Figure 3.2.3. Repeat units involving *Cyp6g1* and *Cyp6g2* are *cis*-eQTL of both transcripts

Read depth of two repeat units involving *Cyp6g1* (fig. 2.2.6) normalized to a structurally stable control region, plotted against transcript levels of *Cyp6g1* and *Cyp6g2* for male and female adult DGRP lines. Colours represent transposable element insertion alleles (*Cyp6g1-M*: pink, *Cyp6g1-AA*: green, *Cyp6g1-BA*: blue).

The 13 structural variants were also tested for association against all 35 insecticide resistance phenotypes. At $p<1\times10^{-5}$, five structural variants were associated directly with insecticide resistance phenotypes (table 3.2.5; fig. 3.2.4-6).

Table 3.2.5. Structural variants in candidate genes associated with insecticide phenotypes

structural variant	insecticide phenotype association (p<1E-5)	p-value
Cyp6g1 Accord LTR insertion	azinphos-methyl 0.25µg/ml mortality	3.03E-33
	malathion female 3-hr mortality	2.61E-13
	azinphos-methyl 0.5µg/ml mortality	1.99E-12
	malathion male 3-hr mortality	4.17E-09
	azinphos-methyl 1µg/ml mortality	5.54E-08
Cyp6g1g2 repeat unit copy number	azinphos-methyl 0.25µg/ml mortality	6.31E-15
	azinphos-methyl 0.5µg/ml mortality	4.52E-10
	azinphos-methyl 1µg/ml mortality	6.67E-08
	malathion female 3-hr mortality	1.98E-07
	azinphos-methyl 2µg/ml mortality	2.08E-06
	azinphos-methyl LC50	8.99E-06
Cyp12d1 duplication	chlorantraniliprole LC50	4.93E-07
	chlorantraniliprole 1µg/ml mortality	9.75E-07
	chlorantraniliprole 4µg/ml mortality	1.32E-06
Cyp6a17/23 deletion	deltamethrin male 48-hr mortality	1.08E-24
Cyp6a17 deletion	permethrin female 3-hr knockdown	5.86E-06

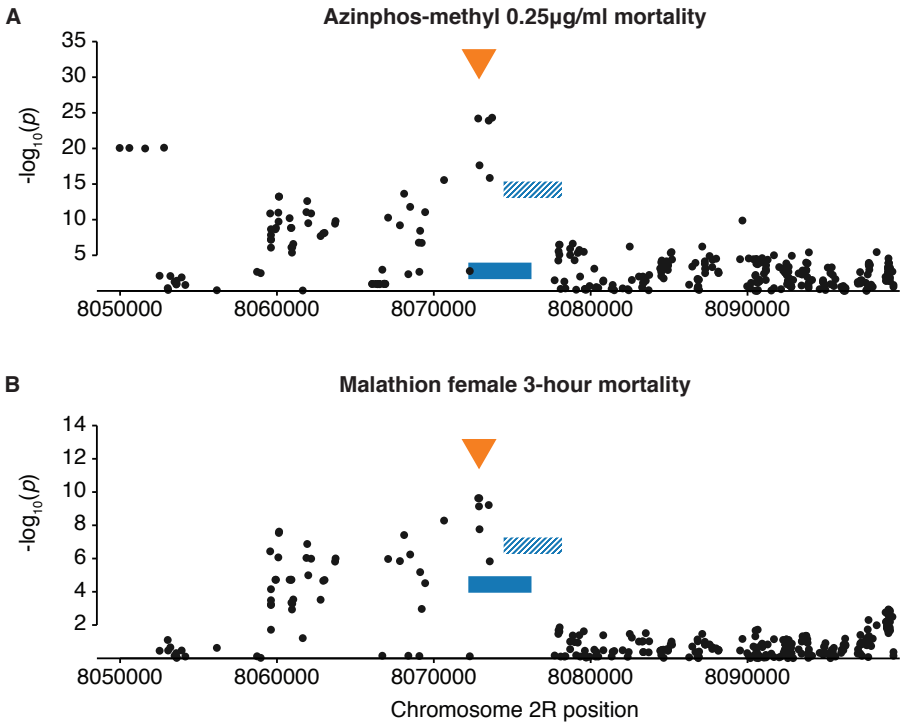


Figure 3.2.4. *Cyp6g1* structural variation associations with organophosphate phenotypes
Manhattan plots (mixed *p*-value against genomic location) for 100kb surrounding organophosphate-associated structural variants involving *Cyp6g1*. Structural variants are also plotted on the y-axis. Blue bars represent relative repeated units *Cyp6g1* (solid) and *Cyp6g1g2* (striped). The *Accord* LTR insertion into the 5'UTR of *Cyp6g1* is indicated by an orange triangle.

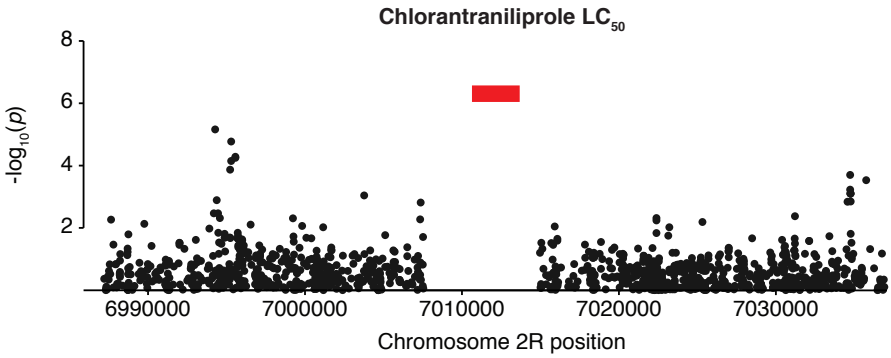


Figure 3.2.5. *Cyp12d1* structural variation association with chlorantraniliprole LC_{50}
Manhattan plot (mixed *p*-value against genomic location) for 100kb surrounding *Cyp12d1*. Structural variants are also plotted on the y-axis. The red bar indicates the *Cyp12d1* duplication (a deletion relative to the reference genome). Note that due to the similarity between the reference genome copies of *Cyp12d1* (*Cyp12d1-p* and *Cyp12d1-d*), genomic variants are not called in the duplicated region.

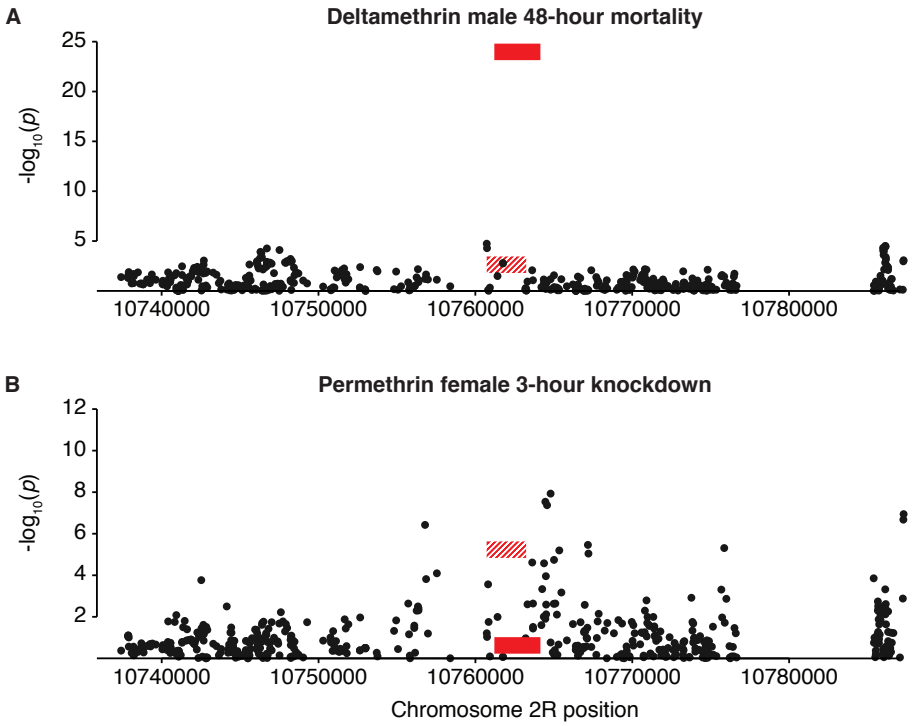


Figure 3.2.6. *Cyp6a17* and *Cyp6a23* structural variation associations with pyrethroid phenotypes
Manhattan plots (mixed *p*-value against genomic location) for 100kb surrounding pyrethroid-associated structural variants involving *Cyp6a17* and *Cyp6a23*. Structural variants are also plotted on the y-axis. Red bars indicate deletions relative to the reference genome *Cyp6a17/23* (solid) and *Cyp6a17*^{del} (striped).

3.2.1.5. Genetically correlated transcriptional modules

Huang *et al.* 2015 demonstrated that a large proportion of the male and female DGRP transcriptomes could be clustered into genetically correlated transcriptional modules (GCTMs; groups of transcripts in which the genetic component of abundance co-varies among DGRP lines; appendix 2). To assess if genetic variation in modular transcription plays a role in DGRP insecticide phenotypes, insecticide-associated transcripts were tested for membership in these modules. Chlorantraniliprole and malathion candidate lists were enriched for transcripts from male GCTM 79 (33 transcripts) and female GCTM 21 (662 transcripts; binomial $p<2\times10^{-6}$ in all cases; fig. 3.2.7; fig 3.2.8), and these associations accounted for four of the 13 transcript associations common to two or more insecticides (table 3.2.2). Among the insecticide-associated members of these modules, multiple transcripts are observed with levels altered by ectopic activation of *cncC* (Misra *et al.* 2011; fig. 3.2.7).

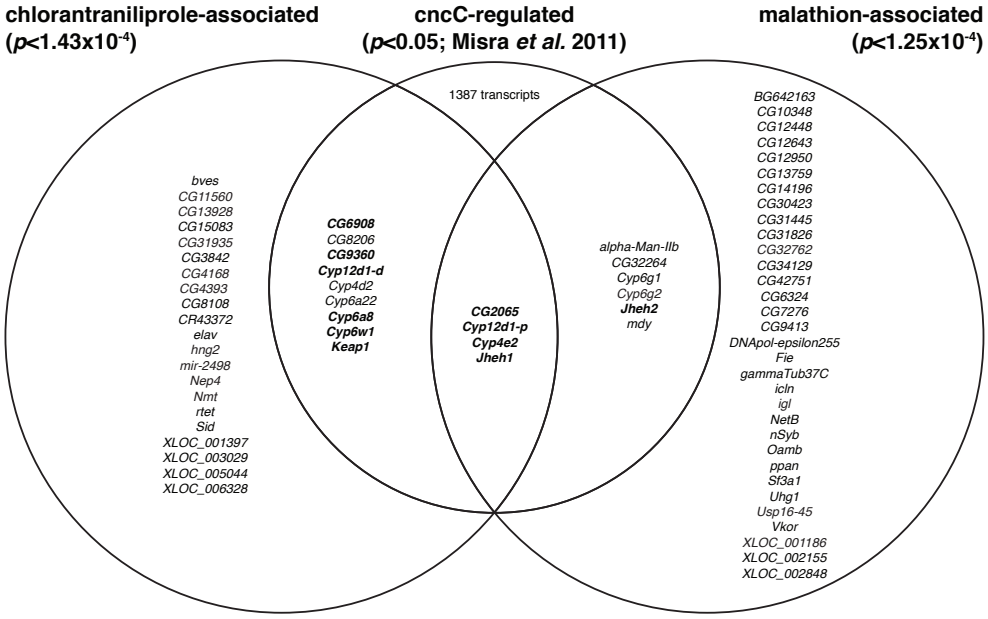


Figure 3.2.7. Intersection of transcripts associated with chlorantraniliprole and malathion phenotypes, and cncC-regulated transcripts identified by Misra *et al.* (2011)

Transcripts in bold indicate membership in male GCTM 79.

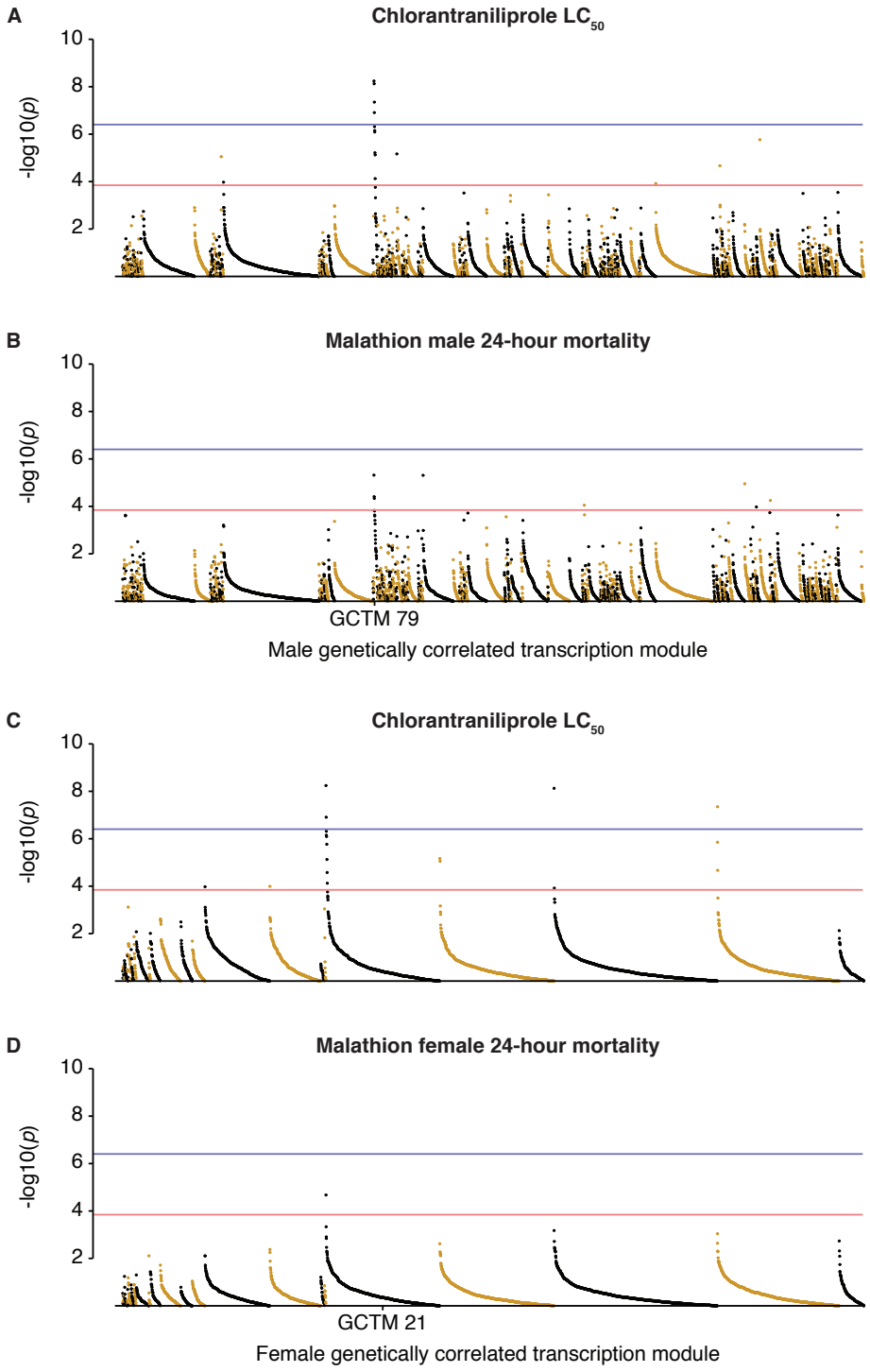


Figure 3.2.8. DGRP transcription module associations with two insecticide phenotypes

Opera house plots (*p*-value against transcript association, organized by Huang *et al.*'s [2015] GCTMs), for chlorantraniliprole LC_{50} (male [A] and female [C] transcriptome associations plotted separately), and malathion male (B) and female (D) 24-hour mortality. GCTMs are demarcated by alternating black and brown dots. The transcriptome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue.

3.2.2. Mapping eQTL of male GCTM 79

Male GCTM 79 and female GCTM 21, enriched in chlorantraniliprole and malathion transcript associations, contain multiple transcripts significantly differentially expressed following ectopic activation of *cncC* (Misra *et al.* 2011). Male GCTM 79 and female GCTM 21 contain 28 (of 33; 84.8%) and 164 (of 632; 25.9%) of such transcripts respectively. Constitutive upregulation of cncC-regulated transcripts has previously been described in resistant pest insect lines (Wilding 2018) and DDT-resistant *D. melanogaster* line *91-R* (Misra *et al.* 2013), however the genetic bases of these phenotypes remain unknown.

To identify novel eQTL regulating the modular transcription associated with chlorantraniliprole and malathion, the covariance among modular transcripts was specifically considered. Given that the male GCTM 79 is smaller and more enriched for cncC-regulated transcripts, this module was used to map modular associations. GCTM 79 comprises the levels of 33 transcripts from DGRP males. To ensure the biological relevance of the mapping, only the 28 transcripts from this module which were also differentially expressed following ectopic activation of *cncC* (Misra *et al.* 2011) were considered (fig. 3.2.9). The first principle component of these 28 transcript levels (explaining 36.52% of the variance) was calculated and used as a phenotype for a GWAS.

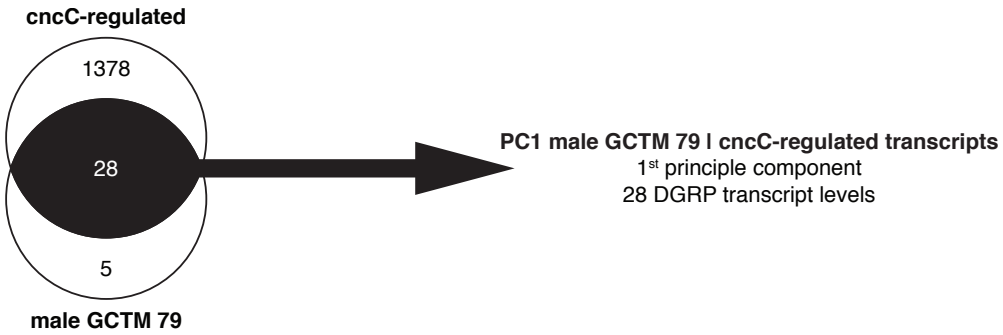


Figure 3.2.9. A phenotype for mapping eQTL of modular cncC-regulated transcripts
To identify candidate eQTL of the putative modular regulation of cncC-regulated transcripts in the DGRP, the first principle component of the overlap of transcripts known to be differentially regulated in response to ectopic cncC activation (Misra *et al.* 2011), and transcripts belonging to GCTM 79 (Huang *et al.* 2015), was used as a phenotype for a GWAS.

106 DGRP variants reached genome-wide significance ($p<1\times10^{-5}$; fig. 3.2.10; appendix 1.11), including four SNPs in introns of *Strn-Mlck*, which were also associated with chlorantraniliprole phenotypes (Green *et al.* 2019; appendix 1.3), a SNP in an intron of *CG10737*, a gene which has been demonstrated (through DGRP GWAS and transgenic experiments) to affect DDT-induced knockdown (Schmidt *et al.* 2017), and two SNPs in an intron of *Keap1*, a directly-interacting inhibitor of cncC. Furthermore, two variants achieved Bonferroni significance ($p<2.64\times10^{-8}$; fig. 3.2.11; appendix 1.11); an intergenic SNP 2,126bp upstream of *CG9922*, and a SNP causing a synonymous substitution in completely uncharacterized gene *CG5225*. These variants represent putative regulators of constitutive *cncC* levels among DGRP lines and make compelling candidates for future studies.

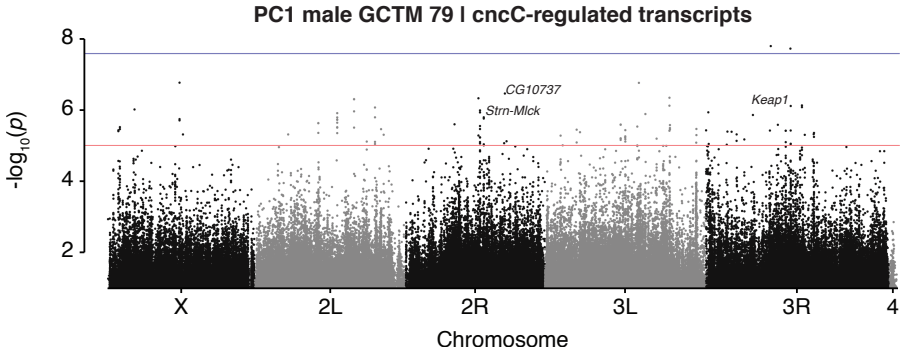


Figure 3.2.10. DGRP genomic variant associations with PC1 male GCTM 79 | cncC-regulated transcripts

Manhattan plot (*p*-value against genomic location) for the first principle component of transcript levels across DGRP lines from 28 transcripts in male GCTM 79 and identified as cncC-regulated by Misra *et al.* (2011). The genome-wide significance threshold is indicated in red, the Bonferroni significance threshold is indicated in blue.

3.2.3. Validation of *Cyp12d1* in chlorantraniliprole and malathion resistance

Increased expression of *Cyp12d1* has previously been implicated in resistance to insecticides. DDT-resistant lines have been demonstrated to overexpress the gene (e.g. *91-R*; Pedra *et al.* 2004), while transgenic manipulation of *Cyp12d1* suggests a potential role in resistance to DDT and dicyclanil (Daborn *et al.* 2007; Gellatly *et al.* 2015). However, natural variation in *Cyp12d1* has not previously been associated with insecticide resistance. In the DGRP, *Cyp12d1* transcript levels were significantly positively correlated with resistance to malathion in adult males at 24 hours, and with resistance to chlorantraniliprole in multiple larval phenotypes (table 3.2.2; appendix 2), but not with either DDT phenotype.

Associations between chlorantraniliprole and malathion phenotypes and *Cyp12d1* transcript levels were validated using GAL4-UAS overexpression of *Cyp12d1* at life stages matching the GWAS for each respective association (i.e. larval screens for chlorantraniliprole and adult screens for malathion). Larvae transgenically overexpressing *Cyp12d1* had a significantly higher (and ~220% greater) chlorantraniliprole LC_{50} than the control cross (fig. 3.2.11A), while adult flies of both sexes overexpressing *Cyp12d1* had significantly higher malathion LC_{50}s (~30% and ~20% increases for males and females respectively) than control crosses (fig. 3.2.11B).

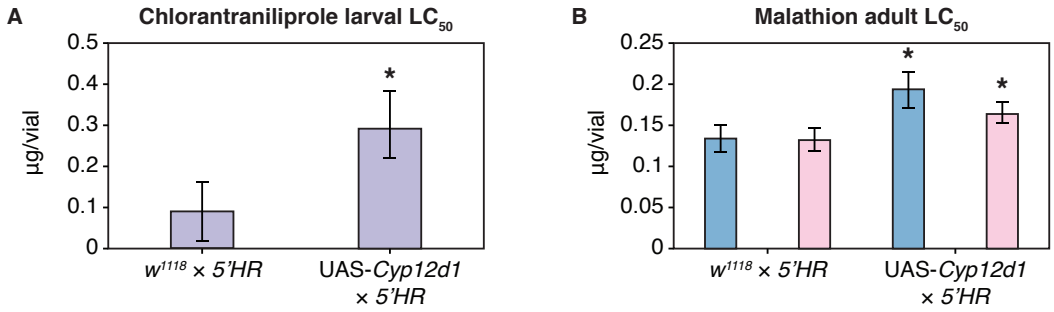


Figure 3.2.11. Functional validation of *Cyp12d1*

Insecticide LC_{50}s for larvae exposed to chlorantraniliprole (A) and male and female adults exposed to malathion (B). Error bars represent 95% confidence interval. Blue bars represent males and pink bars represent females. Transgenic overexpression of *Cyp12d1* in key metabolic tissues confers resistance to both chlorantraniliprole (A) and malathion (B).

3.3. Discussion

3.3.1. The quantitative transcriptomics of insecticide phenotypes

Here TWAS were performed on 35 phenotypes from nine insecticides measured using different exposure methods, at different life stages and collected by different research groups. A total of 170 unique significant transcript associations were identified across the 9 insecticides, of which only six (3.5%) were in genes implicated by GWAS from the same insecticide. For five of these six, putative *cis* regulation was supported by at least one *cis*-eQTL annotation. Inclusion of *trans*-eQTL explained an additional four transcript associations, and analysis of structural variation in genes identified by TWAS allowed a further four associations to be explained by *cis*-acting structural variants. In total of only 13 transcript associations (out of 170; 7.6%) could be linked to individual genomic variants (table 3.3.1); these represent the variants for which a causal regulatory role is most robustly inferred. However, the remaining GWAS variants which cannot be linked to TWAS-identified transcripts demonstrate the utility of TWAS approach to detect associations with genes not directly implicated by GWAS.

Table 3.3.1. eQTL-transcript networks associated with insecticide phenotypes

target transcript	eQTL type	eQTL annotation	insecticide
alpha-Man-IIb	*cis**	-	malathion
CG31157	*cis*†	-	parathion
CG6293	*cis*	-	boric acid
CG7966	*cis*	-	parathion
Cyp12d1-d	*cis*	-	chlorantraniliprole
Cyp6a23	*cis*	-	deltamethrin
Cyp6g1	*cis*†	-	azinphos-methyl; malathion
Cyp6g2	*cis*†	-	azinphos-methyl; malathion
Cyp6t3	*cis*†	-	azinphos-methyl
Hsc70-2	*cis*	-	parathion
CG15083	*trans*	*CG11566; stg1; unc*	chlorantraniliprole
Cyp12d1-d	*trans*	*sli*	chlorantraniliprole
Cyp6a22	*trans*	*Asx*	chlorantraniliprole
Cyp6a23	*trans*	intergenic	deltamethrin
rdgBbeta	*trans*	intergenic	deltamethrin

*GWAS association not annotated as eQTL

†*trans*-eQTL in selective sweep regions which included *cis*-eQTL for the same transcript were excluded as artefacts of linkage disequilibrium

Of the 1693 unique insecticide-associated genomic variants considered in this chapter, only 95 were changes to protein structure (i.e. nonsynonymous substitutions or introduction of start or stop codons [80, ten and five respectively]), predicting the vast majority of GWAS associations to be regulatory. However, only a modest proportion of GWAS associations (281; fig. 3.2.2) were annotated to be eQTL by Huang *et al.* (2015), and of these, most (264; fig. 3.2.2B) act only in *trans*. In contrast, links between genomic and transcriptomic associations (table 3.3.1) are enriched for *cis*-acting variants, however this reflects the bias in power to detect simple regulatory networks; identification of complex trans-eQTL networks requires enormous sample sizes (Liu, Li and Pritchard 2019). These results suggest that the bulk of insecticide GWAS associations are *trans*-regulatory variants with diffuse effects, which is consistent with empirical estimates of *cis* versus *trans* heritability of gene expression from *D. melanogaster* (McManus *et al.* 2010), and meta-analyses of human traits (Liu, Li and Pritchard 2019).

3.3.2. Modular transcription linked to cncC regulation

One strategy to identify *trans*-regulatory networks is to utilize the property that transcript expression levels tend to cluster into genetically correlated transcription modules (GCTM; Huang *et al.* 2015). Annotation of insecticide resistance-associated transcripts with genetically correlated transcription modules (GCTMs) reveals an enrichment for module 79 (male) and module 21 (female), due to module members' strong associations with chlorantraniliprole and malathion (appendix 2; fig. 3.2.8). Associated members of these modules resemble transcripts with levels altered by ectopic activation of *cncC* (Misra *et al.* 2011; fig. 3.2.7). *cncC* and its mammalian homolog *Nrf2* encode CNC-bZIP transcription factors that play a well-conserved role in oxidative stress response (Wilding 2018). In unstressed conditions, cncC is inhibited and restricted to the cytosol by actin-associated Keap1, which can also target cncC for ubiquitination and proteasomal degradation (Sykiotis and Bohmann 2008). Under cellular stress, cysteine residues in Keap1 react to chemical inducers, which results in Keap1 detaching from cncC (Eggler *et al.* 2005). This enables cncC to move from the cytosol to the nucleus, activating oxidative stress-response genes (fig. 3.3.1).

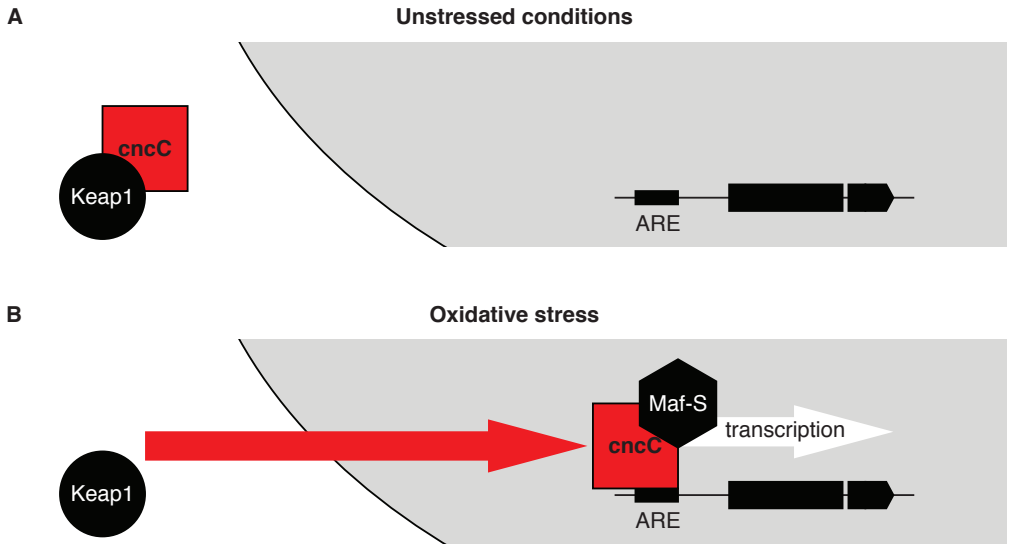


Figure 3.3.1. Oxidative stress response regulation by cncC and Keap1 (after Wilding 2018)
Under normal conditions (A) cncC is sequestered in the cytoplasm by Keap1. Under oxidative stress (B), Keap1 releases cncC which translocates to the nucleus. In the nucleus cncC forms a heterodimer with Maf-S and binds antioxidant response elements (AREs) to initiate transcription of oxidative stress response genes.

While most studies characterize this pathway in terms of its induction in response to oxidative stress, the DGRP transcriptome data comes from unexposed flies, suggesting natural variation among DGRP lines in the baseline expression level of cncC-regulated transcripts is associated with insecticide resistance. Evidence from DGRP transcriptomes implies that the constitutive level of cncC-regulated expression varies among lines. Mapping of genomic variants associated with variation in modular transcription of cncC-regulated transcripts identified multiple candidate genes (fig. 3.2.10; appendix 1.11) including *Keap1*, the product of which directly interacts with cncC. Any variation in *Keap1* affecting its ability to inhibit cncC would be expected to contribute to variation in levels of cncC-activated transcripts. Other candidate genes from this analysis include DGRP insecticide-associated genes *Strn-Mlck* and *CG10737*. Both genes are strong candidates from chlorantraniliprole and DDT GWAS respectively and were validated in insecticide phenotypes using RNAi knockdown (Green *et al.* 2019; Schmidt *et al.* 2017). Furthermore, both genes contain calcium-dependent domains and are implicated in muscle biology, however a mechanistic link between either gene and cncC or the oxidative stress response remains to be determined.

The constitutively upregulated cncC pathway transcripts associated with chlorantraniliprole or malathion phenotypes comprise potential detoxification enzymes including eight cytochrome P450s (*Cyp12d1*, *Cyp4d2, Cyp4e2, Cyp6a22, Cyp6a8, Cyp6g1, Cyp6g2* and *Cyp6w1*), two oxidoreductases (*CG2065* and *CG9360*) and two juvenile hormone epoxide hydrolases (*Jheh1* and *Jheh2*). Guio *et al.* (2014) demonstrated that the insertion of a *Bari1* transposable element upstream of *Jheh1* and *Jheh2* introduces additional cncC binding motifs. This results in increased inducibility both transcripts in response to oxidative stress, which they found also increased resistance to malathion. In this work, constitutive transcript levels of *Jheh1* and *Jheh2* were associated with male malathion mortality at 24 hours and 3 hours respectively, however presence of the *Bari1* insertion was not strongly associated with resistance phenotypes for malathion (or any insecticide). As the ten lines used by Guio *et al.* (2014) were also DGRP lines, the difference in *Bari1* insertion association between the two studies presumably reflects the differences between the malathion exposure times measured in this work (up to 24 hours) and by Guio *et al.* (2014; up to 214 hours). It makes sense that in the more acute assays in this study, the baseline expression level of

Jheh1 and *Jheh2* would be important, whereas over longer assay periods induction capacity would play a more relevant role.

3.3.3. Structural-variant eQTL

Structural variation is known to have large effects on transcript abundance, and therefore it is not surprising that the interrogation of structural variation in TWAS candidates allowed the identification of multiple novel eQTL (table 3.2.4). Of particular interest are structural-variant *cis*-eQTL of five cytochrome P450 genes which showed robust associations with insecticide phenotypes directly, as well as through their target transcripts (table 3.2.5).

3.3.3.1. *Cyp12d1*

Heading lists of insecticide-associated, cncC-activated candidates, *Cyp12d1* exists as a 3935bp duplication in the *y; cn bw sp;* reference genome, with copies annotated as *Cyp12d1-p* and *Cyp12d1-d*. While both *Cyp12d1-p* and *Cyp12d1-d* transcript levels are significantly associated with chlorantraniliprole phenotypes, only *Cyp12d1-p* is significantly associated with any malathion phenotype. These results are supported by the findings that overexpression of a transgene derived from *Cyp12d1-p* (Daborn *et al.* 2007) in key metabolic tissues increases both chlorantraniliprole and malathion resistance (fig. 3.2.11), however the differing associations of the two insecticides with *Cyp12d1-d* remains unresolved. One explanation is that differences between gene copies may differentiate their products at the level of either enzyme specificity or expression profile: In the reference genome, *Cyp12d1-p* and *Cyp12d1-d* are differentiated by three amino acids and have 3' UTRs of different length (Boey 2011). Variation in the *Cyp12d1* duplicated region among DGRP lines is unknown. Due to paired-end reads mapping to both *y; cn bw sp;* reference genome copies, no variants were called in the region by DGRP genotyping (fig. 3.2.5).

An alternative explanation to differentiation between *Cyp12d1* copies, *Cyp12d1*'s role in malathion resistance may be comparatively minor (although transgenically validated), and the malathion association with *Cyp12d1-p* may have been inflated by its strong correlation with cncC-regulated transcripts, amongst which may exist a transcript with a greater effect. *Cyp12d1* is identified in chlorantraniliprole GWAS by a single, downstream SNP in linkage disequilibrium with the duplication (r^2=0.63). The addition of eQTL mapping, TWAS associations and structural variation data allowed the identification of *Cyp12d1* as a demonstrable contributor of resistance to two insecticides

through transcript abundance, a result of both *trans*-acting genetic variation in constitutive modular expression, and *cis*-acting copy number variation.

3.3.3.2. *Cyp6g1* and *Cyp6g2*

Increased *Cyp6g1* transcript level was associated with resistance to azinphos-methyl, imidacloprid and malathion phenotypes. Multiple variants associated with azinphos-methyl and malathion in GWAS were eQTL of *Cyp6g1*, consistent with the premise that the selective sweep surrounding *Cyp6g1* is driven by increased product of the gene. Interestingly, many eQTL of *Cyp6g1* are also eQTL of its tandem paralog, *Cyp6g2*, suggesting the transcripts are co-regulated. This hypothesis is bolstered by the mapping of eQTL using four *Cyp6g1* structural variation genotypes; copy number of two repeat unit regions (estimated by normalized sequence read depth) and presence of two transposable elements in *Cyp6g1*'s 5'UTR (fig. 2.2.6; table 3.2.4). The *Accord* LTR insertion, common to all known resistance alleles of *Cyp6g1* (Schmidt *et al.* 2010), is an eQTL of both *Cyp6g1* and *Cyp6g2*, as is the amplification involving partial copies of *Cyp6g1* and *Cyp6g2* (fig. 3.2.3). The amplification involving only *Cyp6g1* coding region shows weaker associations with both transcripts, with only *Cyp6g1* transcript level reaching significance as an eQTL target in males. The *HMS-Beagle* transposable element insertion is not correlated with either *Cyp6g1* or *Cyp6g2* transcript level, congruent with previous studies indicating that the effects of this transposable element insertion are both transcriptomically and phenotypically subtle (Schmidt *et al.* 2010). The *HMS-Beagle* insertion is instead an eQTL of four transcripts (table 3.2.4), which are extremely highly expressed in the midgut (according to FlyAtlas data; Chintapalli *et al.* 2007), a tissue in which the *HMS-Beagle* insertion has been demonstrated to significantly increase *Cyp6g1* transcript (Schmidt *et al.* 2010). Overall these results reiterate the findings from chapter 2 and previous studies (Daborn *et al.* 2002; Schmidt *et al.* 2010) that naturally occurring alleles at *Cyp6g1* confer insecticide resistance through overexpression of the enzyme. Moreover, the transcript level of *Cyp6g1*'s tandem paralog, *Cyp6g2* is also upregulated in these alleles. Although TWAS associations with *Cyp6g2* are weaker than those with *Cyp6g1*, transgenic overexpression of *Cyp6g2* has been demonstrated to confer resistance to diazinon, nitenpyram and imidacloprid (Daborn *et al.* 2007; Denecke *et al.* 2017).

3.3.3.3. *Cyp6a17* and *Cyp6a23*

Multiple lines of evidence (including GWAS, TWAS, eQTL and veQTL) link DGRP phenotypes from two pyrethroid-class insecticides (deltamethrin and permethrin) with a pair of P450s, *Cyp6a17* and *Cyp6a23*. Two structural variants of these genes are present in the DGRP, both forming a single chimeric gene comprising differing amounts of *Cyp6a17* and *Cyp6a23* coding sequence (fig 2.2.9). Intensity levels of probes mapping to *Cyp6a17* and *Cyp6a23* are potentially unreliable in measuring transcription levels of the chimeric alleles. Nevertheless, eQTL and veQTL of *Cyp6a23* are associated with deltamethrin (appendix 1.5), as is *Cyp6a23* transcript level (table 3.2.2; appendix 2.5), and eQTL and veQTL of *Cyp6a17* are associated with permethrin (appendix 1). By probing associations at the level of the structural variant alleles, it is apparent that the *Cyp6a17/23* allele is associated with *increased* resistance to deltamethrin, while the *Cyp6a17* deletion allele is associated with *decreased* resistance to permethrin (fig 3.3.2). Furthermore, gene disruption and RNAi knockdown experiments demonstrate that reduction in the expression of either *Cyp6a17* or *Cyp6a23* results in reduced pyrethroid susceptibility (fig. 2.2.12; Duneau *et al.* 2018).

Taken together, these results suggest that either differences in the coding sequences of the derived structural variants, or upstream regulatory elements (which are inherited from *Cyp6a17* in the *Cyp6a17/23* allele and from *Cyp6a23* in the *Cyp6a17*^{del} allele) contribute to a deltamethrin resistance allele in *Cyp6a17/23* and a permethrin susceptibility allele in *Cyp6a17*^{del}. This finding not only underscores the importance of structural variation in explaining phenotypic variation, but also demonstrates significant allelic heterogeneity among derived structural variants at a single locus. It also raises the intriguing possibility of chimeric genes driving phenotypic diversification (e.g. Joußen *et al.* 2012).

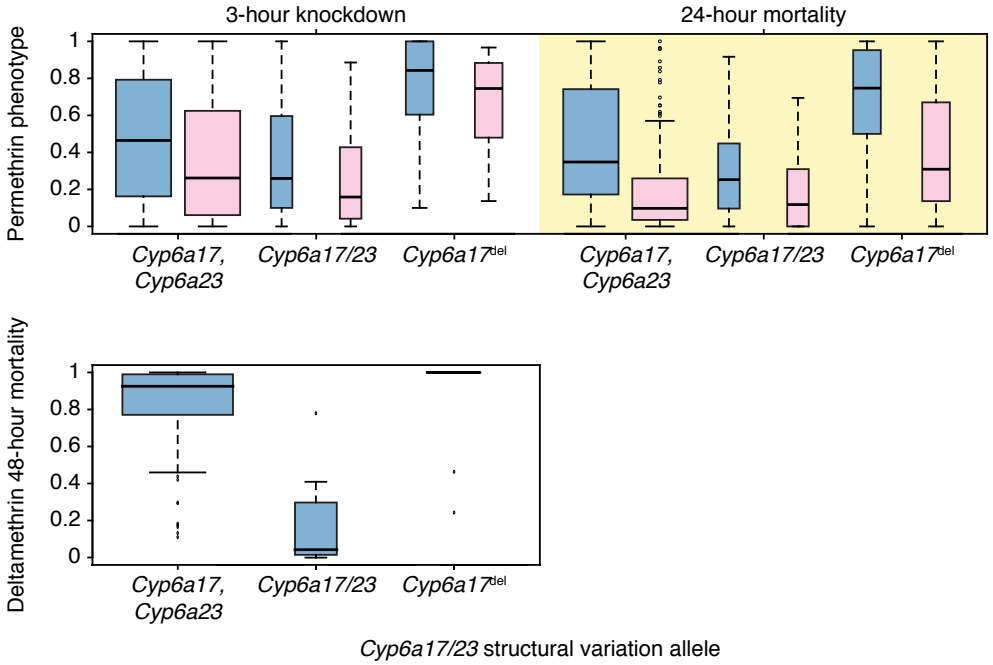


Figure 3.3.2. Influence of structural variation at *Cyp6a17* and *Cyp6a23* on deltamethrin and permethrin phenotypes

Distributions of DGRP deltamethrin and permethrin phenotypes partitioned by structural variation at *Cyp6a17* and *Cyp6a23*. Blue bars represent males and pink bars represent females. Boxes indicate the 25 to 75 percentile range, whiskers indicate the nine to 91 percentile range.

3.3.4. Conclusions

In this chapter, the incorporation of transcriptome data into DGRP association studies of phenotypes from nine insecticides has not only greatly increased the number of candidate genes identified, but allowed the characterization of mechanisms by which top candidates contribute to insecticide resistance. Five cytochrome P450 genes with a total of at least seven structurally variant alleles contribute to multiple DGRP insecticide phenotypes through *cis*-acting effects on transcript abundance. However, despite the presence of large-effect *cis*-eQTL, the majority of genetic variation associated with insecticide resistance appears have more subtle effects on *trans*-regulatory networks. Genetically correlated transcripts allowed the further characterization of one such network, implicating variation in the constitutive levels of an oxidative stress response module in resistance to multiple insecticides.

3.4. Materials and methods

3.4.1. Fly lines

The UAS-*Cyp12d1* line was generated by Daborn *et al.* (2007) and the *6g1*HR-GAL4 driver line was generated by Chung *et al.* (2007). Both lines were obtained from the Batterham Lab at the University of Melbourne.

3.4.2. Phenotype data

35 raw DGRP phenotypes from 9 insecticides (table 3.2.1) were obtained from §2.2.1.1, §2.2.1.2, Najarro *et al.* 2017, Green *et al.* 2019, Schmidt *et al.* (2017), Duneau *et al.* (2018) and Denecke *et al.* (2017). GWAS for eQTL and veQTL analyses were performed under standardized conditions using the Mackay lab DGRP2 GWAS pipeline (`http://dgrp2.gnets.ncsu.edu/`; Huang *et al.* 2014), described in §2.4.6 and §4.4.1.

3.4.3. Transcriptome data

Transcriptome data for 1-3 day old adult flies from 185 DGRP lines were recovered from the DGRP website (`http://dgrp2.gnets.ncsu.edu/data.html`; Huang *et al.* 2015). Mean transcription level was calculated for each gene in each sex from two biological replicates, to give a mean level for each of the 18,139 transcripts measured by Huang *et al.* (2015) in each DGRP line, for both males and females. eQTL, veQTL and genetically correlated transcription module data was generated by Huang *et al.* (2015; tables S7, S11*, S12*, S15, S16).

*corrected versions obtained from `http://dgrp2.gnets.ncsu.edu/data.html`

3.4.4. Transcriptome-wide association studies

A linear model was fit between mean transcription level of each gene measured by Huang *et al.* (2015) and each insecticide phenotype using the `lm` function in R (R Core Team 2018). In each case, the transcriptome from the relevant sex was used; for larval phenotypes where sex was not discriminated, the mean of male and female levels for each transcript was taken. 1×10^{-3} (which roughly corresponds to the genome-wide significance threshold used in GWAS adjusted for the smaller number of tests performed against the transcriptome compared to the genome) was used as a base significance threshold for transcriptome associations after correction for the number of phenotypes tested for each insecticide (ranging from one for boric acid and deltamethrin

to eight for malathion). Bonferroni significance thresholds were calculated as 0.05 divided by the product of the number of transcripts (18,139) and phenotypes tested.

3.4.5. Mapping eQTL of male GCTM 79

The first principle component of transcript levels from the intersect set of male GCTM 79 (Huang *et al.* 2015) and transcripts differentially expressed following ectopic activation of *cncC* (Misra *et al.* 2011) was calculated for 185 DGRP lines using the `pca` command from the 'FactoMineR' package (Lê, Josse and Husson 2008) in R (R Core Team 2018). This was then submitted as a phenotype file to the Mackay lab DGRP2 GWAS pipeline (`http://dgrp2.gnets.ncsu.edu/`; Huang *et al.* 2014).

3.4.6. Structural variation in TWAS candidates

Structural variants identified by Zichner *et al.* (2013) in 38 DGRP lines were filtered for variants affecting coding regions of all TWAS candidates. BAM files containing alignments of DGRP line sequences from Illumina platforms to the *y; cn bw sp;* reference genome were recovered from the Baylor College of Medicine website (`https://www.hgsc.bcm.edu/content/dgrp-lines`; Mackay *et al.* 2012). Local alignments at TWAS candidates were visualized with IGV software (Thorvaldsdóttir, Robinson, and Mesirov 2013) to manually score structural variants identified by Zichner *et al.* (2013) in 205 DGRP genomes. After filtering for minor allele frequency >0.04, associations of structural variants with the DGRP transcriptome and insecticide phenotypes were measured by fitting linear models using the `lm` function in R (R Core Team 2013).

Amplification events involving *Cyp6g1* and *Cyp6g2* were inferred from local read depth in DGRP BAM files. Read depth at each nucleotide position under consideration were recovered using the Genome Analysis Toolkit 'DepthOfCoverage' utility (McKenna *et al.* 2010). Regions interrogated were non-overlapping portions of the *Cyp6g1* repeat unit (`2R:8072727-8074976`), the *Cyp6g1g2* repeat unit (`2R:8075688-8077656`), and a control region of similar size just upstream of *Cyp6g1* which does not exhibit structural variation in the DGRP (`2R:8070657-8072656`). Mean read depth for each repeat unit was calculated for each DGRP line, and normalized to mean read depth of the control region.

3.4.7. Transgenic overexpression of *Cyp12d1*

Cyp12d1 was overexpressed using the GAL4/UAS system (Brand and Perrimon 1993) and the *6g1*HR-GAL4 driver described by Chung *et al.* (2007). *6g1*HR-GAL4 virgin females, in which GAL4 is regulated by *Cyp6g1* upstream sequence originating from Hikone-R line flies, were crossed to males carrying an additional copy of *Cyp12d1* under control of a UAS promoter. w^{1118} was used in crosses as a control for the UAS-*Cyp12d1* line (Daborn *et al.* 2007). F1 progeny were exposed to chlorantraniliprole and malathion using bioassays matching those used in DGRP phenotyping for each compound (chlorantraniliprole larval survival to adulthood [Green *et al.* 2019.]; malathion male and female mortality at 24 hours [§2.2.1.2]). LC_{50}s were calculated using the method described in §2.4.4.

4

A meta-analysis of insecticide resistance in the Drosophila Genetic Reference Panel

A powerful and often underutilized aspect of mapping panels is that phenotypes within the panel are directly comparable, and accumulate over time as the panel is studied. As described in chapter 3, phenotypes for nine insecticidal compounds have been measured among the lines of the DGRP. Although results from GWAS of these phenotypes have been studied individually, a systematic comparative analysis of these phenotypes has yet to be performed. In this chapter, 35 phenotypes from assays of nine insecticides, reanalyzed under standardized GWAS parameters, are assessed for correlations and cross-resistance-associated variants. The evidence for positive selection at insecticide-associated loci is evaluated, and finally, the relevance of insecticide-associated DGRP variants to insecticide resistance more broadly is quantified by comparisons with worldwide population samples of *D. melanogaster*.

4.1. Introduction

4.1.1. The DGRP as a population genomics resource

Despite its primary role as a trait-mapping panel, the DGRP was designed to allow it to function as a resource for population genomics. Progenitor lines were sampled from a wild population within the same year (2003), and each line was derived from inbreeding the progeny of a single, gravid female over 20 generations (Mackay *et al.* 2012). Although this process purged alleles deleterious when homozygous along with more than 90 percent of the heterozygosity in each line, the DGRP's design preserved the bulk of common alleles and their frequencies in a living population sample. The trade-off in preserving the allele frequency spectrum of the DGRP's progenitor population is that low-frequency alleles, which cannot be utilized in standard association study designs, comprise over 50 percent of the genetic variants in the panel. This in contrast to other mapping panels (e.g. the Drosophila Synthetic Population Resource; King, Macdonald and Long 2012) which minimize genetic variation and incorporate crossing designs to maximize the ability to attribute phenotypic effects to all genetic variants.

One of the questions central to the study of population genomics is the inference of evolutionary history from signatures in a population's genomes. As one of *D. melanogaster*'s first population genomics samples comprising resequenced individuals, the DGRP has been utilized in inferring genomic regions targeted by selection, including analysis by Garud *et al.* (2015) using the H12 statistic, powered to detect recent selective events in the form of both 'hard' and 'soft' selective sweeps. Under the hard sweep model, a single haplotype bearing a beneficial allele rises in frequency, leaving a footprint of reduced diversity in linkage disequilibrium with the sweep's focal variant. Soft sweeps follow a similar trajectory, however instead of the beneficial allele initially existing on a single haplotype, it exists on multiple haplotypes. This is due either to multiple origins of the allele (Pennings and Hermisson 2006), or the allele being sufficiently ancient and common as to appear in multiple haplotypes in the population before the initiation of the sweep (Hermisson and Pennings 2005; fig. 4.1.1).

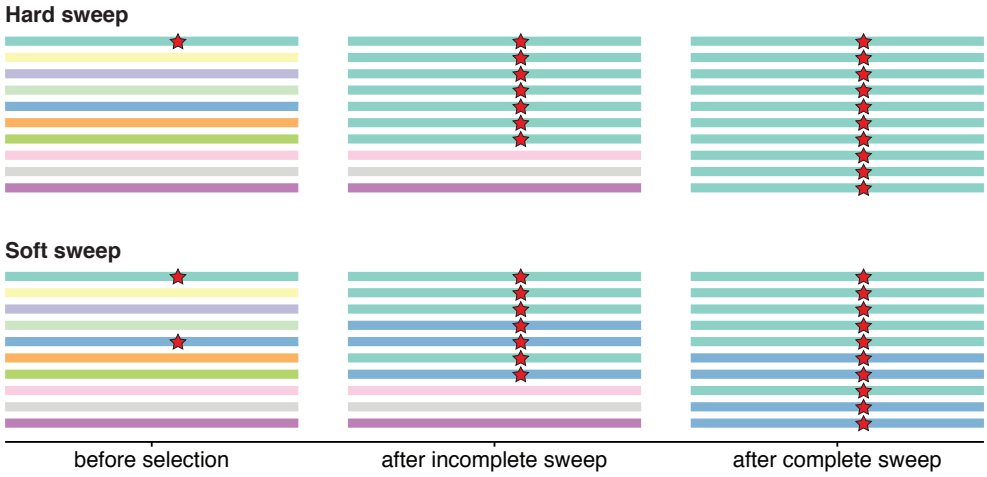


Figure 4.1.1. Models of haplotype structure in selective sweeps

Two models of haplotype structure produced by selective sweeps shown before selection, after an incomplete sweep and a complete sweep. Distinct haplotypes are represented by coloured bars, with the positively selected mutation represented by red stars.

4.1.2. Signatures of selection in the DGRP

Garud *et al.*'s (2015) analysis identified 50 strong candidates for recent selection in the DGRP's progenitor population, the three most pronounced of which included loci previously implicated in resistance to insecticides (*Ace* [Mutero *et al.* 1994], *CHKov1* [Aminetzach, Macpherson and Petrov 2005] and *Cyp6g1* [Daborn *et al.* 2001]). Subsequent analysis at *Ace* and *Cyp6g1* in the DGRP has largely supported the involvement of these loci in insecticide resistance, although at the level of specific compounds significant nuance has been observed (§2.3.2; Schmidt *et al.* 2017). In contrast, evidence suggests that the sweep including *CHKov1* is more likely due to selection for Sigma virus resistance (Magwire *et al.* 2011) than insecticide resistance (§2.2.2.3). These findings highlight the insight to be gained by validating alleles hypothesized to be driving selective sweeps in the population background, rather than relying solely on annotations of gene function or transgenic manipulation of candidates. The DGRP, as a trait mapping panel and populations genomics resource, provides an unparalleled resource for such experiments.

To date, 35 phenotypes measuring the effect of nine insecticidal compounds have been quantified across the lines of the DGRP, providing a powerful dataset for elucidating the specific selective forces imposed on the DGRP progenitor population by insecticides. Given that insecticide resistance loci have been proposed to explain two of the three most pronounced signatures of selection among DGRP lines (Garud *et al.* 2015; Duneau *et al.* 2018; §2.3), it is of interest to assess what proportion of all insecticide-associated variation in the DGRP lies within selective sweeps, and furthermore ask whether candidate alleles in these regions show patterns consistent with being the drivers of strong, recent selection.

4.1.3. Phenomics

A further dimension of the DGRP resource is the accumulating dataset of phenotypes measured across the mapping panel. Although this dataset includes hundreds of phenotypes from over 50 studies (as of late 2017; Anholt and Mackay 2018), the full potential of these data is seldom utilized in DGRP publications. Leveraging the DGRP 'phenome' promises to reveal the connectivity between phenotypes in this mapping panel, as well as allowing the identification of pleiotropic variants, a phenomenon long described in studies of insecticide resistance in the form of cross-resistance (e.g. Kikkawa 1961), as well as in the fitness cost of resistance alleles (reviewed in Kliot and Ghanim 2012).

Results describing cross-resistance patterns have important applications for insecticide usage. Insecticides are classified by their target site, and hence mutations modifying target proteins generally confers cross-resistance within an insecticide class, and to any class that shares the target (e.g. Pittendrigh *et al.* 1997; Menozzi *et al.* 2004). However, class has proven to be a less reliable predictor of cross-resistance mediated by metabolic enzymes (e.g. Daborn *et al.* 2007). This complicates predictions, based on previous insecticide exposure, as to which compounds a population may be preadapted to. Fitness costs associated with insecticide resistance alleles have been widely reported and likewise have implications for resistance management as well as providing insights into the endogenous functions of genes associated with insecticide phenotypes. However, despite abundant observations of both phenomena, neither fitness costs nor pleiotropy of insecticide resistance have previously been assessed with quantitative genetics approaches.

4.1.4. Beyond the DGRP

Since 2012 population genomics in *D. melanogaster* has intensified, producing multiple population sequence datasets to compliment the DGRP (Pool *et al.* 2012; Langley *et al.* 2012; Campo *et al.* 2013; Kao *et al.* 2015; Bergman and Haddrill 2015; Grenier *et al.* 2015; Lack *et al.* 2015; Kapun *et al.* 2018). The characterization of these datasets has revealed geographic structure between populations and higher genetic diversity in all sub-Saharan African populations (Lack *et al.* 2015). Furthermore, American populations, including the DGRP, contain African and European admixture, consistent with the higher genetic diversity observed in these populations relative to those from Europe (Lack *et al.* 2015; Mateo, Rech and González 2018). Parallel analysis of selective sweep peaks in Zambian and DGRP populations (using the H12 statistic; Garud and Petrov 2016) revealed only one peak (containing *Cyp6g1*) was shared between the top 25 peaks from each population.

In light of the apparent diversity between worldwide populations of *D. melanogaster*, these population datasets provide both the opportunity, and the imperative, to assess the relevance of genetic variants identified in the DGRP at a worldwide scale. Such an analysis using the 35 insecticide phenotypes considered in this chapter will provide a measure of the population specificity of insecticide-associated variation, and assess the utility of the DGRP for modelling associations in other *D. melanogaster* populations the world over.

4.2. Results

4.2.1. A comparison of DGRP insecticide phenotypes

4.2.1.1. Genome-wide association studies

Thirty-five DGRP insecticide phenotypes from nine insecticides were compiled (§3.2.1.1; table 3.2.1). The number of DGRP lines screened in each of the nine insecticide experiments ranged from 151 to 194 (table 4.2.1). Of the 212 DGRP lines screened with at least one insecticide (including lines from freeze 1 and 2 of the DGRP), 97 lines were common to all insecticide phenotypes (fig. 4.2.1). GWAS including Wolbachia infection status and five common chromosomal inversion genotypes as covariates were performed on all phenotypes using the DGRP2 pipeline (Huang *et al.* 2014; §2.2.1.3; appendix 1). Significant associations were determined using the 'genome-wide significance threshold' (1×10^{-5}), corrected for the number of phenotypes tested for each insecticide. The total number of unique genomic variants significantly associated with phenotypes from each insecticide ranged from 16 (imidacloprid) to 612 (parathion; table 4.2.1). A total of 1693 unique associations were identified, of which 1572 (93%) were private to a single insecticide. 118 variants were common to two insecticides and three variants were common to three insecticides (table 4.2.2; appendix 1). All but three of the variants associated with multiple insecticides are attributable to associations with organophosphate insecticides, and within selective sweeps around *Ace* and *Cyp6g1* (table 4.2.2).

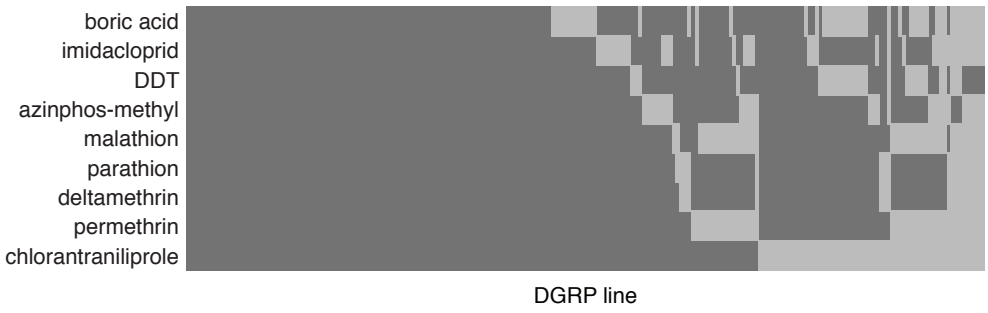


Figure 4.2.1. DGRP insecticide phenotype line sharing

DGRP lines screened (dark grey) in each of the nine insecticide experiments utilized in this chapter.

Table 4.2.1. Significantly associated variants from GWAS of 35 insecticide phenotypes (results pooled for each insecticide)

insecticide class	insecticide	phenotypes	DGRP lines screened	associated variants
borate	boric acid	1	162	22 (p<1E-5)
neonicotinoid	imidacloprid	2	175	16 (p<5E-6)
organochloride	DDT	2	176	36 (p<5E-6)
organophosphate	azinphos-methyl	5	178	193 (p< 2E-6)
organophosphate	malathion	8	169	273 (p<1.25E-6)
organophosphate	parathion	5	193	612 (p<2E-6)
pyrethroid	deltamethrin	1	194	148 (p<1E-5)
pyrethroid	permethrin	4	168	39 (p<2.5E-6)
ryanoid	chlorantraniliprole	7	151	481 (p<1.43E-6)

Table 4.2.2. DGRP variants associated with phenotypes from more than one insecticide

location	annotation	insecticides
2R:8059576-8073503	*Cyp6g1* sweep (15 variants)	azinphos-methyl; malathion
2R:10460322	*ttv* (intron); *LamC* (synonymous)	chlorantraniliprole; deltamethrin
2R:11852285	*Strn-Mlck* (intron)	chlorantraniliprole; deltamethrin
2R:12858731	*CR44377* (exon)	DDT; imidacloprid
3R:8979756-9188878	*Ace* sweep (99 variants)	malathion; parathion
3R:8988291	*Ace* sweep; *timeout* (stop gained)	azinphos-methyl; malathion; parathion
3R:8988335	*Ace* sweep; *timeout* (nonsynonymous)	azinphos-methyl; malathion; parathion
3R:9031423	*Ace* sweep; intergenic	azinphos-methyl; malathion; parathion

4.2.1.2. Association studies utilizing triallelic sites

Previously, a single state of a triallelic variant in *Cyp6w1* was fortuitously identified as a top DDT resistance candidate (Schmidt *et al.* 2017), motivating the incorporation of triallelic sites into association studies of all insecticide phenotypes. As sites with greater than two alleles are not included in DGRP variant files, multiallelic sites were recalled from DGRP BAM files. After applying filters for minor allele frequency (>0.05) and missing data (<20%) 15,487 triallelic sites remained. These variants were tested for association with each insecticide phenotype using ANOVA. Applying the same significance threshold as GWAS for each insecticide, a total of 35 significant associations were identified at 20 unique triallelic sites, two of which were associated with phenotypes from two different insecticides (table 4.2.3). 15 of the 20 triallelic associations were annotated to genes, ten of which were novel (i.e. genes not identified by GWAS using bialleles). Additionally, two trialleles were annotated as nonsynonymous substitutions, both in cytochrome P450 genes (*Cyp6w1* and *Cyp6a17*).

Table 4.2.3. Triallelic DGRP variants associated with insecticide phenotypes

location	annotation	reference allele (frequency)	alternate alleles (frequency)	insecticide phenotype association
X:6793168	-	T (0.83)	C (0.10); A (0.06)	chlorantraniliprole LC50
X:14549290	-	A (0.55)	G (0.40); T (0.05)	permethrin female 24-hr mortality
2L:4211388	*ft* (intron)	C (0.87)	T (0.08); G (0.05)	boric acid female lifespan
2L:10526460	*Lip4* (upstream)	G (0.51)	C (0.40); A (0.08)	DDT female 4-hr knockdown
2L:13265552	*CG31848* (synonymous); *CG44085* (intron)	G (0.13)	C (0.80); T (0.07)	boric acid female lifespan
2R:2062449	*Cyp6w1* (nonsynonymous)	T (0.51)	C (0.33); G (0.15)	DDT female 24-hr mortality
2R:10543936	*Adgf-E* (downstream)	A (0.30)	G (0.56); T (0.13)	deltamethrin male 48-hr mortality
2R:10761746	*Cyp6a17* (nonsynonymous)	C (0.88)	A (0.07); G (0.06)	deltamethrin male 48-hr mortality; malathion male 24-hr mortality
2R:12697064	-	C (0.81)	G (0.11); A (0.08)	chlorantraniliprole 5µg/ml mortality; chlorantraniliprole LC50
2R:13893421	*dpr13* (intron); *CR44419* (exon)	T (0.74)	C (0.21); G (0.06)	malathion female 3-hr mortality
2R:14754672	*DptB* (upstream); *CG43109* (downstream); *Dpt* (downstream)	G (0.88)	A (0.07); C (0.05)	chlorantraniliprole LC50
2R:19657305	*CG9850* (intron)	T (0.52)	G (0.30); C (0.18)	chlorantraniliprole 3µg/ml mortality
3R:8933601	*timeout* (intron)	G (0.58)	A (0.31); T (0.11)	parathion male 11-hr mortality; parathion male 24-hr mortality; parathion male 48-hr mortality
3R:8998707	-	T (0.47)	C (0.40); G (0.12)	malathion male 6-hr mortality; malathion female 6-hr mortality; malathion male 12-hr mortality; malathion female 12-hr mortality; malathion female 24-hr mortality; parathion male 11-hr mortality; parathion male 24-hr mortality; parathion male 48-hr mortality
3R:9059297	*Ace* (intron)	A (0.57)	G (0.29); C (0.14)	parathion male 11-hr mortality; parathion male 24-hr mortality; parathion male 48-hr mortality
3R:9158285	*PK2-R2* (5'UTR)	C (0.31)	A (0.61); T (0.08)	parathion male 24-hr mortality; parathion male 48-hr mortality
3R:9258198	*yrt* (synonymous)	G (0.49)	C (0.28); A (0.22)	parathion male 11-hr mortality; parathion male 24-hr mortality; parathion male 48-hr mortality
3R:9451645	-	G (0.36)	A (0.49); C (0.15)	parathion male 24-hr mortality; parathion male 48-hr mortality
3R:10018514	*Dop1R1* (intron)	C (0.84)	A (0.09); T (0.08)	malathion male 12-hr mortality
3R:25519156	*dmrt99B* (synonymous)	T (0.24)	A (0.67); C (0.10)	azinphos-methyl 0.25µg/ml mortality

4.2.1.3. Cross resistance

After correcting the direction of phenotypes such that increasing values indicated increasing resistance, correlations between all 35 insecticide phenotypes were assessed (fig. 4.2.2). The correlation coefficient (*r*) between insecticide phenotypes ranged from -0.19 to ~1 (extremely strong correlations are observed between late-timepoint parathion phenotypes). Predictably, the strongest correlations were observed between phenotypes from the same insecticide, followed by phenotypes from within the same insecticide class (organophosphates azinphos-methyl, malathion and parathion, and pyrethroids deltamethrin and permethrin). Boric acid and imidacloprid phenotypes showed the weakest positive and strongest negative correlations with other insecticide phenotypes.

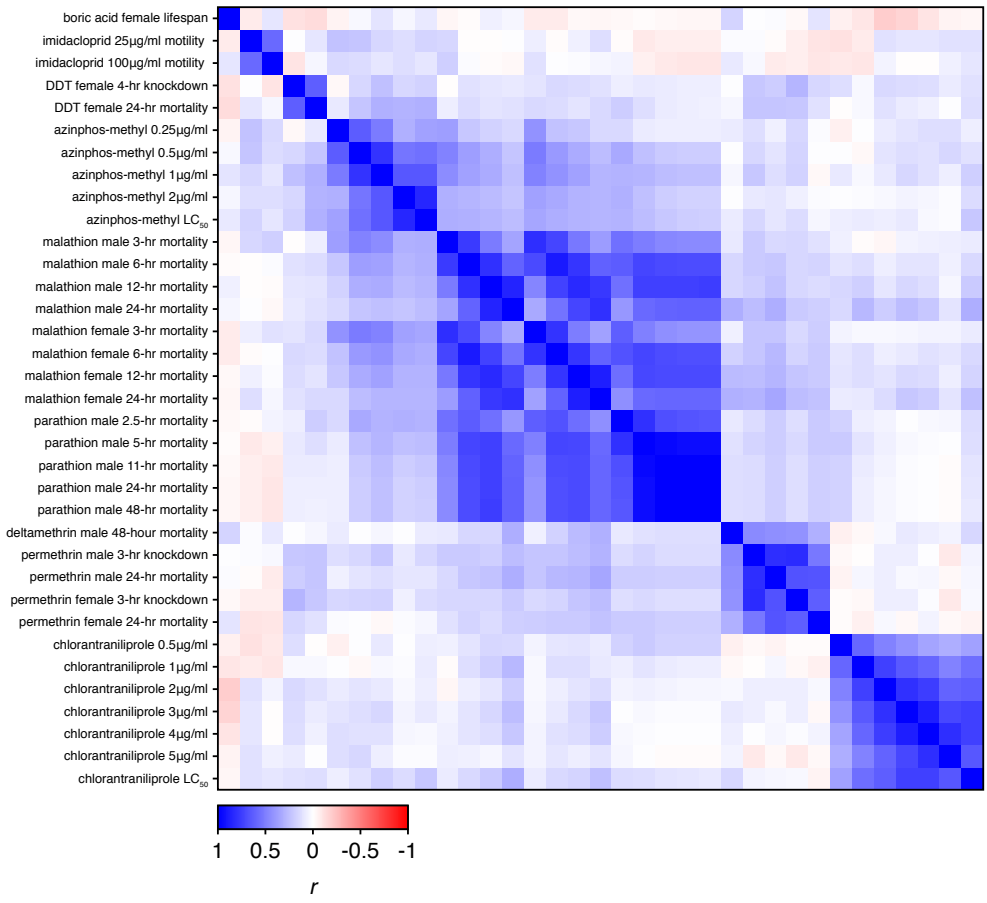


Figure 4.2.2. Insecticide phenotype correlations

Pairwise correlations (Pearson's *r*) between 35 phenotypes from DGRP studies of nine insecticides.

4.2.1.4. Pleiotropy and fitness costs

One hundred and thirteen DGRP phenotypes from 27 publications were assembled and divided into four categories: Behavior, physiology, stress, and toxic stress (table 4.2.4). As with the insecticide phenotypes, all survival-based phenotypes were transformed such that increasing values would reflect increasing survival. Therefore positive correlations between these phenotypes reflect the same lines exhibiting greater survival under both conditions. Abbott's correction for control mortality was applied to methylmercury/caffeine phenotypes from Montgomery *et al.* 2014, and correlations were calculated between each insecticide phenotype and each non-insecticide phenotype (appendix 3). These *r* values ranged from -0.28 to 0.23 for behavioral phenotypes, -0.28 to 0.26 for physiological phenotypes, -0.25 to 0.32 for stress phenotypes, and -0.24 to 0.37 for toxic stress phenotypes. Based on the mean DGRP line sample size across all phenotypes assessed (178) and a Bonferroni-corrected *p*-value of 1.26×10^{-5}, only three correlations reached a critical *r* of +/-0.32. These were all positive correlations, between permethrin male 24-hour mortality and both male paraquat survival (r=0.37) and male starvation resistance (r=0.32), and boric acid female lifespan and caffeine female lifespan (r=0.35). Although not reaching statistical significance, strong, negative correlations were observed between multiple fecundity phenotypes (Durham *et al.* 2014) and adult insecticide survival phenotypes (in particular female malathion phenotypes; appendix 3), suggesting a trade-off between fecundity and insecticide resistance in adults.

Table 4.2.4. Sources of DGRP phenotypes

Trait	Phenotype	Reference
behaviour		
courtship song	1	Turner *et al* 2013
food intake	2; male and female	Garlapow *et al.* 2015
male aggression	1	Shorter *et al.* 2015
phototaxis	5; male and female, 3 timepoints	Carbone *et al.* 2016
sleep	14; male and female	Harbison *et al.* 2013
startle response	2; male and female	Mackay *et al.* 2012
physiology		
abdominal pigmentation	2; tergite 5 and 6	Dembeck *et al.* 2015
benzaldehyde olfaction	2; male and female	Swarup *et al.* 2013
eclosion	1	Montgomery *et al.* 2014
fecundity	5; 4 timepoints and lifetime	Durham *et al.* 2014
genome size	1	Huang *et al.* 2014
mated lifespan	1	Durham *et al.* 2014
nutritional indices	6; glucose, glycerol, glycogen, triglyceride, protein, weight	Unckless, Rottschaefer & Lazzaro 2015
olfactory perception	28; male and female perception of 14 odorants	Arya *et al.* 2015
ovariole physiology	2; number and asymetry	Lobell *et al.* 2017
recombination rate	2; *e-ro*, *y-v*	Hunter *et al.* 2016
virgin lifespan	1	Ivanov *et al.* 2015
wing morphology	4; male and female, centroid size and interocular distance	Vonesch *et al.* 2016
stress		
brain injury survival	1	Katzenberger *et al.* 2015
C virus infection survival	1	Magwire *et al.* 2012
chill coma recovery	2; male and female	Mackay *et al.* 2012
Flock House virus infection survival	1	Magwire *et al.* 2012
Ma549 infection survival	2; male and female	Wang, Lu & Leger 2017
P. entomophila infection survival	1	Bou Sleiman *et al.* 2015
radiation survival	1	Vaisnav *et al.* 2014
Sigma virus infection resistance	1	Magwire *et al.* 2011
starvation resistance	2; male and female	Mackay *et al.* 2012
toxic stress		
caffeine survival	1	Najarro *et al.* 2015
caffeine survival	1	Montgomery *et al.* 2014
behaviour on MSB	4; male and female, startle response and negative geotaxis	Jordan *et al.* 2012
ethanol sensitivity	6; male and female, three timepoints	Morozova *et al.* 2015
methylmercury survival	4; 3 doses, 1 dose including caffeine	Montgomery *et al.* 2014
oxidative stress survival	4; male and female, MSB and paraquat	Weber *et al.* 2012

4.2.2. Population genomics of insecticide resistance

4.2.2.1. Insecticide resistance and selective sweeps

Previously, selective sweep peaks surrounding *Ace* and *Cyp6g1* have been associated with insecticide resistance phenotypes in the DGRP (§2.2.1; Duneau *et al.* 2018). To ascertain if any other insecticide-associated loci show evidence of selection, insecticide-associated variants were tested for overlap with the 50 most pronounced signatures of selection on chromosomes 2 and 3 identified by Garud *et al.* (2015) using the H12 statistic. Insecticide-associated variants were present in 13 of these 50 windows, however this is not in itself indicative of whether these variants are involved in these sweeps. Therefore, insecticide-associated variants identified in H12 sweeps were further filtered by asking whether a significant enrichment of resistance-associated alleles is observed in lines carrying the most common haplotypes (H1 and H2) at each H12 peak (Fisher's exact test; $p<0.05$). After applying this filter to the data, there remained a total of nine overlaps between significant associations from five insecticides and seven H12 peaks (table 4.2.5).

Table 4.2.5. Insecticide-associated variants which overlap H12 selective sweep peaks, and show an enrichment for resistance alleles in DGRP lines carrying the two most common haplotypes

H12 peak	azinphos-methyl	chlorantraniliprole	DDT	malathion	parathion
2L:14840204-14892944	-	1	-	-	-
2R:1982987-2232677	-	-	1	-	-
2R:7722921-8486706 (*Cyp6g1*)	94	-	-	18	-
3R:8333340-8562798	-		-	-	2
3R:8563358-9432483 (*Ace*)	-	-	-	127	596
3R:13195146-13304663	-	-	-	4	-
3R:26016366-26044781	-	3	-	-	-

To further validate H12 overlaps, two other selective sweep statistics, *i*HS (Voight *et al.* 2006) and nS_L (Ferrer-Admetlla *et al.* 2014), were calculated for DGRP variants (Joshua Schmidt pers. comm). *i*HS and nS_L attempt to identify loci where the derived allele is present on a longer haplotype than the ancestral allele (*i*HS defines haplotype length in base pairs, while nS_L defines it in by the number of variant sites). Genome-wide outliers of the absolute values of each of these statistics (empirical p-value<0.05) were cross-referenced with insecticide-associated variants (appendix 1). Only insecticide-associated variants in H12 sweeps around *Ace, Cyp6g1* and *Cyp6w1* showed signs of selection by all three statistics, while an additional ten variants showed signs of selection by two statistics (table 4.2.6).

Table 4.2.6. Insecticide-associated variants showing evidence of selection from two or more selective sweep statistics

location	annotation	insecticide	statistic
2L:7020566	*ade3* (intron)	chlorantraniliprole	*i*HS; *n*SL
2L:11174340	*Ca-beta* (intron)	azinphos-methyl	*i*HS; *n*SL
2L:15641504		azinphos-methyl	*i*HS; *n*SL
2R:2062449	*Cyp6w1* (nonsynonymous)	DDT	H12; *i*HS; *n*SL
2R:7722921-8486706	*Cyp6g1* H12 sweep	azinphos-methyl; malathion	H12; *i*HS; *n*SL
2R:10153715	*CG8547* (intron)	deltamethrin	*i*HS; *n*SL
2R:11639298	*CG30089* (intron)	malathion	*i*HS; *n*SL
3L:2041056	*sls* (synonymous)	chlorantraniliprole	*i*HS; *n*SL
3L:2525372	*Spn* (synonymous); *CG32295* (downstream)	chlorantraniliprole	*i*HS; *n*SL
3L:3383322	-	chlorantraniliprole	*i*HS; *n*SL
3R:14349974	*fru* (intron)	chlorantraniliprole	*i*HS; *n*SL
3R:6077388	*CG3999* (synonymous); *CG42795* (downstream)	malathion	*i*HS; *n*SL
3R:8563358-9432483	*Ace* H12 sweep	azinphos-methyl; malathion; parathion	H12; *i*HS; *n*SL

4.2.2.2. *91-R* and *91-C*

Additional evidence for variants responding to DDT-based selection is provided by the genome sequences of *D. melanogaster* lines *91-R* and *91-C* (Steele *et al.* 2014). *91-R* has been selected for DDT resistance for ~30 years (Strychartz *et al.* 2013) and as a result, it is ~1500 times more resistant to the compound than its control, *91-C*. Direct comparison of variants differentiated between *91-R* and *91-C* is insufficient to distinguish neutral stochasticity from DDT selection, and as such the genetic basis of *91-R*'s DDT resistance remains largely unaccounted for (Strychartz *et al.* 2013). However, multiple candidate mechanisms for DDT resistance have been identified in the line, including constitutive upregulation of cncC-regulated transcripts (Misra *et al.* 2013).

Genome sequences of *91-R* and *91-C* generated by Steele *et al.* (2014) were compared and contrasted with DGRP variants associated with phenotypes from DDT and PC1 GCTM 79|cncC-activated transcripts (the phenotype used to map constitutive cncC upregulation among DGRP lines; §3.2.2). A total of 864,878 variants differentiated between *91-R* and *91-C* were compared with the 36 and 106 DGRP variants associated with DDT and PC1 GCTM 79|cncC-activated transcripts phenotypes respectively, of which seven and 47 respectively were differentiated between *91-R* and *91-C*, and four and 13 respectively were differentiated in the expected direction (i.e. with the allele state in *91-R* associated with increased DDT resistance or cncC-activated transcript abundance in the DGRP; table 4.2.7). These included variants in *CG10737* and *Cyp6w1*, genes with transgenically validated roles in DDT resistance (Schmidt *et al.*

2017). This result adds further weight to the DDT associations with these variants in the DGRP, and also suggests that although they may not be associated with selective signatures in the DGRP, they may have been selected by DDT during the generation of *91-R*.

Table 4.2.7. DGRP variants associated with DDT or PC1 GCTM 79|cncC-activated transcripts, and also differentiated between *91-R* and *91-C* in the expected direction

location	annotation	*91-C* allele	*91-R* allele
DDT			
2L:14188350	*Dyrk2* (intron)	G/G	G/C
2R:2062449	*Cyp6w1* (nonsynonymous)	T/T	C/C
2R:15056934	*CG10737* (intron)	T/T	G/G
3R:26794618	*5-HT7* (downstream)	A/A	ATA/ATA
PC1 GCTM 79|cncC-activated transcripts			
X:1626076	-	T/T	A/A
X:1626161	-	T/T	C/C
2R:11277198	-	G/G	T/T
2R:11277297	-	G/G	A/A
2R:11277331	-	T/T	C/C
2R:11277343	-	G/G	A/A
3R:4660636	*pyd* (downstream)	G/G	A/A
3L:12237872	*app* (intron); *CR43991* (downstream)	C/T	C/C
3L:12248962	*app* (intron)	G/G	A/A
3L:12254655	*app* (intron)	C/C	G/G
3L:12254733	*app* (intron)	G/G	C/C
3L:12254736	*app* (intron)	T/T	G/G
3R:10944757	*CG42788* (intron)	G/G	A/A

4.2.2.3. DGRP insecticide-associated variants in worldwide populations

To gain insight into the extent to which insecticide associations in the DGRP may be attributable to *D. melanogaster* in general, population genome sequence datasets from the Drosophila Genome Nexus (DGN; Lack *et al.* 2015) were interrogated. The 15 populations containing ten or more individuals (fig. 4.2.3A) were examined for single nucleotide polymorphisms (SNPs) associated with DGRP phenotypes from nine insecticides. For each of the nine insecticides, between 97.9 and 100% of associated SNPs were observed at least once in a DGN population (table 4.2.8). To ascertain whether this represents a departure from neutral expectations, the total number of DGRP SNP variants which could be utilized in GWAS (i.e. MAF≥0.05) were tested for presence in any of the 15 DGN populations. 97.8% of all DGRP SNPs (MAF≥0.05) were observed in at least one of the 15 DGN populations, a proportion which is not significantly different from any of the proportions of insecticide-associated SNPs (binomial $p>0.60$ in all cases).

At the level of individual populations, the proportion of insecticide-associated DGRP SNPs present ranged from 35.9 to 95.5% (fig. 4.2.3B), values which were only partially explained by population sample size (r^2=0.006-0.19 for correlations between the proportion of DGRP variants represented for each population, and the population's sample size). Geographic distance from Raleigh, NC likewise explained only a modest proportion of the variance in the proportion of insecticide-associated SNPs in each population (r^2=0.02-0.27).

Table 4.2.8. Number of SNPs associated with each insecticide, and percentage that are observed at least once in any of 15 DGN populations

Insecticide	Associated SNPs	% present in any of 15 DGN populations
azinphos-methyl	185	99.46
boric acid	22	100.00
chlorantraniliprole	429	99.07
DDT	32	100.00
deltamethrin	141	97.87
imidacloprid	13	100.00
malathion	246	99.19
parathion	569	99.12
permethrin	37	100.00

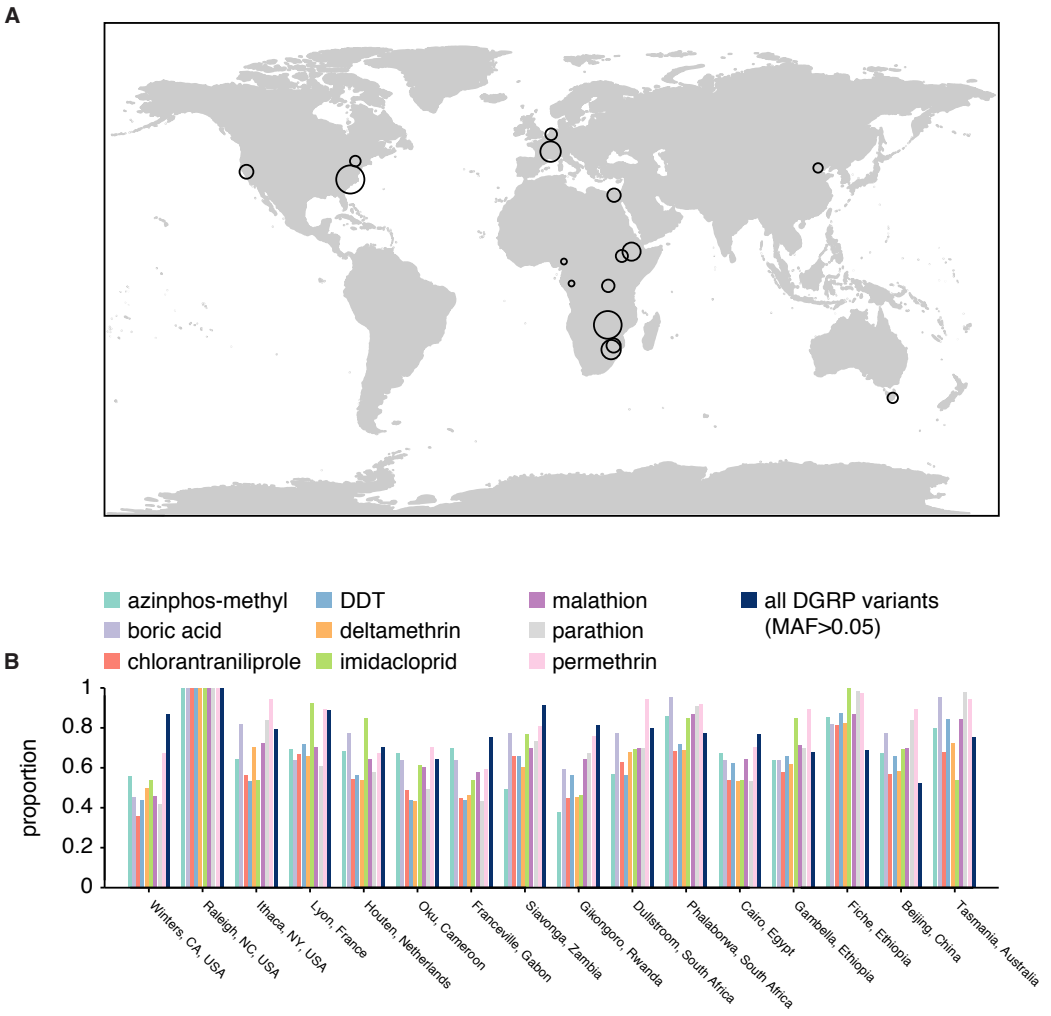


Figure 4.2.3. Presence of insecticide-associated DGRP variants in 15 DGN populations
Location (A) and proportion of insecticide-associated DGRP variants present (B) for 15 DGN populations and the DGRP (Raleigh, NC).

Allele frequencies for top insecticide candidates *Ace*, *Cyp6g1, Cyp6a17* and *Cyp6w1* were also calculated for each population (fig. 4.2.4), and were generally more frequent in populations outside of Africa, consistent with a population bottleneck during the expansion of the species out of Africa. At least one of the four resistance substitutions at *Ace* was observed in ten of the 15 DGN populations (fig. 4.2.4A), although combinations of these alleles are often highly distinct from the DGRP (e.g. Ace^{G303A}, which is not observed alone on any DGRP haplotype, is the only substitution observed in the Beijing population). Derived alleles of *Cyp6g1* as defined by evidence of the *Accord* insertion and gene duplication events (fig. 2.2.6) are observed in all DGN populations, and are almost at fixation in all populations outside of Africa (fig. 4.2.4B).

Triallelic variants at *Cyp6a17* and *Cyp6w1* are also widely observed among DGN populations (fig. 4.2.4C,D) but show highly variable frequencies.

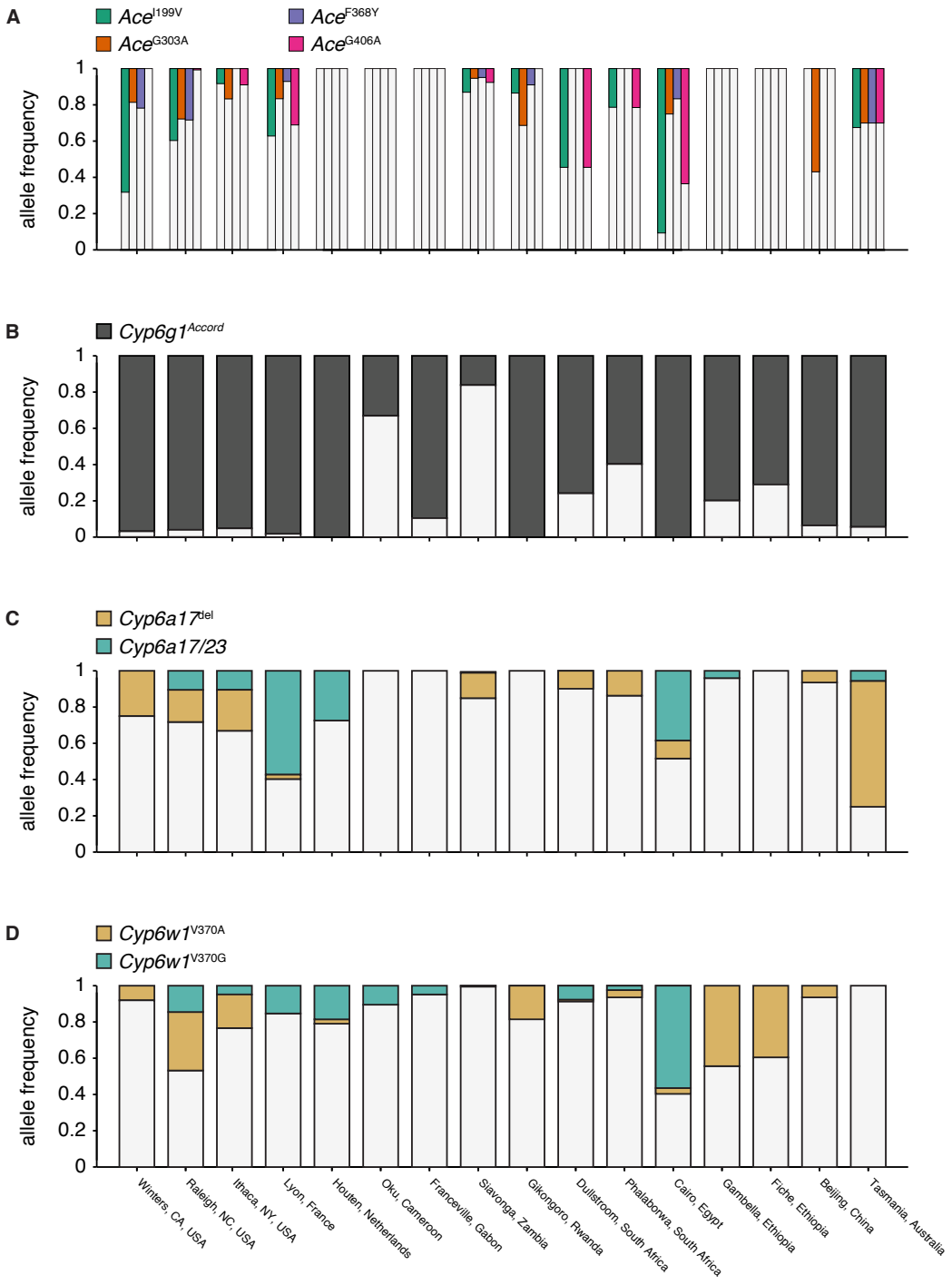


Figure 4.2.4. Allele frequencies of top DGRP insecticide candidates in 15 DGN populations
Derived allele frequencies of top insecticide candidate loci *Ace* (A), *Cyp6g1* (B), *Cyp6a17* and *Cyp6a23* (C) and *Cyp6w1* (D) in 15 DGN populations and the DGRP (Raleigh, NC).

H12 scans of the four DGN populations with the largest sample size (including the previously analysed Zambia population [Garud and Petrov 2016]) reveal discernible evidence of a selective sweep at *Cyp6g1* in all four scans (as well as the DGRP; fig. 4.2.5), concordant with the high frequency of derived alleles observed in DGN populations (fig. 4.2.4B). Sweeps at *Ace* are also evident in three of the four populations, the absence of the sweep in Ethiopia consistent with the absence of *Ace* resistance substitutions in this population sample (fig. 4.2.4A).

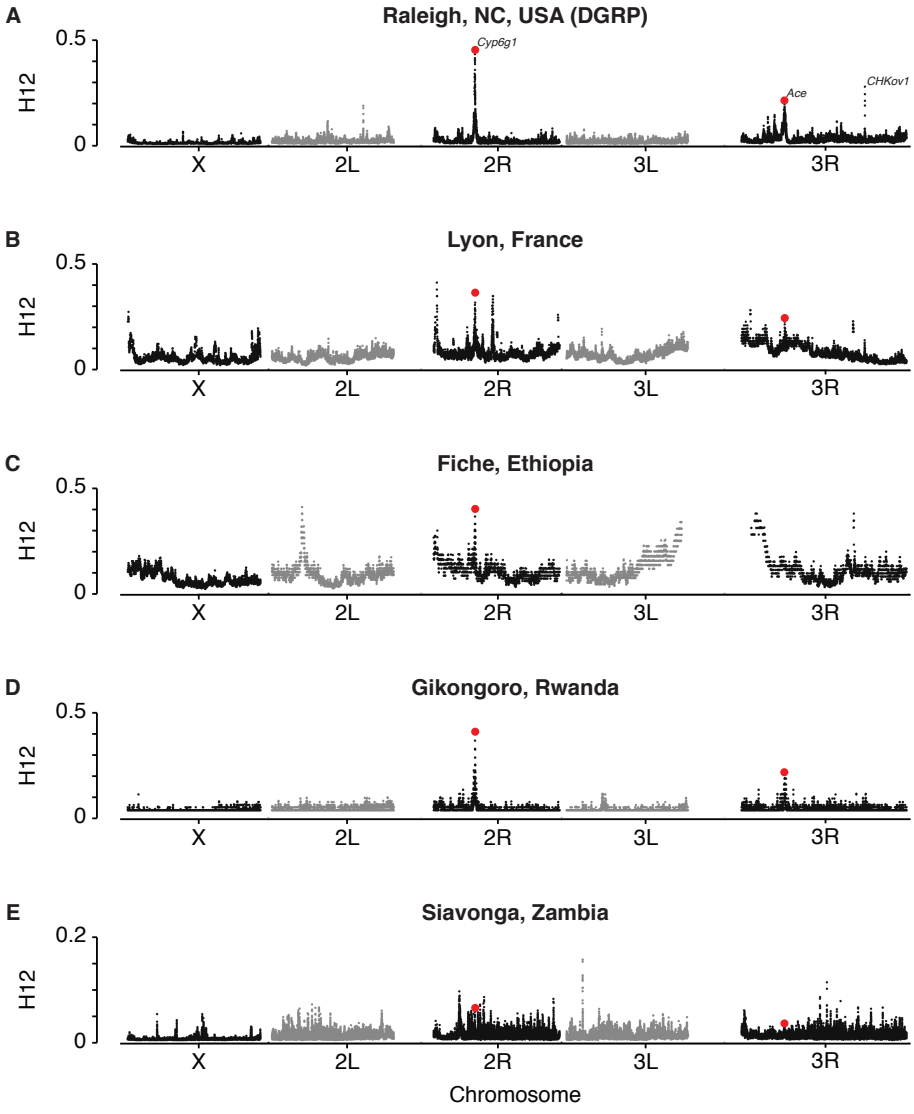


Figure 4.2.5. H12 across five populations

Genome-wide H12 scans for the four largest DGN population samples (B-E) in comparison to the DGRP (A). Peaks of sweeps containing *Ace* and *Cyp6g1* are represented in red.

4.3. Discussion

4.3.1. The quantitative nature of phenotypes from multiple insecticides

In this chapter, insecticide resistance in the DGRP is examined in the context of other DGRP phenotypes, population-wide selection, and worldwide variation. These analyses are achieved by leveraging the DGRP's utilities beyond association mapping; for population-genomic analyses and phenomic analyses, and moreover by assessing variation in the multiple population samples present in the Drosophila Genome Nexus (DGN; Lack *et al.* 2015).

Despite associations with alleles of large effect, some of which feature footprints of recent selection, GWAS of insecticide phenotypes suggests a quantitative trait architecture in all cases, with associated variants ranging from 16 to 612 for individual compounds. For the majority of insecticides, transgenic experiments support the involvement of more than one gene (table 4.3.1).

Table 4.3.1. Transgenically validated insecticide GWAS candidates

insecticide	candidate	validation tool	reference
imidacloprid	*Cyp6g1*	knockout (CRISPR)	Denecke *et al.* 2017
	Cyp6g2	overexpression	Denecke *et al.* 2017
DDT	*CG10737*	RNAi	Schmidt *et al.* 2017
	Cyp6w1	mutagenesis; overexpression	Schmidt *et al.* 2017
azinphos-methyl	*Cyp6g1*	knockout (CRISPR)	§2.2.2
malathion	*Ace*	mutagenesis; *in vitro* expression	Menozzi *et al.* 2004
	Cyp12d1	overexpression	§3.2.3
	Cyp6g1	knockout (CRISPR)	§2.2.2
parathion	*Ace*	mutagenesis; *in vitro* expression	Menozzi *et al.* 2004
	Dscam1	gene disruption	Duneau *et al.* 2018
	trpl	overexpression	Duneau *et al.* 2018
deltamethrin	*CG7627*	RNAi	Duneau *et al.* 2018
	Cyp6a17	gene disruption; RNAi	Duneau *et al.* 2018
	Cyp6a23	RNAi	Duneau *et al.* 2018
permethrin	*Cyp6a17*	gene disruption	§2.2.2
chlorantraniliprole	*Cyp12d1*	overexpression	§3.2.3
	Strn-Mlck	RNAi	Green *et al.* 2019

Liu, Li and Pritchard (2019) demonstrate that even in human GWAS (which have far more statistical power than association studies in the DGRP), much of the heritability of quantitative traits lies below statistical significance. While this 'omnigenic' perspective on quantitative traits suggests false negatives are prevalent in GWAS, it generates an issue for assessment of false positives: Transgenic manipulation of GWAS candidate genes is insufficient to assess the validity of associations, as under an omnigenic model

most genes (or at least those with trait-relevant expression patterns) should have a significant effect on any phenotype if they are aggressively manipulated. Zhang, Reeves and Tautz (2019) provide support for this premise using *D. melanogaster*. After performing a DGRP GWAS on pupation site choice, they found that of 20 randomly selected non-candidate genes, knockouts of twelve had significant effects on the phenotype.

4.3.2. Between-phenotype correlations

Correlations were assessed among insecticide phenotypes and a library of DGRP phenotypes. As would be expected, correlations within the insecticide dataset were generally strongest between phenotypes from the same insecticide, and then between phenotypes from insecticides from the same class (fig. 4.2.2). Correlations between phenotypes from different insecticides suggest the presence of cross-resistance, although an effect of the assay method may also be involved, as the most distinct phenotypes were also measured using the most distinct methods (the boric acid phenotype measured adult lifespan at a relatively low concentration [Najarro *et al.* 2017], and the imidacloprid phenotypes measured larval motility at two non-lethal concentrations [Denecke *et al.* 2017]). This assay effect is reinforced in the significant correlation between boric acid and caffeine phenotypes (appendix 3), which were performed in the same lab using the same assay methods (Najarro *et al.* 2015; Najarro *et al.* 2017).

Given mortality and eclosion success are the basis for many insecticide phenotypes analyzed in this chapter, it was encouraging that no statistically significant or strong patterns were observed in correlations between insecticide phenotypes and mated lifespan (Durham *et al.* 2014), virgin lifespan (Ivanov *et al.* 2015) or eclosion phenotypes (Montgomery *et al.* 2014), suggesting that, on the whole, variation in life history traits is not confounding the toxin's effect in measures of insecticide phenotypes. Strong correlations (statistically significant in males) were, however, observed between permethrin 24-hour mortality and both paraquat survival and starvation resistance (r=0.37 and 0.32 respectively). Starvation resistance makes sense as a pyrethroid phenotype correlate, as pyrethroid insecticides exert a strong knockdown effect. Mortality is expected to occur as one of the results of extended knockdown, which may include starvation. It is unclear why the correlation between survival on the herbicide paraquat would be especially strong with permethrin mortality; there were no overlaps between the 475 variants associated with Weber *et al.*'s (2012) paraquat phenotype and any permethrin phenotype. The answer likely lies with oxidative stress. Like paraquat, permethrin appears to cause oxidative stress in *D. melanogaster* (Terhzaz *et al.* 2015), and *iPLA2-VIA*, a phospholipase involved in oxidative stress sensitivity, is among the top permethrin GWAS candidates (appendix 1.9; §2.3.3).

4.3.3. Evidence for insecticide-driven selection in the DGRP

In contrast to the correlations observed between insecticide phenotypes, after performing GWAS only seven percent of insecticide-associated variants were common between insecticides. Of these 121 variants associated with more than one insecticide, 118 are within selective sweep regions containing *Ace* and *Cyp6g1* (table 4.2.2), suggesting the bulk of these variants are artefacts of linkage disequilibrium with the insecticide resistance alleles in these regions. The explanation for this disparity between correlations at the phenotypic level and common genomic associations likely lies among the genomic variation excluded by these GWAS; low-frequency variants, excluded by the minor allele frequency threshold, and variants not reaching statistical significance for association. Although rare, associated variants common to multiple insecticides are enriched for signatures of selection, which prompts the detailed analysis of these regions, and DGRP insecticide-associated variants in general, in the context of population-scale evidence of selection.

4.3.3.1. *Ace* and *timeout*

The largest source of between-insecticide associations, the H12 selective sweep region which includes *Ace* (Garud *et al.* 2015), contains 102 variants associated with malathion and parathion phenotypes, of which three are also associated with azinphos-methyl. Malathion and parathion-associated variants include the three *Ace* resistance substitutions present in the DGRP at moderate frequencies, and the number and strength of associations observed in the sweep region are consistent with the chemical inhibition of each insecticide by mutant *Ace* enzymes described by Menozzi *et al.* (2004; fig. 4.3.1).

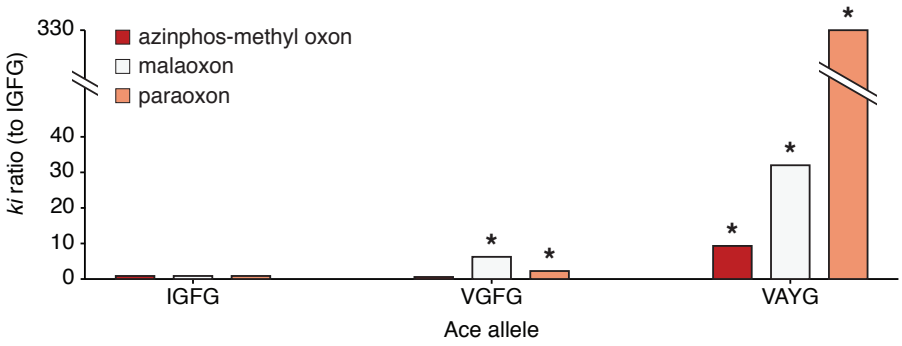


Figure 4.3.1. Ace inhibitory constants (data from Menozzi *et al.* 2004)

ki ratios (relative to the ancestral IGFG allele) of activated forms of azinphos-methyl (methyl-azinphos oxon), malathion (malaoxon) and parathion (paraoxon) for insecticide-resistant Ace alleles present in the DGRP. Asterisks represent statistically significant differences from the IGFG allele.

While it might be expected that the three variants from the *Ace* sweep which are also associated with azinphos-methyl could be could be attributed to the Ace^{VAYG} allele, examination of the locations of these variants reveals that they are not in strong LD with the *Ace* resistance substitutions (r^2<0.08 in all cases), and two of the three are predicted to be missense and nonsense mutations in *timeout* (fig. 4.3.2). *timeout* is involved in light entrainment of the adult circadian clock (Benna *et al.* 2010), and circadian rhythm has long been known to affect insecticide resistance (e.g. Shipp and Otton 1976). Resistance to malathion, fipronil (a phenylpyrazole) and propoxur (a carbamate) has been shown to vary with the circadian clock in *D. melanogaster*, as have multiple transcripts associated with insecticide resistance including *Cyp6g1* and *Ace* (Hooven *et al.* 2009).

Although the potential contribution of variation in *timeout* to insecticide resistance remains unresolved, and cannot be excluded as a contributing factor, comparisons between insecticide-associations and haplotype structure at the *Ace* sweep provide strong evidence for insecticide-driven selection in this region though *Ace* insensitivity alleles: The peaks of GWAS and H12 are aligned (fig. 4.3.3), and *Ace* insensitivity alleles are almost perfectly explained by the two most common haplotypes at the sweep peak (fig. 4.3.4).

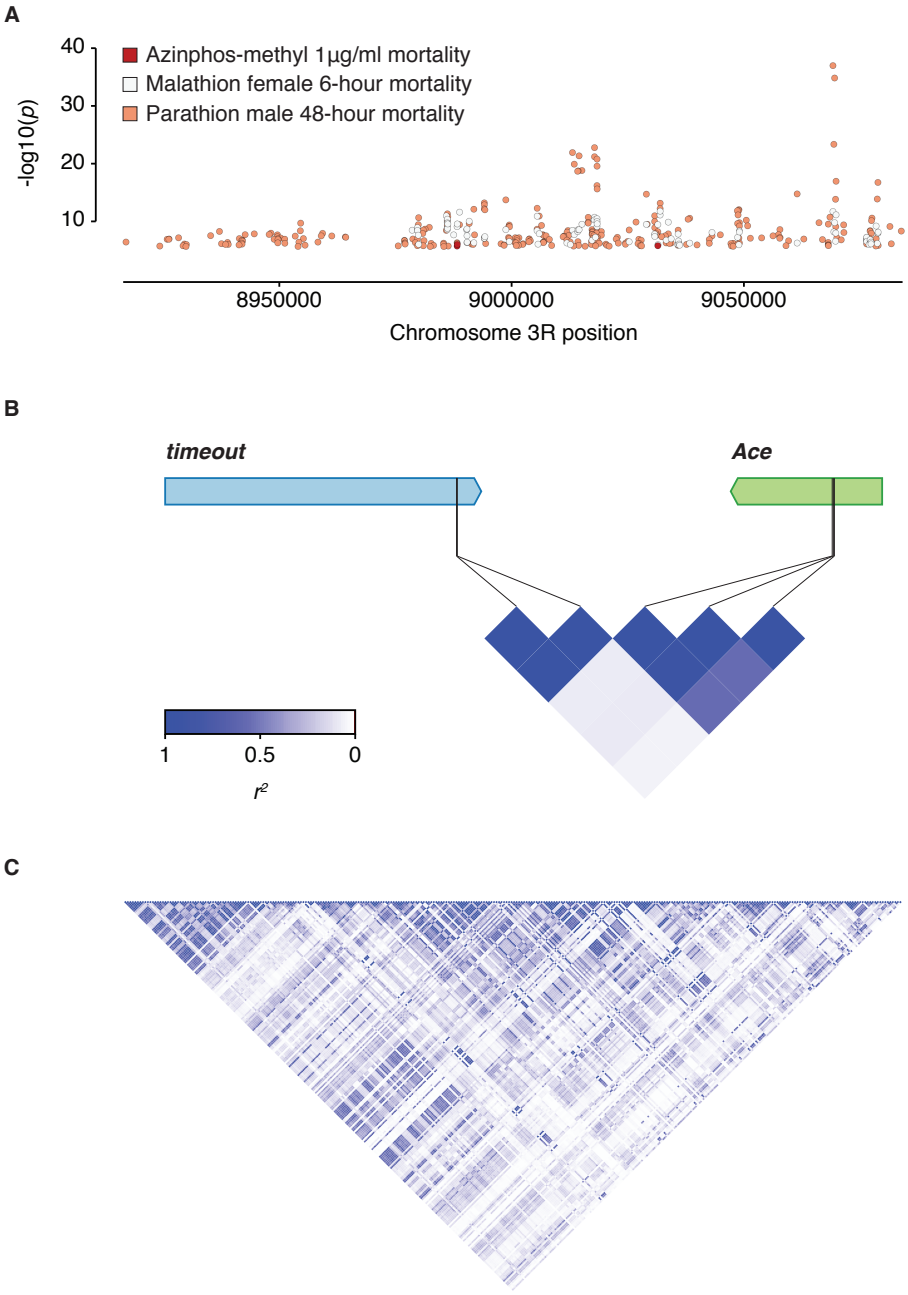


Figure 4.3.2. Organophosphate associations with *Ace* and *timeout*

GWAS associations for representative phenotypes from three organophosphate insecticides (A) in the genomic region encompassing *Ace* and *timeout* genes (positions on the x-axis shown in B). Despite widespread linkage disequilibrium among these GWAS variants in the region (C), top azinphos-methyl candidates in *timeout* are not in linkage disequilibrium with *Ace* resistance substitutions (B).

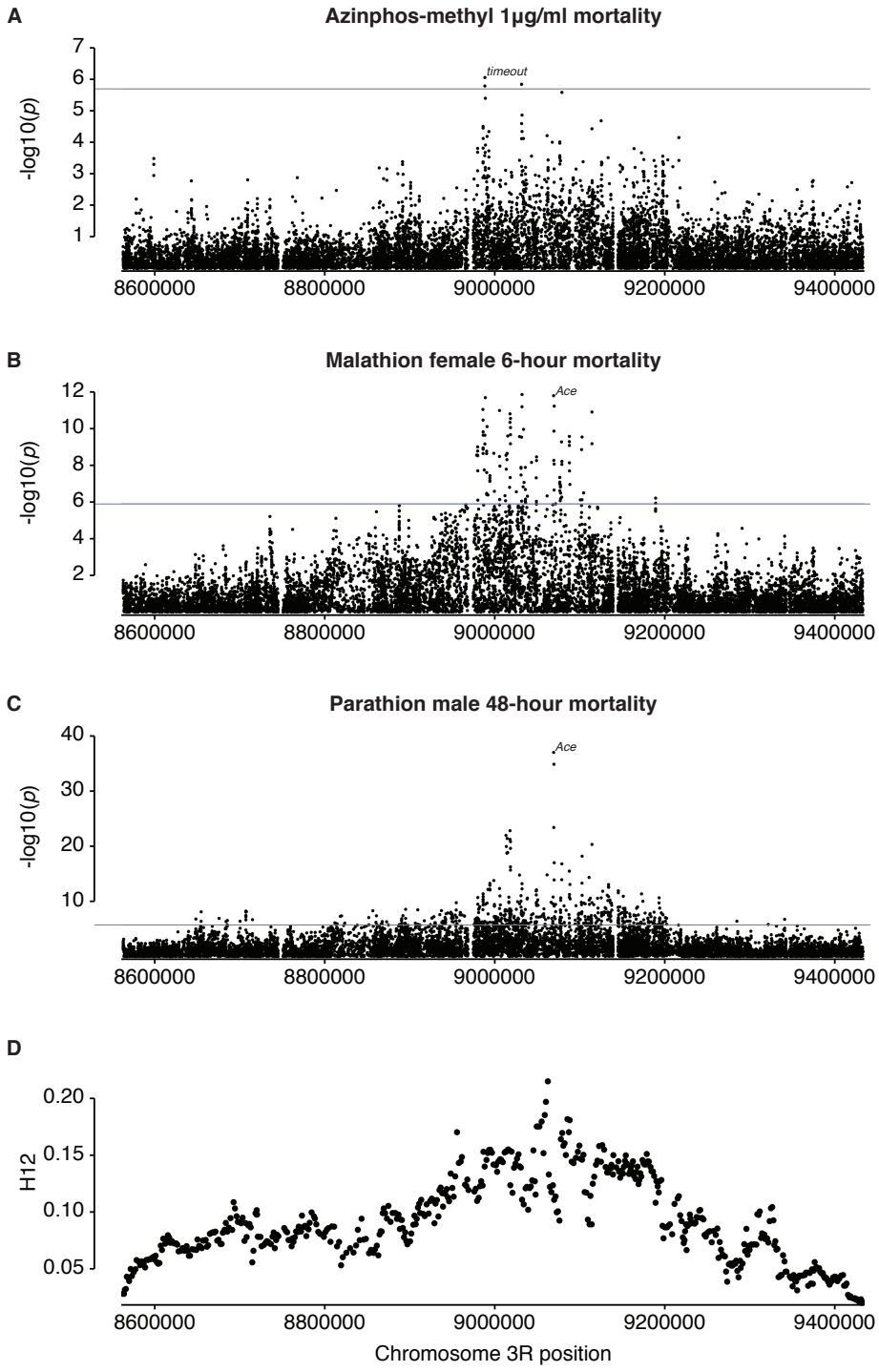


Figure 4.3.3. The *Ace* selective sweep

Comparison between extreme insecticide associations (A-C) and local H12 statistic values (D) in the selective sweep peak encompassing *Ace*.

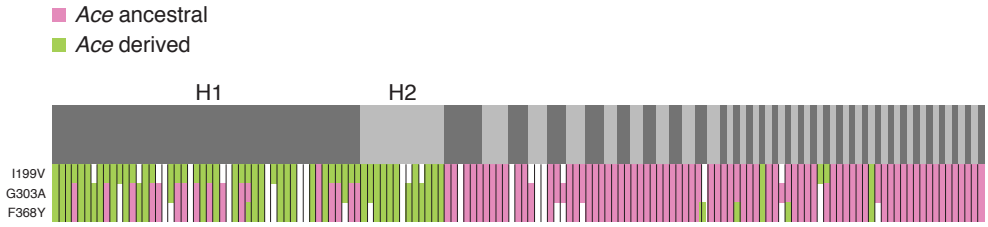


Figure 4.3.4. Resistance allele structure at the *Ace* selective sweep

The distribution of each of the three *Ace* resistance substitutions among haplotype groups in the selective sweep peak encompassing *Ace*. White blocks represent missing genotype data.

4.3.3.2. *Cyp6g1*

The selective sweep at *Cyp6g1* explains a further 16 of the variants associated with phenotypes from multiple insecticides; in this case, due to associations with organophosphates azinphos-methyl and malathion. As observed at the *Ace* sweep, peaks of insecticide association at *Cyp6g1* correspond to the H12 peak in the region (fig. 4.3.5), and haplotype structure at the peak (fig. 4.3.6) is consistent with selection for derived *Cyp6g1* alleles, with all ancestral *Cyp6g1*^{M} alleles carried on rare haplotypes. Note that Garud *et al.*'s (2015) data, which is based on 145 DGRP lines, only includes five *Cyp6g1*^{M} alleles (fig. 4.3.6A). Recalculating H12 for the region utilizing 205 DGRP lines shows that this pattern holds, as lines carrying the remaining four *Cyp6g1*^{M} alleles have singleton haplotypes at this peak (fig. 4.3.6B).

While strong associations with *Cyp6g1* and its associated selective sweep are observed for azinphos-methyl and malathion phenotypes, previous studies would lead to the expectation of associations with phenotypes from other insecticides considered in these analyses. *Cyp6g1* was originally mapped as a resistance locus using DDT (Daborn *et al.* 2001), and an association study using Australian populations of *D. melanogaster* identified a strong *Cyp6g1* effect on DDT resistance (Schmidt *et al.* 2010). *Cyp6g1*-mediated cross-resistance to imidacloprid has also been described (Daborn *et al.* 2001), and parathion resistance has previously been mapped to a region including *Cyp6g1* (Kikkawa 1961). The transcript level of *Cyp6g1* was significantly associated with DGRP imidacloprid phenotypes (table 3.2.2; appendix 2), however no significant genomic or transcriptomic associations linked either DDT or parathion phenotypes to *Cyp6g1* (Schmidt *et al.* 2017; Duneau *et al.* 2018; appendix 1; appendix 2).

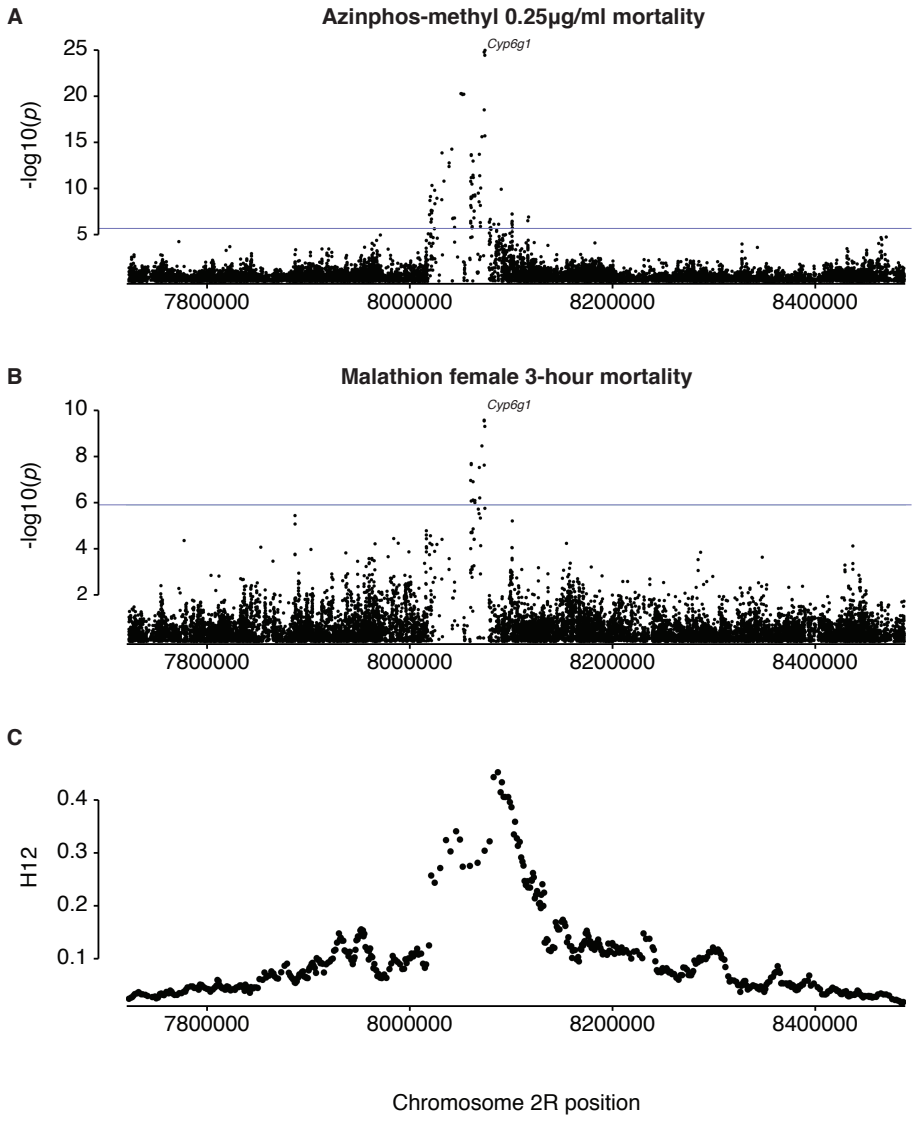


Figure 4.3.5. The *Cyp6g1* selective sweep

Comparison between extreme insecticide associations (A,B) and local H12 statistic values (C) in the selective sweep peak encompassing *Cyp6g1*.

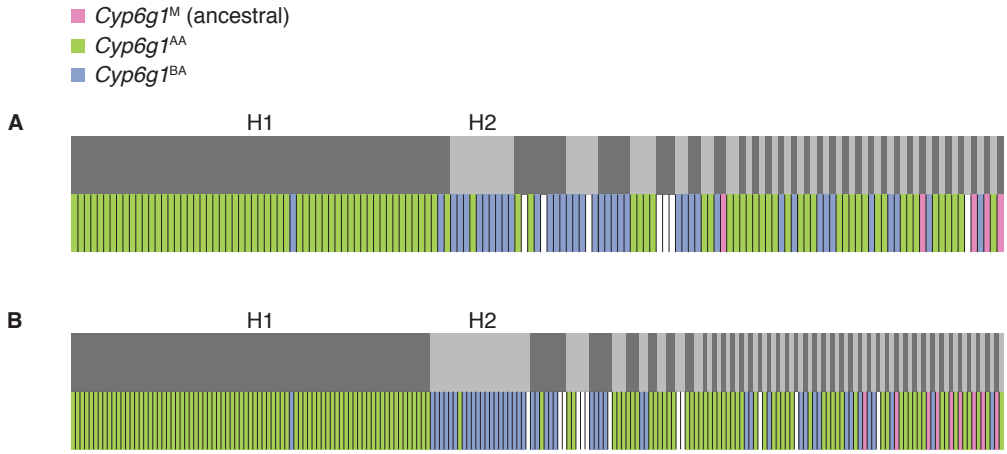


Figure 4.3.6. Resistance allele structure at the *Cyp6g1* selective sweep

The distribution of *Cyp6g1* alleles among haplotype groups in the selective sweep peak encompassing *Cyp6g1* for the 140 lines utilized by Garud *et al.* (2015) and present in DGRP freeze 2 (A), and all 205 DGRP freeze 2 lines (B). White blocks represent missing genotype data.

One explanation for the lack of expected associations between *Cyp6g1* and these other compounds is the frequency of the alleles in the DGRP. Only nine DGRP lines are homozygous for the ancestral, susceptible *Cyp6g1*^{M} allele, and phenotypes must therefore be heavily skewed towards survival to maximize the detection of a low-frequency minor allele associated with susceptibility. This is consistent with azinphos-methyl and malathion *Cyp6g1* associations, which are only significant for certain phenotypes (the three lowest concentrations of azinphos-methyl and the earliest malathion timepoint; fig. 2.2.1; fig. 2.2.2). Such phenotypes were not measured in DDT and imidacloprid studies, however Duneau *et al.*'s (2018) parathion phenotypes included a 2.5-hour phenotype skewed towards survival. Using the population mean (μ) as a measure of skew, 2.5-hour parathion mortality μ=0.27, which should have been sufficient to detect an association with *Cyp6g1* under this hypothesis (§2.3.1).

Another factor that may account for population-level differences in *Cyp6g1*-mediated insecticide resistance is that *Cyp6g1* does not act alone in detoxifying insecticides. Biochemical studies of interactions between *Cyp6g1* and insecticides reveal that metabolites produced by *Cyp6g1* are not necessarily less toxic (Joußen *et al.* 2008; Hoi *et al.* 2014; Fusetto *et al.* 2017) and detoxification therefore requires additional

metabolic steps and/or transport. These actors downstream of *Cyp6g1* remain to be identified, but epistatic interactions between their activity and that of *Cyp6g1* may dictate the importance of derived *Cyp6g1* alleles to insecticide resistance in different genetic background.

Furthermore, the rich allelic diversity at *Cyp6g1* undoubtedly complicates between-population comparisons of the gene's contribution to insecticide resistance. For example, *Cyp6g1*^{BP} alleles (a more derived form of the *Cyp6g1*^{BA} allele characterized by an additional insertion of a partial *P-element* sequence) segregate in Australian populations and explain 16 percent of the variation in DDT resistance in this population (Schmidt *et al.* 2010). This allele is absent from the DGRP (Schmidt *et al.* 2017), and indeed most populations investigated outside of Australia; Catania *et al.* (2005) identified ten *P-element* insertions from their global sample of 673 *D. melanogaster* lines (which included 38 lines from Australia), but only reported the origin of one of these alleles: New Delhi. Interestingly, Catania *et al.* (2005) also identified a *Cyp6g1* insertion variant in two lines from Zimbabwe (Harara and Lake Kariba) lacking an *Accord* insertion. Analysis of the *Cyp6g1* locus in the DGRP has also revealed significant variation in the copy number of both amplicons associated with derived alleles (fig. 3.2.3), however the extent of this variation in other population remains unknown.

4.3.3.3. *Cyp6a17* and *Cyp6a23*

Cyp6g1 is not the only cytochrome P450 locus involved with phenotypes from multiple insecticides. Previously, two deletions involving *Cyp6a17* and *Cyp6a23* were described among DGRP lines (Good *et al.* 2014; fig. 2.2.9), and both have been associated with pyrethroid insecticide phenotypes. *Cyp6a17*^{del}, an almost complete deletion of *Cyp6a17*, is associated with increased susceptibility to permethrin (fig. 2.2.10; fig. 3.2.6B), while *Cyp6a17/23*, a deletion which creates a single, chimeric gene derived of *Cyp6a17* and *Cyp6a23* sequence is associated with increased resistance to deltamethrin (fig. 3.2.6A). No overlap was found between variants significantly associated with deltamethrin and permethrin (table 4.2.2), consistent with associations with different alleles. However, the association between deltamethrin and the *Cyp6a17*^{del} allele sits just below the significance threshold, and the two SNPs identified as top candidates in the deltamethrin GWAS performed by Duneau *et al.* (2018; `2R:10763630` and

2R:10764362) are in LD with the *Cyp6a17*^{del} allele (r^2=0.96 and 0.73 respectively). This suggests that *Cyp6a17*^{del} and *Cyp6a17/23* may have opposing effects on resistance to both pyrethroid insecticides, with the differences in associations at this locus attributable to the skew of the phenotypes (§2.3.1): The deltamethrin phenotype is skewed towards mortality (mortality μ=0.79) and hence more likely to detect a minor allele associated with resistance, whereas the permethrin phenotypes are generally more skewed towards survival (mortality μ=0.24-0.52), making them more likely to detect a minor allele associated with susceptibility (fig. 4.3.7). In support of this hypothesis are the results that a transgenic line harboring a disruption of *Cyp6a17* is significantly more susceptible to both deltamethrin (Duneau *et al.* 2018) and permethrin (fig. 2.2.11E), demonstrating that the effect of *Cyp6a17* on phenotypes from both insecticides is in the same direction.

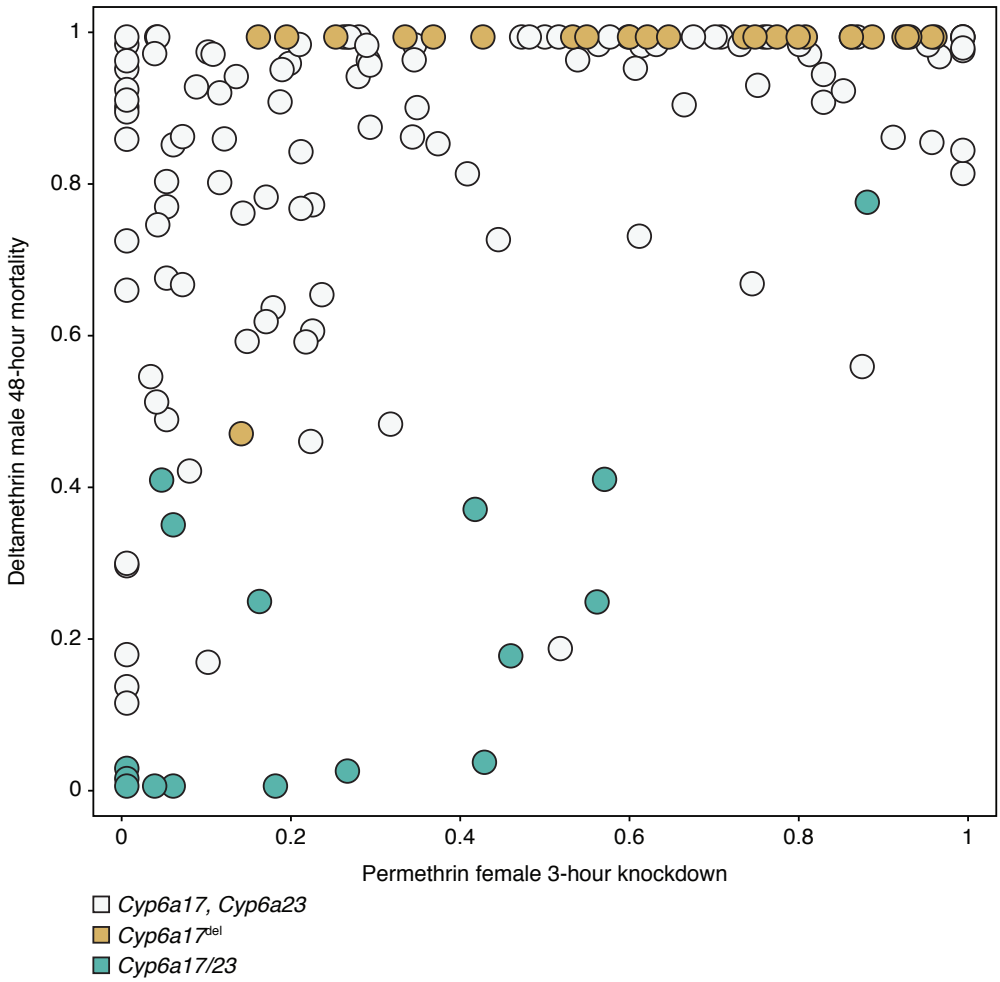


Figure 4.3.7. Comparison of pyrethroid phenotypes and *Cyp6a17/Cyp6a23* alleles
Scatter plot of deltamethrin male 48-hour mortality and permethrin female 3-hour knockdown phenotypes highlighting variation in structure at *Cyp6a17* and *Cyp6a23.*

Association studies utilizing triallelic DGRP variants also identify a triallelic nonsynonymous substitution in *Cyp6a17,* an allele of which is associated with both deltamethrin resistance (fig. 4.3.8A), and malathion resistance at 24 hours in males (fig. 4.3.8B). The associated G allele at this site is in strong linkage disequilibrium with the *Cyp6a17/23* allele (r^2=0.71; fig. 4.3.9A), and all G alleles are in the *Cyp6a17/23* allele background (fig. 4.3.8C), therefore the individual contributions of these alleles to resistance is unclear. However, these results suggest that variation in *Cyp6a17* and *Cyp6a23* may also be involved in resistance to malathion.

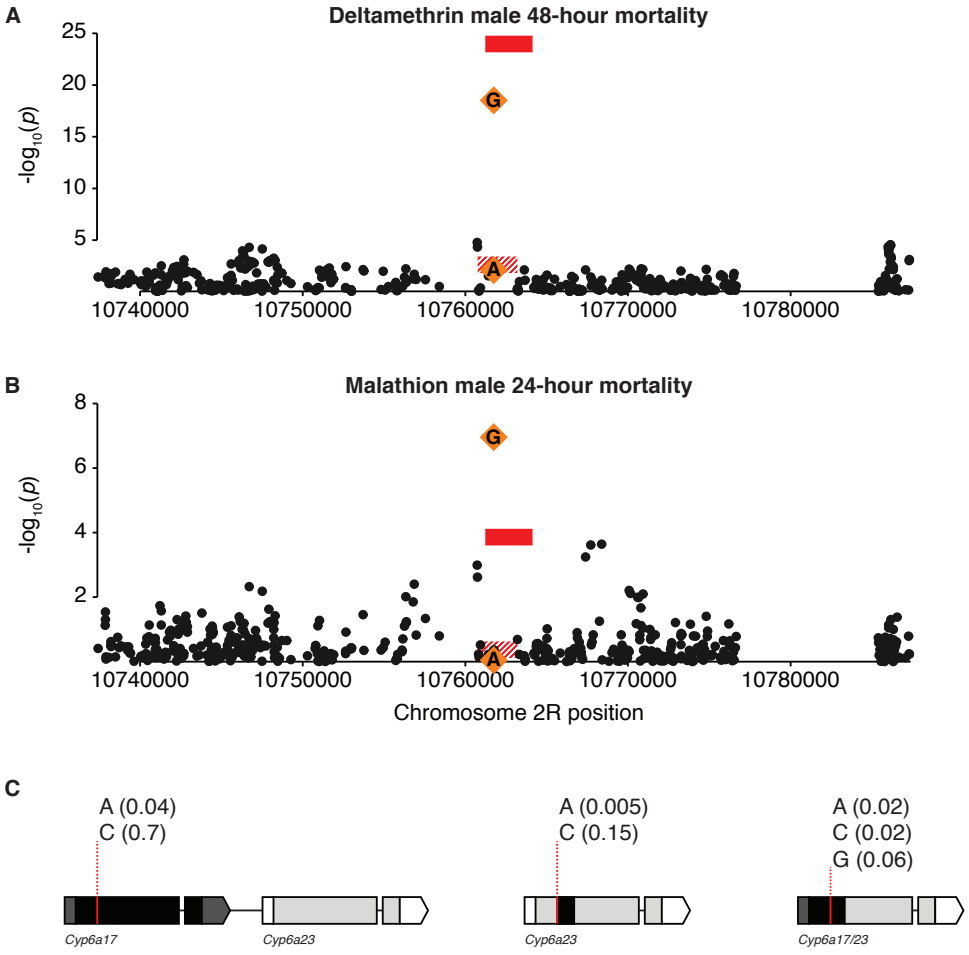


Figure 4.3.8. *Cyp6a17* triallele associations in context

Local Manhattan plots showing associations with annotated DGRP variants (black), *Cyp6a17/23* allele (solid red), *Cyp6a17*^{del} allele (striped red) and the *Cyp6a17* triallele (orange) for deltamethrin male 48-hour mortality (A) and malathion male 24-hour mortality (B). The overall DGRP frequency of the *Cyp6a17* triallele in each of the structural variant alleles is shown in C.

Duneau *et al.* (2018) identified evidence of selection using the H12 statistic in a region encompassing *Cyp6a17* and *Cyp6a23*. This signal was not detected in the H12 analysis performed by Garud *et al.* (2015) utilized in this chapter, and the difference is attributable to discrepancies in the DGRP lines used and the size of analysis windows in each of the analyses. However, linkage disequilibrium in the region suggests that the extended haplotype is associated with the *Cyp6a17*^{del} allele (fig. 4.3.9B) and not the *Cyp6a17/23* allele (fig. 4.3.9A). This therefore refutes the hypothesis that selection in this region is driven by insecticide resistance, as the *Cyp6a17*^{del} allele is associated with increased insecticide susceptibility. *Cyp6a17* has also been implicated in temperature

preference behavior (Kang, Kim and Choi 2011; Chakraborty *et al.* 2018) which in the light of these results appears to be a more likely explanation of the selection observed at this locus. Although apparently not driven by insecticide-based selection in the DGRP's progenitor population, *Cyp6a17* does appear to be a locus of major effect on pyrethroid phenotypes with derived alleles conferring both susceptibility and resistance in the DGRP.

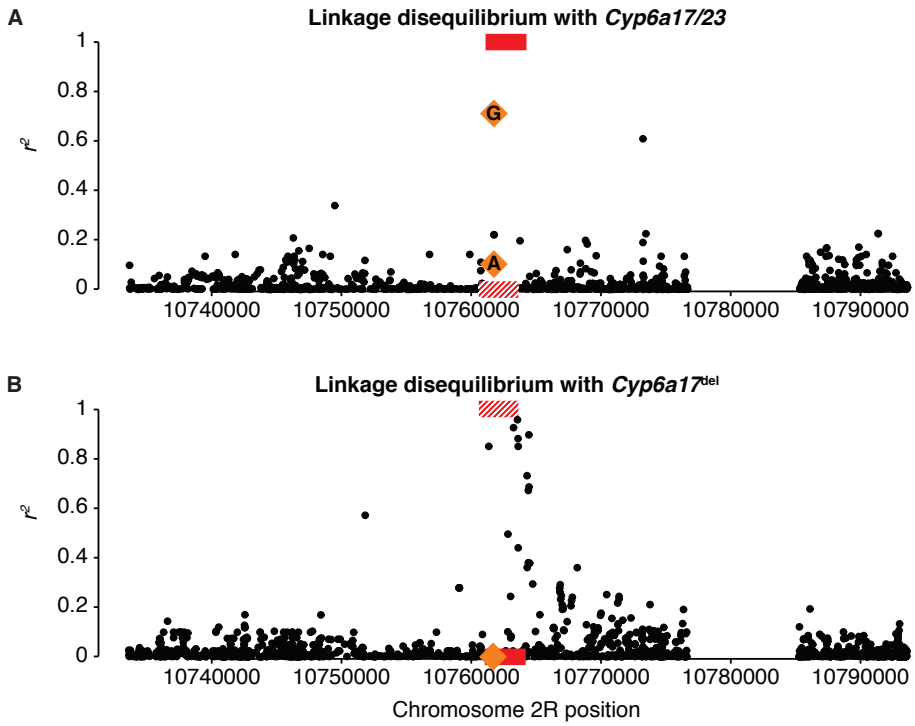


Figure 4.3.9. Linkage disequilibrium with *Cyp6a17/Cyp6a23* alleles

Linkage disequilibrium (r^2) of local variation with *Cyp6a17/23* allele (A; solid red) and *Cyp6a17*^{del} allele (B; striped red).

4.3.3.4. *Cyp6w1*

Another nonsynonymous triallelic site in a cytochrome P450 gene, *Cyp6w1*, was associated with resistance to DDT. Unlike *Cyp6a17,* structural variation involving *Cyp6w1* was not observed among DGRP lines, nor was variation in *Cyp6w1* transcript level associated with DDT phenotypes (appendix 2.4). Transgenic overexpression of *Cyp6w1* constructs containing each of the three states at this triallele revealed that one derived allele, *Cyp6w1*^{ALA}, conferred more resistance to DDT than the ancestral allele (*Cyp6w1*^{VAL}) and the other derived allele (*Cyp6w1*^{GLY}; Schmidt *et al.* 2017). The

association with *Cyp6w1* lies within a H12 selective sweep window (table 4.2.5), and near one of two local H12 maxima (fig 4.3.10). However, analysis of haplotype structure demonstrates that the selective sweep involves the derived allele that is *not* associated with resistance to DDT, *Cyp6w1*^{GLY} (fig 4.3.11). This is consistent with findings by Schmidt *et al.* (2017) using *i*HS and *n*S_L statistics, and suggests that although amino acid variation at *Cyp6w1* may be driving the selection observed in this region, DDT resistance is unlikely to be the pressure responsible. Selection for the DDT resistance allele *Cyp6w1*^{ALA} can, however, be inferred (albeit without statistical power) in *91-R,* as the allele is present in this DDT-selected line, while the ancestral *Cyp6w1*^{VAL} allele is present in the unselected control *91-C* (table 4.2.7).

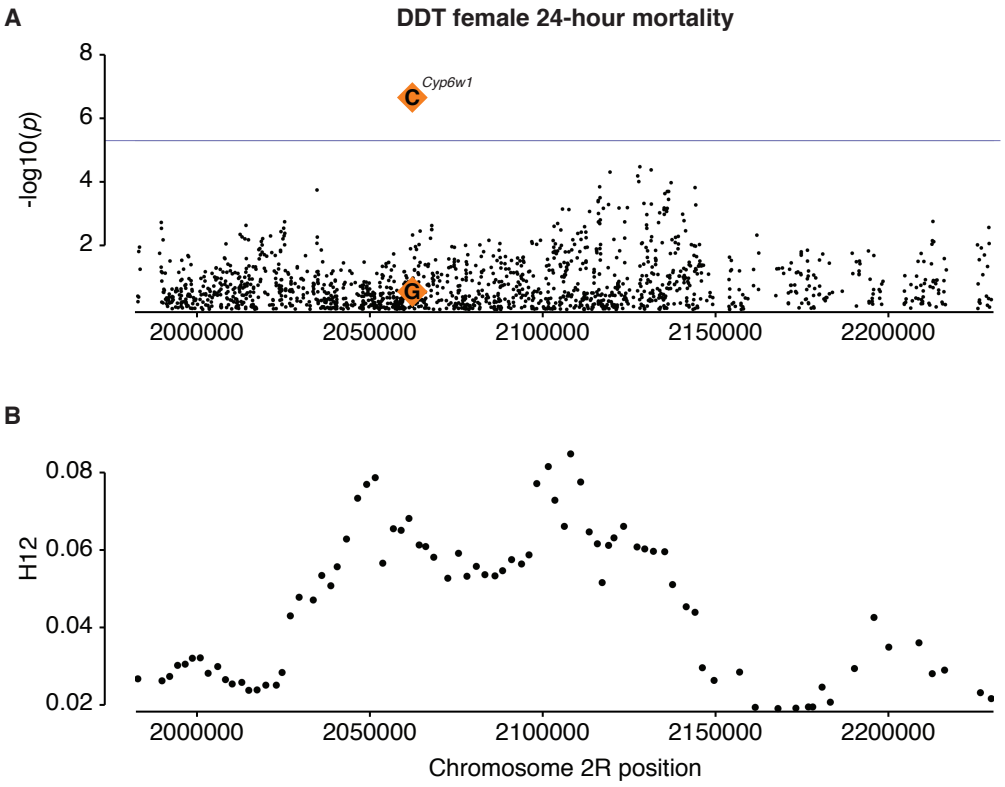


Figure 4.3.10. Evidence for selection at *Cyp6w1*

Comparison between DDT female 24-hour mortality phenotype associations (A) and local H12 statistic values (B) in the selective sweep peak encompassing *Cyp6w1*.

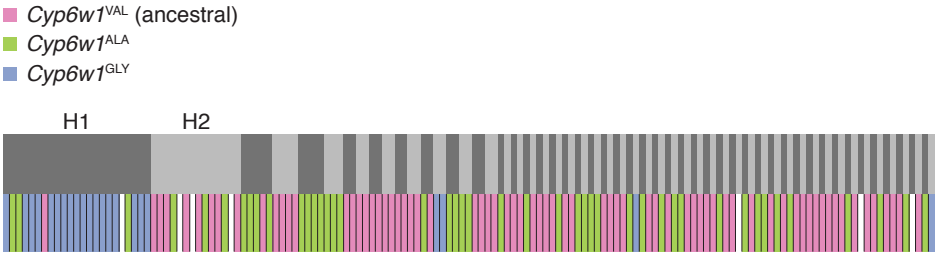


Figure 4.3.11. Resistance allele structure at the selective sweep including *Cyp6w1*
The distribution of the *Cyp6w1* triallele among haplotype groups in the selective sweep peak encompassing *Cyp6w1* (white blocks represent missing data).

4.3.3.5. Evidence for selection at other insecticide-associated loci

The preceding paragraphs consider the evidence for selection at four major insecticide-associated loci in the DGRP. To generalize this approach and assess the selective impact of insecticides on the genomes of the DGRP, all cases in which insecticide-associated DGRP variants overlapped with the 50 most pronounced signatures of selection based on the H12 statistic (Garud *et al.* 2015) were considered. After filtering out overlaps in which the alleles involved were not correlated with the two most common haplotypes at the sweep peaks, nine overlaps at seven peaks remained (table 4.2.5). Excluding the previously discussed sweep regions involving *Ace*, *Cyp6g1* and *Cyp6w1*, the remaining four overlaps do not show concordance between peak location and insecticide association (fig. 4.3.12; fig. 4.3.13), nor do *i*HS or nS_L statistics support selective signatures in insecticide-associated variants within these sweeps. This suggests that sweeps at *Ace* and *Cyp6g1* are the only DGRP H12 peaks associated with the insecticide phenotypes considered here. With the caveats that DGRP insecticide phenotypes are not exhaustive and selective sweep statistics are underpowered, these data suggest that insecticide-based selection in the DGRP's progenitor population is limited to two loci, and can be explained by a single class of insecticides, the organophosphates.

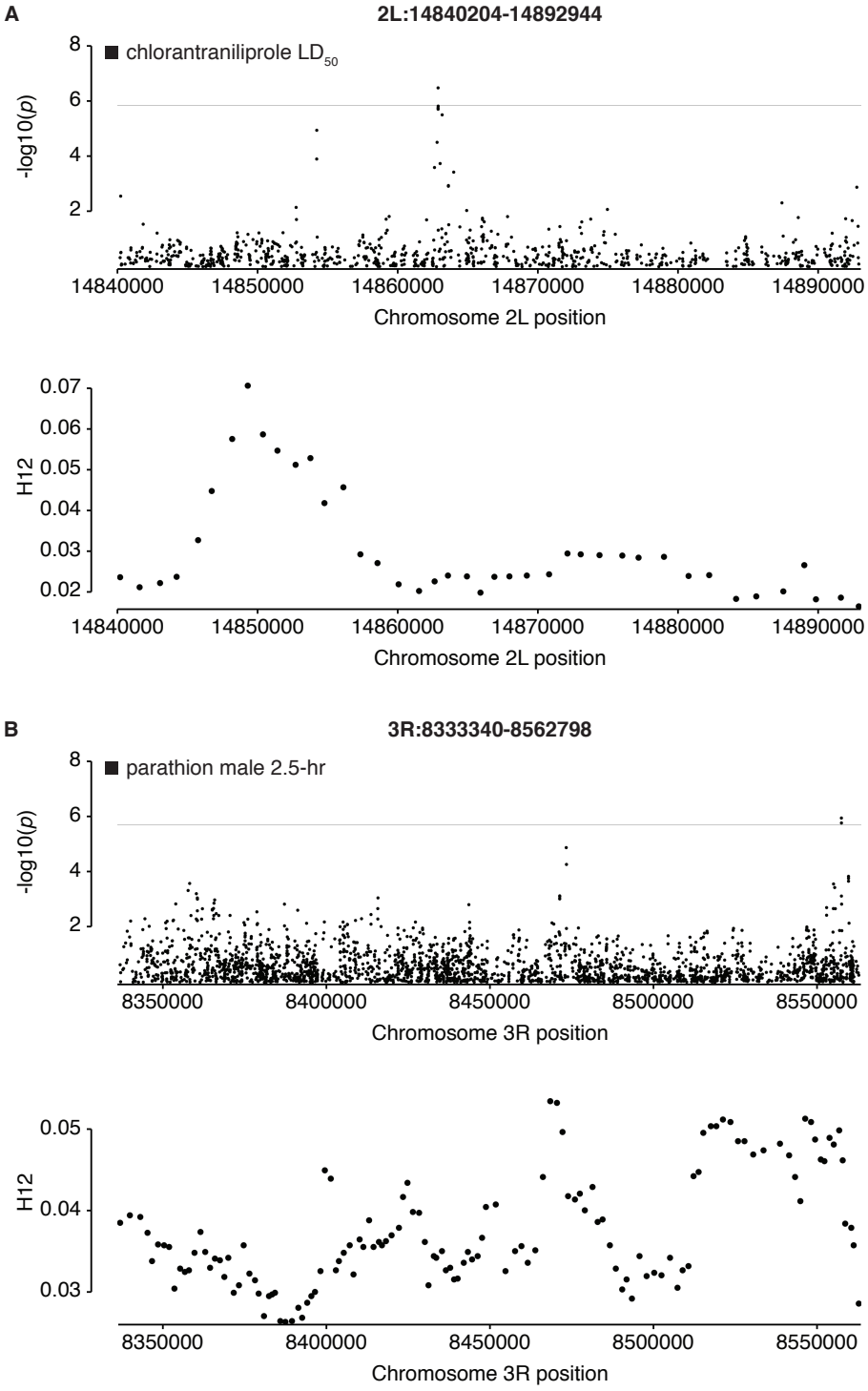


Figure 4.3.12. Lack of evidence for selection at other insecticide candidates

Comparison of significant phenotype associations and local H12 statistic values for selective sweep peaks encompassing insecticide-associated variants.

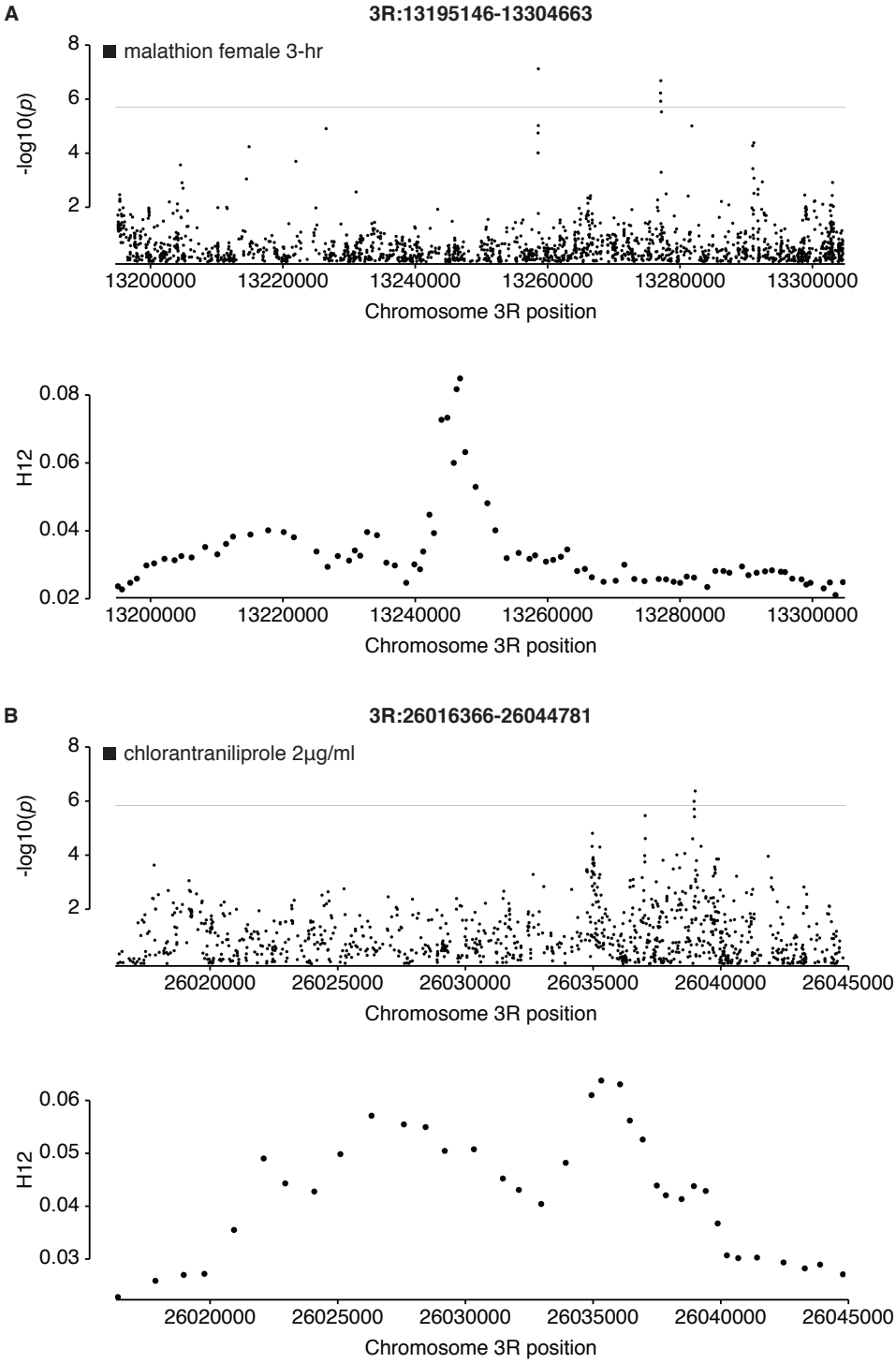


Figure 4.3.13. Lack of evidence for selection at other insecticide candidates

Comparison of significant phenotype associations and local H12 statistic values for selective sweep peaks encompassing insecticide-associated variants.

4.3.4. Inferences from populations outside the DGRP

While the studies of insecticide resistance utilizing association studies in the DGRP mapping panel have provided multiple insights into the quantitative architecture of insecticide phenotypes, a major question remains as to the applicability of these results to *D. melanogaster* in general, and more broadly, to other insect species targeted by insecticide use or suffering as collateral damage thereof.

At the level of selective signatures, the two DGRP sweeps attributable to insecticide-driven selection (*Cyp6g1* and *Ace*) are certainly not confined to the DGRP. Evidence for selection at the *Cyp6g1* locus has been reported in multiple populations of *D. melanogaster* (Catania *et al.* 2004; Schmidt *et al.* 2010; Kapun *et al.* 2018) and a parallel sweep has also been described in *Drosophila simulans* (Schlenke and Begun 2004; which also appears to be driven by an insertion of a transposable element in the 5' region of *Cyp6g1*). A sweep at *Ace* has likewise been identified outside the DGRP (Karasov, Messer and Petrov 2010) and in *D. simulans* (Signor, New and Nuzhdin 2017). Comparison of H12 peaks between the DGRP and a population sample from Zambia revealed *Cyp6g1* as the only common sweep among the top 25 peaks from each population (Garud and Petrov 2016; evidence for selection at *Ace* was observed in the Zambian population but this peak was ranked 35th). Detailed analysis by Garud and Petrov (2016) of *Ace* and *Cyp6g1* haplotypes in both populations reveal incomplete sharing of common haplotypes, suggesting that at least some of the resistance allele-bearing haplotypes in both sweeps share a common origin.

Utilizing a subset of the Drosophila Genome Nexus (Lack *et al.* 2015; DGN) comprising 15 population samples each with ten or more individuals, it was determined that fewer than 2.2% of the alleles associated with any insecticide were private to the DGRP (table 4.2.7), and analyses of populations individually revealed that this was not due to the contribution of alleles from a single, highly similar population (fig. 4.2.3B). Frequencies of top resistance candidate alleles at *Ace, Cyp6g1, Cyp6a17* and *Cyp6w1* were observed in most DGN populations, often at comparable frequencies to the DGRP (fig. 4.2.4). Most strikingly, derived *Cyp6g1* alleles are at extremely high frequencies in almost all DGN populations, consistent with an H12 peak observed around *Cyp6g1* in all four DGN populations scanned with the statistic (fig. 4.2.5). Similarly, evidence of an *Ace* sweep is observed in three of these four population scans. These results offer

insights into the history and demography of *D. melanogaster*: Presence of the vast majority of DGRP GWAS variants outside of the DGRP suggests that either this variation has been persisted through the original events of the species' diaspora, or it has been maintained by high gene flow between populations. In the specific cases of the insecticide-driven sweeps at *Ace* and *Cyp6g1*, the latter is most likely given the recent introduction of insecticides into the species' environment, and is supported by the identification of shared haplotypes at both sweeps in Zambia and Raleigh populations (Garud and Petrov 2016).

These results must be interpreted with caution, however, as they are subject to the limits of DGRP association studies, and hence exclude the effects of alleles of low frequency (which are expected to vary more between populations) and alleles below significance thresholds (which, under the omnigenic model may have a large cumulative effect). Epistatic interactions between alleles, which are common in complex traits in the DGRP (Huang *et al.* 2012), also complicate the inference of insecticide phenotypes in these other populations from DGRP-associated variants.

4.3.5. Conclusions

This chapter allows several conclusions to be drawn about the nature of insecticide resistance variation in the DGRP. Utilizing a relatively permissive significance threshold and 35 phenotypes from nine insecticides, insecticide-associated variants are generally distinct to phenotypes from a single compound. Exceptions to this are the variants involved in selective sweeps at *Ace* and *Cyp6g1*, which are associated with organophosphate insecticide phenotypes and represent the only selective sweeps in the DGRP attributable to any of the insecticide phenotypes investigated. Furthermore, insecticide-associated variants from the DGRP are observed in sequences from the DGN, with cases of DGRP-private variants being extremely rare, suggesting the results of these analyses are broadly applicable to populations worldwide.

The association studies performed in this chapter have identified a large number of candidates (appendix 1). Only a fraction of these have been discussed; even fewer have had their roles in insecticide phenotypes independently validated. The novel candidates identified here will hopefully inspire candidate gene studies and guide interpretation of datasets from future genome-wide analyses. Beyond the identification of specific candidates, this thesis makes global inferences about the genetic architecture of insecticide resistance; these are caveated by the power of the DGRP to map highly quantitative traits, which is limited by its sample size as well as the highly-inbred nature of the lines. It would be comforting to see the general patterns identified in this work, along with specific gene-level associations, replicated in studies with larger sample sizes and utilizing outbred individuals.

While the DGRP may be underpowered to map quantitative traits, the scans of selection applied in this work are certainly more so. The haplotype-based methods employed to provide evidence of selection on insecticide-associated alleles are sensitive only to recent adaptation that has left a large genomic footprint: It is unsurprising, therefore, that only two such footprints were linked to insecticides. Intriguingly, recent studies have been successful in identifying recent adaptation with more polygenic architecture by utilizing signatures of convergence (Yeaman et al. 2016) or temporal sampling (Alves *et al.* 2019).

4.4. Materials and methods

4.4.1. Genome-wide association studies

Thirty-five phenotype files from nine DGRP insecticide studies (table 4.2.1) were individually submitted to the Mackay lab DGRP2 GWAS pipeline (`http://dgrp2.gnets.ncsu.edu/`; Huang *et al.* 2014). The genome wide significance threshold (1×10^{-5}) was corrected for the number of phenotypes tested for each insecticide and applied to the Mackay lab pipeline 'mixed *p*-value' (association after correction of the phenotype for the effects of Wolbachia and major chromosomal inversions).

4.4.2. Associations with multiallelic variants

The 201 DGRP BAM files available from the Baylor College of Medicine website processed into a VCF file using the `mpileup` command in SAMtools (Li *et al.* 2009). This VCF file was filtered for sites with a minimum of three alleles and a minimum minor allele frequency of 0.05 using VCFtools (Danecek *et al.* 2011). Heterozygous sites were masked to NA and sites with greater than 20% missing data were filtered out, resulting in 15,487 triallelic sites for further analysis (no remaining site had greater than three alleles). For each triallele and insecticide phenotype, ANOVAs were performed in R (R Core Team 2018) using the `aov` function.

4.4.3. DGRP phenome assembly

One hundred and thirteen phenotypes were collected from the references in table 4.2.4 and aligned by DGRP line using the `merge` function in R (R Core Team 2018).

4.4.4. Selective sweep statistics

The top 50 signals of selection among DGRP genomes were obtained from Garud *et al.* (2015). DGRP-line breakdowns of haplotype structure at each peak from the original scane were obtained from Nandita Garud (pers. comm.). To plot local H12 for figures, scripts described in Garud *et al.* (2015) were run on DGRP data for 140 lines (the 145 analyzed by Garud *et al.* [2015] excluding five lines missing from DGRP freeze 2) with the following parameters: `-w 400 -j 50 -d 0`

For DGN populations, sites with > 2 alleles, and sites with > 30% missing data were removed, and scripts described in Garud *et al.* (2015) were run on data for each population separately using the same parameters as the DGRP scan. Values within heterochromatic regions were masked from the scan using estimates from Fiston-Lavier *et al.* (2010).

Genome-wide *i*HS and nS_L values for the DGRP were calculated by Joshua Schmidt (pers. comm) using the selscan program (Szpiech and Hernandez 2014) with default parameters. A detailed description of the approach is published in Schmidt *et al.* (2017).

4.4.5. Analysis of *91-R* and *91-C* genomes

BAM files containing alignments of *91-R* and *91-C* sequencing reads to the *y; cn bw sp;* reference genome were obtained from the Sequence Read Archive (Leinonen *et al.* 2010; `SRR1237973`; `SRR1237974`) and variants were called using the `call` command in BCFtools (Li *et al.* 2009).

4.4.6. Worldwide population data

FASTA sequences from Drosophila Genome Nexus (DGN) version 1.1 (Lack *et al.* 2015) were obtained from `http://www.johnpool.net/genomes.html` along with scripts for masking population admixture and identity by descent regions, which were applied to the FASTA files before further analysis. A modified version of the dataslice.pl script, also obtained from `http://www.johnpool.net/genomes.html`, was used to extract genotypes at sites of interest across the DGN genomes.

To call structural variants and the triallelic site in *Cyp6w1* in DGN genomes, FASTQ files for DGN lines were obtained from the Sequence Read Archive (Leinonen *et al.* 2010) and aligned to 40kb sections of the reference genome surrounding each of *Cyp6g1*, *Cyp6a17* and *Cyp6w1* using the `mem` command in Burrows-Wheeler Aligner (Li and Durbin 2009). The resulting BAM files were visualized with IGV software (Thorvaldsdóttir, Robinson, and Mesirov 2013) and manually scored for variants of interest.

5

General discussion

Studies of insecticide resistance in the Drosophila Genetic Reference Panel have provided multiple insights into signatures of strong, recent selection in the genome of this model organism. Based on genome-wide scans of the DGRP population, two of the most prominent peaks could be clearly attributed to insecticide-based selection, each one driven by variants displaying highly-significant GWAS associations with phenotypes from organophosphate-class insecticides. However, in addition these sweeps at *Ace* and *Cyp6g1*, work described in this thesis has contributed to the characterization of a further three selective sweep peaks with putative foci at *CHKov1*, *Cyp6w1* and *Cyp6a17*. Interestingly, all five proposed selective sweep foci are complex, exhibiting a combination of multiallelism, soft sweep architecture and structural variation. This chapter seeks to review the evidence for selection in each of these five regions, infer the adaptive forces shaping them, and ascertain what this information can teach us about recent evolution in *D. melanogaster*.

5.1. Geometric adaptation and selective sweeps

Although Fisher's (1918) model of the genetic architecture of traits has been largely supported by modern quantitative genetics (Boyle, Li and Pritchard 2017; Visscher and Goddard 2019), his geometric model of adaptive evolution is incongruent with the evidence for selection at large-effect loci provided by selective sweeps. Fisher found that the likelihood of a genetic factor improving fitness rapidly decreased as its effect size increased, and thus reasoned that evolution would act on changes with extremely small effect sizes.

The existence of adaptive alleles of large effect can be explained by the 'adaptive walk' model of evolution (reviewed in Orr 2005), whereby a population gradually approaches a phenotypic optimum through successive mutations. Under this model, the first evolutionary step may be large, and the effect of this step on the 'fitness landscape' (Wright 1932) is then fine-tuned by steps of smaller effect, the distribution of which would ultimately be exponential (Orr 1998).

As statistical methods for detecting selective sweeps have been refined, the relative contribution of soft selective sweeps to adaptation has been debated, with studies suggesting they are a significant (Prtichard, Pickrell and Coop 2010) or even dominant (Messer and Petrov 2013) force in the architecture of adaptation. Intriguingly, an adaptive walk at a single locus should produce similar patterns to soft sweeps, as each new step in the walk will begin its own divergent haplotype defined by a new causal variant (fig. 5.1.1).

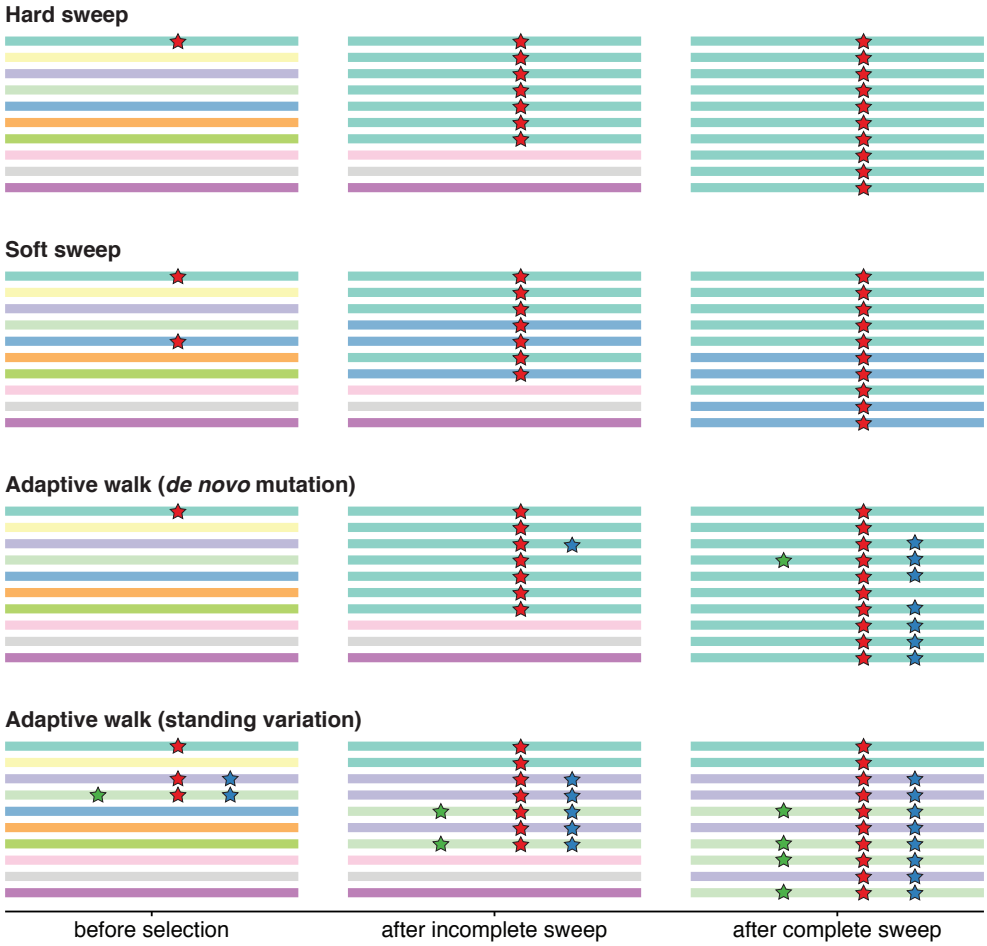


Figure 5.1.1. Adaptive walk models of haplotype structure in selective sweeps

Models of haplotype structure produced by selective sweeps, including adaptive walks. Shown before selection, after an incomplete sweep, and a complete sweep. Distinct haplotypes are represented by coloured bars, with the positively selected mutation(s) represented by stars.

5.2. *Acetylcholinesterase*

Examination of the DGRP haplotype structure at *Ace* shows clear signs of an adaptive walk at least partially explaining the soft sweep architecture, with the most common haplotype at the *Ace* peak almost exclusively composed of lines with one or three *Ace* resistance substitutions, whereas the second most common haplotype is entirely composed of lines with three *Ace* substitutions (fig. 4.3.4). Interestingly, analysis of *Ace* alleles at a worldwide scale reveals at least one adaptive walk has occurred, along with multiple soft sweeps from standing variation: Shared haplotypes surrounding three-substitution *Ace* alleles have been observed between populations from North America and Australia (Karasov *et al.* 2010), and between Raleigh and Zambia populations (Garud and Petrov 2016), while population-specific haplotypes have also been described in each case.

While the adaptive walk at *Ace* is almost certainly driven in part by increases in insecticide insensitivity, it may also be influenced by a relative fitness advantage (in the context of the endogenous function of Ace) of multiple substitutions in the same enzyme: Shi *et al.* (2004) demonstrated that certain combinations of *Ace* substitutions reduce the enzyme's catalytic activity towards its native substrate, while others (including the three-substitution Ace^{VAYG} allele observed in the DGRP) maintain ancestral levels. In contrast, increases in the number of Ace substitutions reduce enzymatic stability (Shi *et al.* 2004), however this appears to be compensated for by increased production of the enzyme, as Ace abundance (up to a 100% increase) is positively correlated with both parathion LD_{50} and enzymatic inhibition of paraoxon in a worldwide sample of 82 *D. melanogaster* samples (Charpentier and Fournier 2001). It remains unknown whether increased Ace abundance is a direct result of the resistance substitutions, or due to an additional step in the adaptive walk, however gene duplication events which include *Ace*, described in multiple populations, make intriguing candidates (Turner *et al.* 2008; Schrider *et al.* 2016).

Duplications involving *Ace* have been described in *D. melanogaster* population samples from Australia and the USA, with distinct alleles significantly differentiated between tropical and temperate climates on each continent (fig. 5.2.1). Interrogation of DGRP data cannot inform this issue; only a single line appears to carry a gene duplication

including *Ace*, (fig. 5.2.1) making the polymorphism practicably absent for any quantitative genetics analysis. Huang *et al.* (2015) mapped multiple *cis*- and *trans*-eQTL of *Ace*, however none of these were associated with any organophosphate phenotype (although two *cis*-veQTL and one *trans*-veQTL of *Ace* were associated with parathion; appendix 1).

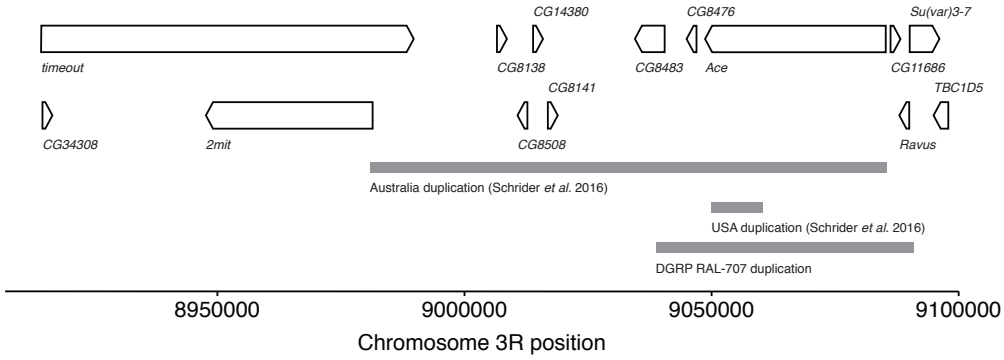


Figure 5.2.1. Duplications involving *Ace*

Duplications involving *Ace* as described by Schrider et al. (2016), which show significantly different frequencies in tropical and temperate populations in Australia and the USA, along with the *Ace* duplication observed in a single DGRP line, RAL-707.

5.3. *Cyp6g1*

A region containing *Cyp6g1* represents the strongest signature of selection, genome-wide, in the DGRP population (Garud *et al.* 2015), and evidence for selection on this region has been identified in multiple other populations worldwide (Schmidt *et al.* 2010; Kapun *et al.* 2018, fig. 4.2.5), consistent with high derived allele frequencies observed in worldwide samples (fig. 4.2.4). Among DGRP phenotypes, strong associations were observed between organophosphate insecticides and *Cyp6g1*, both at the level of genomic variation and transcript-level variation, while *Cyp6g1* transcript level was also associated with imidacloprid phenotypes. A strong association between DDT resistance and *Cyp6g1* has previously been identified in Australia (Schmidt *et al.* 2010), however this result was not replicated in the DGRP (Schmidt *et al.* 2017), potentially due to the greater allelic diversity in the Australian sample.

Analyses of DGRP data performed in this thesis have identified additional insights into the allelic variation at the *Cyp6g1* locus. Two duplicated regions, identified by Schmidt *et al.* (2010) in all derived *Cyp6g1* alleles, vary substantially in copy number among DGRP lines, and copy number of each region is correlated with increased *Cyp6g1* transcript level (fig. 3.2.3). Moreover, *cis*-acting variation correlated with increased *Cyp6g1* transcript levels is also correlated with transcript levels of *Cyp6g1*'s tandem paralog, *Cyp6g2* (appendix 2), which is also capable of conferring resistance to the insecticides diazinon, nitenpyram and imidacloprid when overexpressed (Daborn *et al.* 2007; Denecke *et al.* 2017), and is significantly correlated with azinphos-methyl phenotypes in the DGRP (appendix 1.1).

These results demonstrate a greater allelic diversity among the *Cyp6g1* alleles under selection, but they are not exhaustive; the methods used to genotype structural variation at the *Cyp6g1* locus in the DGRP are blind to any polymorphism within the known transposable element insertions. Additional insights into structural variation at *Cyp6g1* are provided by the first long-read sequences of lines carrying derived alleles at the locus (Chakraborty *et al.* 2019; fig. 5.3.1). Two DSPR founder strains, *A6* (collected in Georgia, USA in 1966) and *B4* (collected in California, USA in 1963) exhibit the *Accord* insertion and tandem duplication events indicative of derived *Cyp6g1* alleles. The repeated units in these lines appear to be homologous to those identified in the

DGRP and other populations (fig. 2.2.6), and as in the DGRP, they vary in copy number. Furthermore, variation in the size of *Accord* insertions is observed between copies, and an insertion of a *Gypsy* element is described in *B4*.

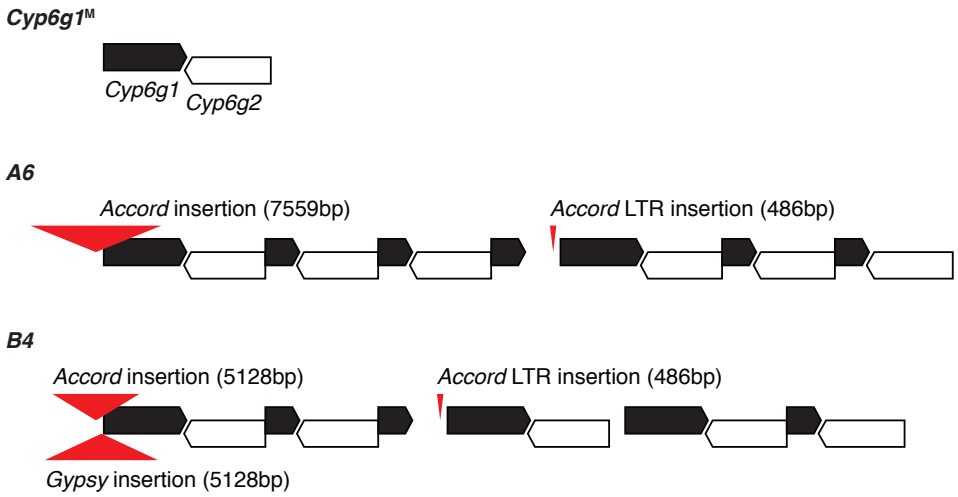


Figure 5.3.1. Structure of derived *Cyp6g1* alleles in the DSPR (after Chakraborty *et al.* 2019)
Gene amplifications and transposable element insertions at *Cyp6g1* inferred from long-read sequencing of DSPR founder lines *A6* and *B4*.

An adaptive walk of successive transposable element insertions in Australian populations has previously been described at *Cyp6g1* (Schmidt *et al.* 2010; fig. 2.2.6), with successive allelic steps associated with increases in *Cyp6g1* transcript abundance and DDT resistance. The ultimate step in this walk, $Cyp6g1^{BP}$, has only been described in samples from Australia and India (Catania *et al.* 2005; Schmidt *et al.* 2010), and is absent from the DGRP. The other derived alleles, $Cyp6g1^{AA}$ and $Cyp6g1^{BA}$, however, are at high frequencies in the DGRP, and signatures of selection in the region surrounding *Cyp6g1* are consistent with a soft sweep and an adaptive walk: The most common haplotype group at the peak is almost entirely composed of lines carrying $Cyp6g1^{AA}$ alleles, the second most common haplotype is almost entirely composed of lines carrying $Cyp6g1^{BA}$ alleles, and lines carrying the ancestral $Cyp6g1^{M}$ allele have rare haplotypes at the peak (fig. 4.3.6).

However, the softness of the *Cyp6g1* sweep is not entirely explained by an adaptive walk, as both $Cyp6g1^{AA}$ and $Cyp6g1^{BA}$ alleles appear on multiple DGRP haplotypes

(fig. 4.3.6). Due to the similarity of the *Cyp6g1* alleles between populations from Australia and the USA, the rare nature of these mutational events, and the nested transposable insertions, it is almost certain that the *Cyp6g1*^{AA} and *Cyp6g1*^{BA} alleles arose along a single lineage, which suggests that both alleles were swept from standing variation. This hypothesis is reinforced by the fact that the sizes of selective sweep footprints surrounding *Cyp6g1*^{AA} and *Cyp6g1*^{BA} alleles in the DGRP are not significantly different (Schmidt 2014).

It appears that *Cyp6g1*^{AA} and *Cyp6g1*^{BA} alleles predate modern insecticides, and the frequency of these alleles has risen rapidly in response to widespread modern insecticide use. By the early 1950's, lines carrying *Cyp6g1*^{BA} alleles had been established from sampling on two separate continents (*91-R*, established in 1952 from St Paul, Minnesota, USA [Merrell and Underhill 1956]; *Hikone-R*, established 1952 from Hikone, Shiga, Japan [Oshima 1958]). It is unclear whether derived *Cyp6g1* alleles were present at low frequencies in *D. melanogaster* populations worldwide before selection by insecticides (although Schmidt *et al.* [2010] found no evidence of derived alleles in six samples collected in the 1930's), or if insecticide-driven selection facilitated increased gene flow at the locus from a single population. Nevertheless, it appears that these alleles were maintained in at least one population at a frequency and timescale that allowed soft sweeps from standing variation to occur at both *Cyp6g1*^{AA} and *Cyp6g1*^{BA}.

One explanation for this is that other less pervasive xenobiotic compounds maintained derived *Cyp6g1* alleles in populations. This hypothesis is supported by *Cyp6g1*'s wide substrate specificity (Daborn *et al.* 2007; fig. 2.2.11). Particularly, *Cyp6g1* overexpression confers resistance to nicotine (Li, Bai and Cass 2012), which is produced by a range of plants, and has been used as an insecticide since at least the 1690s (Tomizawa and Casida 2005).

While another xenobiotic compound makes a compelling candidate for the initial maintenance of derived *Cyp6g1* alleles, an alternative selective agent comes from evidence that a *Cyp6g1*^{BA} allele introgressed from DDT-resistant line *Hikone-R* into an insecticide-susceptible genetic background (*Canton-S*) confers sexually antagonistic fitness effects. Females carrying a *Cyp6g1*^{BA} allele receive a fitness benefit in the form

of an apparent maternal effect increasing egg and larval survival (McCart, Buckling and Ffrench-Constant 2005), while males with the allele suffer a fitness cost in the form of reduced size, and lower metrics of courtship success and aggression (Rostant *et al.* 2017). However, interrogation of existing DGRP data does not support these findings: male aggression (Shorter *et al.* 2015) and female fecundity (Durham *et al.* 2014) phenotypes are not associated with either transposable element insertion at *Cyp6g1* ($p>3.4\times10^{-1}$ for each pairwise comparison). Moreover, the strategy to introgress the *Cyp6g1*^{BA} allele from *Hikone-R* into *Canton-S* (McCart, Buckling and Ffrench-Constant 2005) utilized DDT resistance as the marker for the allele, which could have resulted in multiple factors contributing to DDT resistance being crossed into the resulting line.

5.4. *Cyp6a17* and *Cyp6a23*

Associations with phenotypes from two pyrethroid insecticides in the region of the cytochrome P450 cluster on the right arm of chromosome two implicate structural variation involving tandem paralogs *Cyp6a17* and *Cyp6a23* in these insecticide phenotypes. Although not among Garud *et al.*'s (2015) top 50 peaks of selection, a similar use of the H12 statistic with different parameters (specifically DGRP line choice and statistic window size; Duneau *et al.* 2018) identified evidence for selection in this region. In addition to the reference genome arrangement of these genes, two common structural variants exist in the DGRP, both derived from relative deletion events producing a single chimeric replacement of *Cyp6a17* and *Cyp6a23* (Good *et al.* 2014; fig. 2.2.9). Both these alleles appear to be derived states, as the gene duplication leading to *Cyp6a17* and *Cyp6a23* seems to be at least as old as the divergence between *D. melanogaster* and *Drosophila ananassae*.

Unlike those observed at *Cyp6g1*, structural variants involving *Cyp6a17* and *Cyp6a23* do not appear to be successive mutations, as they do not share a common breakpoint (fig. 2.2.9). A non-homologous recombination event between *Cyp6a17* and *Cyp6a23* likely produced the *Cyp6a17/23* allele, however multiple events must be invoked in the formation of the *Cyp6a17*^{del} allele. Also contrasting the derived alleles at *Ace* and *Cyp6g1*, evidence for selection is only observed for one derived allele, *Cyp6a17*^{del} (fig. 4.3.9), which is not associated with an adaptive insecticide phenotype: While the *Cyp6a17/23* appears to increase pyrethroid resistance, *Cyp6a17*^{del} increases susceptibility (fig. 3.3.2, fig. 4.3.7). Further complicating inferences from these alleles is a nonsynonymous, triallelic site which is not in complete linkage disequilibrium with the structural variation (fig. 4.3.8).

Selection in the DGRP's progenitor population at *Cyp6a17* and *Cyp6a23* does not appear attributable to the use of insecticides, however these genes have also been linked to climate adaptation. *Cyp6a17* and *Cyp6a23* show sequence differentiation between temperate and subtropical populations in parallel in both Australia and the USA (Turner *et al.* 2008), and structural variation appears to be involved (Schrider, Hahn and Begun 2016). At the transcript level, *Cyp6a17* and *Cyp6a23* also display significant geographic origin-dependent plasticity, with both genes responding transcriptionally to

temperature in Australian tropical population samples but not temperate ones (Levine, Eckert and Begun 2011). Interestingly, these responses were opposing, with *Cyp6a17* transcription positively correlated with rearing temperature and *Cyp6a23* transcription negatively correlated. At the phenotypic level, a temperature preference behavior has been linked to *Cyp6a17*. Flies with *Cyp6a17* disrupted showed reduced avoidance of low temperatures relative to a genetic background control (Kang, Kim and Choi 2011). This result was supported by Chakraborty *et al.* (2018), who demonstrated that flies from the DSPR founder line *A4*, which carries a *Cyp6a17*^{del} allele, prefer colder temperatures than w^{1118} flies which have an intact *Cyp6a17* gene. These results suggest that selection on the *Cyp6a17*^{del} allele is potentially driven by climate adaptation, but not pyrethroid resistance.

5.5. *Cyp6w1*

A single, triallelic, nonsynonymous substitution in *Cyp6w1* is associated with resistance to DDT, and a signal of strong, recent selection at this locus is also proposed by multiple selective sweep statistics (Schmidt *et al.* 2017; table 4.2.6). However, like *Cyp6a17* and *Cyp6a23*, the hypothesis of insecticide-driven selection at this locus is not supported, as only one of the derived states at this site, *Cyp6w1*^{GLY}, shows evidence of selection (Schmidt *et al.* 2017; fig. 4.3.11), while the other derived allele, *Cyp6w1*^{ALA}, confers resistance to DDT (Schmidt *et al.* 2017).

The *Cyp6w1* triallele at amino acid site 370 of Cyp6w1 is predicted to occur in Substrate Recognition Domain SRS5 in the canonical cytochrome P450 model (Gotoh 1992, Zawaira *et al.* 2011), adding credibility to the findings that a mutation at this position increases the affinity of the enzyme to a particular substrate (Schmidt *et al.* 2017). While the *Cyp6w1*^{ALA} allele appears to increase affinity to DDT, this suggests that the selective sweep observed at the *Cyp6w1*^{GLY} allele in the DGRP is driven by increasing *Cyp6w1*'s affinity to an as yet undetermined compound, potentially an insecticide. Furthermore, allele frequencies at this site are quite variable across worldwide population data (Schmidt *et al.* 2017; fig. 4.2.4), suggesting different alleles may be favored by selection in different populations.

5.6. *CHKov1*

The strong selective sweep at *CHKov1* was initially identified in a screen of long interspersed element-like *Doc* transposable element insertions identified in the *D. melanogaster* reference genome and present at a high frequency worldwide. From this screen, Aminetzach, Macpherson and Petrov (2005) identified a *Doc* insertion into the coding region of a gene they named *CHKov1* as putatively adaptive in the expansion of *D. melanogaster* out of Africa. Evidence of a selective sweep around $CHKov1^{Doc}$ allele was also identified by Aminetzach, Macpherson and Petrov (2005), and replicated at a genome-wide level among DGRP lines by Garud *et al.* (2015).

The $CHKov1^{Doc}$ allele introduces a substantial change to the ancestral allele, encoding a chimeric protein composed of both ancestral sequence and a translation of the *Doc* element. Aminetzach, Macpherson and Petrov (2005) initially speculated that an organophosphate insecticide may be driving the selection at *CHKov1*, and this was supported by the result that a single line with different *CHKov1* alleles introgressed showed increased resistance to azinphos-methyl attributable to the $CHKov1^{Doc}$ allele. However, none of the DGRP population level azinphos-methyl phenotypes (or any other insecticide phenotype) described in this thesis show any association with $CHKov1^{Doc}$ presence, making it unlikely that insecticide resistance is driving the sweep at *CHKov1*.

In 2011, Magwire *et al.* proposed an alternative driver for the selection at *CHKov1*: resistance to the sigma virus (Rhabdoviridae). Recombination mapping identified a region including *CHKov1* as the previously described *ref(3)D* sigma virus-resistance factor, and a subsequent GWAS of sigma virus resistance in the DGRP revealed the $CHKov1^{Doc}$ allele as the top candidate. Magwire *et al.* (2011) also identified multiple alleles at *CHKov1* showing evidence of an adaptive walk. Among DGRP lines, resistance is associated with presence of the $CHKov1^{Doc}$ allele, however the allele from Magwire *et al.*'s (2011) initial mapping was found to be a complex rearrangement of the $CHKov1^{Doc}$ allele, including at least two duplication events (fig. 5.6.1). The 'susceptible' line in the initial mapping experiment carried $CHKov1^{Doc}$, indicating that this additional step in the adaptive walk conferred an additional increase in sigma virus resistance relative to the $CHKov1^{Doc}$ allele. Additionally, Aminetzach, Macpherson and

Petrov (2005) described seven nonsynonymous substitutions that are fixed among *CHKov1*^{Doc} alleles, which appear to represent additional allelic steps following the *Doc* insertion. This is partially supported by DGRP data; presence of the *Doc* insertion in *CHKov1* is indeed correlated with seven nonsynonymous substitutions in the gene (r^2>0.1), four of which are upstream of the *Doc* insertion and three of which are downstream. However, the substitutions are not fixed in *CHKov1*^{Doc} alleles, with r^2 between the *Doc* insertion and each SNP ranging from 0.22 to 0.33.

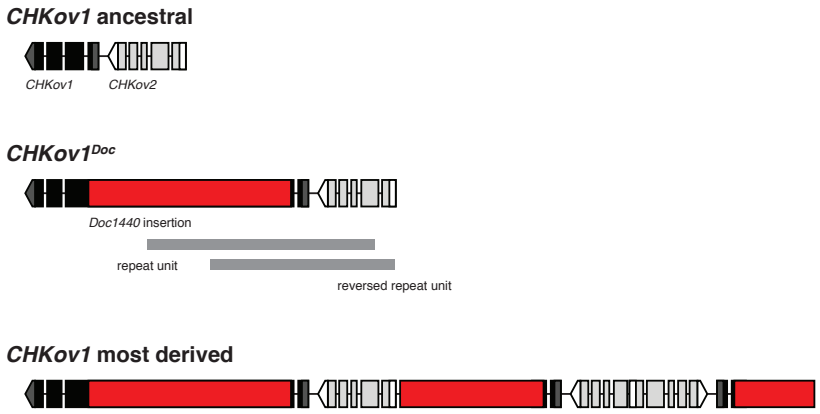


Figure 5.6.1. Structurally variant alleles of *CHKov1* (after Magwire *et al.* 2011)
Structurally variant alleles described at the *CHKov1* locus.

Although demonstrated to be absent from the DGRP (Magwire *et al.* 2011), the frequency of the most derived, most resistant allele at *CHKov1* is undetermined in other populations. The complexity of the variation at this allele makes its detection from population samples nontrivial, however given Magwire *et al*'s. (2011) *ref(3)D* strain was originally isolated in Paris, and there is evidence of a selective sweep at *CHKov1* in the Lyon population from the DGN (fig. 4.2.4), perhaps population samples from the DrosEU (Kapun *et al.* 2018) are a good place to start.

5.7. Conclusions

A detailed analysis of five top selective sweep candidate regions from the DGRP population has revealed substantial evidence for multiallelism and adaptive walks. Interestingly, most of the putatively focal alleles at these sites would preclude detection by standard paired-end read genotyping workflows, and are only characterized due to the large amount of research interest directed at these regions encouraged by their dramatic phenotypes and selective signatures. For this reason, it remains an open question as to whether complex allele structures are a common feature of highly adaptive regions, or whether such alleles are prevalent across the genome and currently remain hidden due to lack of focused interest. However, recent work in *D. melanogaster* suggests that a significant proportion of genome structural variation cannot even be inferred, let alone characterized, without the use of long-read sequencing technologies (Chakraborty *et al.* 2018; Chakraborty *et al.* 2019).

Instances of pleiotropy have been described at sweep candidates, for example the effect of *Ace* alleles on endogenous enzyme function, the putative sexually antagonistic effects of $Cyp6g1^{BA}$, and the involvement of *Cyp6a17* in temperature preference. Although limited data points are available it seems that a negative correlation between fitness and insecticide resistance is an oversimplification of complex allele-specific pleiotropic effects on a range of potentially adaptive phenotypes. Such patterns of pleiotropy are unlikely to be specific to targets of selection; as traits are understood to be more quantitative, they are necessarily more codependent on the same genetic variants and hence pleiotropy is predicted to be rampant.

It remains unknown how much of the recent selection in *D. melanogaster* can be explained by selective sweeps. Garud *et al.* (2015) identified 50 significant sweep signatures in their H12 analysis of the DGRP population, however it has been argued that this statistic cannot distinguish true selective sweeps from neutral patterns in the DGRP data when a more rigorous demographic model is used (Harris, Sackman and Jensen 2018). This criticism is levelled at selective sweep statistics more generally, along with the fact that they are known to be underpowered in detecting true positives (Pavlidis and Alachiotis 2017). Pritchard, Pickrell and Coop (2010) suggest that characterization of phenotypes and selective agents for putative sweeps is an important

tool in validating true positives. As work in this thesis has demonstrated, the DGRP provides an unparalleled system for such experiments. The importance of this validation has been underscored by significant departures from expectations observed in DGRP validations of putative selective pressures. For example, *Cyp6g1* sweeps worldwide were assumed to be driven by DDT selection, whereas work described here has shown that, at least in the DGRP population, organophosphate insecticides are a much more likely driver of selection at this locus.

Nevertheless, signatures of intense selection at *Ace* and *Cyp6g1* remain attributable to insecticide use, and it is of great concern that the two most pronounced signatures of selection in a non-pest species would be caused by collateral exposure to insecticides. As sequencing projects in non-model insects intensify, the applicability of these findings in *D. melanogaster* may be tested for similarities in other species at the population genomics scale. Recently, the *Anopheles gambiae* 1000 Genomes Consortium identified at least three insecticide-resistance associated regions showing selective sweeps in African population samples (The *Anopheles gambiae* 1000 Genomes Consortium, 2017). These include *Vssc*, the molecular target of DDT and pyrethroid insecticides, and a cluster of *Cyp6p* genes, the latter showing multiple structurally variant alleles which appear to be the foci of the sweep (Lucas *et al.* 2019). These findings from the *A. gambiae* complex and the results from the DGRP show compelling similarities.

In this thesis, investigation of the genetic basis of insecticide resistance, a classical model of microevolution, has led to the characterization of multiple selective sweep signals. These sweeps provide evidence that the evolution of quantitative traits can involve alleles of large effect, but it appears that models of hard sweeps around single nucleotide polymorphisms fail to capture the complexity and diversity observed in the *D. melanogaster* sweeps considered here. Instead, it appears that adaptation resulting from strong, recent selection is characterized by multiallelic, often sequential loci which exhibit complex structural variation.

References

Akhund-Zade, J., Bergland, A.O., Crowe, S.O. and Unckless, R.L., 2017. The genetic basis of natural variation in Drosophila (Diptera: Drosophilidae) virgin egg retention. *Journal of Insect Science*, *17*(1).

Aminetzach, Y.T., Macpherson, J.M. and Petrov, D.A., 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science*, *309*(5735), pp.764-767.

Anadón, A., Martínez-Larrañaga, M.R. and Martínez, M.A., 2009. Use and abuse of pyrethrins and synthetic pyrethroids in veterinary medicine. *The Veterinary Journal*, *182*(1), pp.7-20.

Anholt, R.R. and Mackay, T.F., 2018. The road less traveled: From genotype to phenotype in flies and humans. *Mammalian genome*, *29*(1-2), pp.5-23.

The *Anopheles gambiae* 1000 Genomes Consortium, 2017. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*, *552*, pp 96-100.

Arya, G.H., Magwire, M.M., Huang, W., Serrano-Negron, Y.L., Mackay, T.F. and Anholt, R.R., 2015. The genetic basis for variation in olfactory behavior in *Drosophila melanogaster*. *Chemical senses*, *40*(4), pp.233-243.

Atwood, D. and Paisley-Jones, C., 2017. Pesticides industry sales and usage: 2008–2012 Market Estimates. *US Environmental Protection Agency, Washington, DC*, *20460*.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., Rollmann, S.M., Duncan, L.H., Lawrence, F., Anholt, R.R. and Mackay, T.F., 2009. Systems genetics of complex traits in *Drosophila melanogaster*. *Nature genetics*, *41*(3), p.299.

Ayroles, J.F., Buchanan, S.M., O'Leary, C., Skutt-Kakaria, K., Grenier, J.K., Clark, A.G., Hartl, D.L. and De Bivort, B.L., 2015. Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proceedings of the National Academy of Sciences*, *112*(21), pp.6706-6711.

Balabanidou, V., Kampouraki, A., MacLean, M., Blomquist, G.J., Tittiger, C., Juárez, M.P., Mijailovsky, S.J., Chalepakis, G., Anthousi, A., Lynd, A. and Antoine, S., 2016. Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences*, *113*(33), pp.9268-9273.

Barnes, K.G., Weedall, G.D., Ndula, M., Irving, H., Mzihalowa, T., Hemingway, J. and Wondji, C.S., 2017. Genomic footprints of selective sweeps from metabolic resistance to pyrethroids in African malaria vectors are driven by scale up of insecticide-based vector control. *PLoS genetics*, *13*(2), p.e1006539.

Bass, C., Puinean, A.M., Andrews, M., Cutler, P., Daniels, M., Elias, J., Paul, V.L., Crossthwaite, A.J., Denholm, I., Field, L.M. and Foster, S.P., 2011. Mutation of a nicotinic acetylcholine receptor β subunit is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *Bmc Neuroscience*, *12*(1), p.51.

Battlay, P., Schmidt, J.M., Fournier-Level, A. and Robin, C., 2016. Genomic and transcriptomic associations identify a new insecticide resistance phenotype for the selective sweep at the *Cyp6g1* locus of *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics*, *6*(8), pp.2573-2581.

Battlay, P., Leblanc, P.B., Green, L., Garud, N.R., Schmidt, J.M., Fournier-Level, A. and Robin, C., 2018. Structural variants and selective sweep foci contribute to insecticide resistance in the Drosophila genetic reference panel. *G3: Genes, Genomes, Genetics*, *8*(11), pp.3489-3497.

Bellen, H.J., Levis, R.W., He, Y., Carlson, J.W., Evans-Holm, M., Bae, E., Kim, J., Metaxakis, A., Savakis, C., Schulze, K.L. and Hoskins, R.A., 2011. The Drosophila gene disruption project: progress using transposons with distinctive site specificities. *Genetics*, *188*(3), pp.731-743.

Benna, C., Bonaccorsi, S., Wülbeck, C., Helfrich-Förster, C., Gatti, M., Kyriacou, C.P., Costa, R. and Sandrelli, F., 2010. Drosophila *timeless2* is required for chromosome stability and circadian photoreception. *Current Biology*, *20*(4), pp.346-352.

Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B. and Truong, A., 2010. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome research*, *20*(2), pp.281-290.

Bergman, C.M. and Haddrill, P.R., 2015. Strain-specific and pooled genome sequences for populations of *Drosophila melanogaster* from three continents. *F1000Research*, *4*.

Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E. and Hirschhorn, J.N., 2004. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, *74*(6), pp.1111-1120.

Boey, H.K.A., 2011. *The regulation and function of the Drosophila melanogaster Cytochrome P450 gene, Cyp12d1.* PhD thesis, The University of Melbourne.

Bonner, M.R., Coble, J., Blair, A., Beane Freeman, L.E., Hoppin, J.A., Sandler, D.P. and Alavanja, M.C., 2007. Malathion exposure and the incidence of cancer in the agricultural health study. *American journal of epidemiology*, *166*(9), pp.1023-1034.

Bou Sleiman, M.S., Osman, D., Massouras, A., Hoffmann, A.A., Lemaitre, B. and Deplancke, B., 2015. Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. *Nature communications*, *6*, p.7829.

Boyle, E.A., Li, Y.I. and Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell*, *169*(7), pp.1177-1186.

Brand, A.H. and Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *development*, *118*(2), pp.401-415.

Campo, D., Lehmann, K., Fjeldsted, C., Souaiaia, T., Kao, J. and Nuzhdin, S.V., 2013. Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection. *Molecular ecology*, *22*(20), pp.5084-5097.

Carbone, M.A., Yamamoto, A., Huang, W., Lyman, R.A., Meadors, T.B., Yamamoto, R., Anholt, R.R. and Mackay, T.F., 2016. Genetic architecture of natural variation in visual senescence in Drosophila. *Proceedings of the National Academy of Sciences*, *113*(43), pp.E6620-E6629.

Carino, F., Koener, J.F., Plapp Jr, F.W. and Feyereisen, R., 1992. 'Expression of the cytochrome P450 gene CYP6A1 in the housefly, *Musca domestica*', in Mullin, C.A. and Scott, J.G. (ed.) *Molecular Mechanisms of Insecticide Resistance,* American Chemical Society, Washington DC.

Catania, F., Kauer, M.O., Daborn, P.J., Yen, J.L., Ffrench-Constant, R.H. and Schlötterer, C., 2004. World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. *Molecular ecology*, *13*(8), pp.2491-2504.

Chakraborty, M., VanKuren, N.W., Zhao, R., Zhang, X., Kalsow, S. and Emerson, J.J., 2018. Hidden genetic variation shapes the structure of functional elements in Drosophila. *Nature genetics*, *50*(1), p.20.

Chakraborty, M., Emerson, J.J., Macdonald, S.J. and Long, A.D., 2019. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nature communications*, *10*(1), pp.1-11.

Charpentier, A. and Fournier, D., 2001. Levels of total acetylcholinesterase in *Drosophila melanogaster* in relation to insecticide resistance. *Pesticide Biochemistry and Physiology*, *70*(2), pp.100-107.

Chen, N., Pei, X.J., Li, S., Fan, Y.L. and Liu, T.X., 2020. Involvement of integument-rich CYP4G19 in hydrocarbon biosynthesis and cuticular penetration resistance in Blattella germanica (L.). *Pest management science*, *76*(1), pp.215-226.

Chintapalli, V.R., Wang, J. and Dow, J.A., 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nature genetics*, *39*(6), p.715.

Chung, H., Bogwitz, M.R., McCart, C., Andrianopoulos, A., Batterham, P. and Daborn, P.J., 2007. *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics*, *175*(3), pp.1071-1077.

Cohen, M.B., Keener, J.F. and Feyereisen, R., 1994. Structure and chromosomal localization of CYP6A1, a cytochrome P450-encoding gene from the house fly. *Gene*, *146*(2), pp.267-272.

Comins, H.N., 1986. Tactics for resistance management using multiple pesticides. *Agriculture, ecosystems & environment*, *16*(2), pp.129-148.
Coyne, F.P., 1951. Proper use of insecticides. *British medical journal*, *2*(4736), p.911.

Crow, J.F., 1954. Analysis of a DDT-resistant strain of Drosophila. *Journal of Economic Entomology*, *47*(3).

Daborn, P., Boundy, S., Yen, J. and Pittendrigh, B., 2001. DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Molecular Genetics and Genomics*, *266*(4), pp.556-563.

Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., ... & Feyereisen, R. (2002). A single P450 allele associated with insecticide resistance in Drosophila. *Science*, *297*(5590), 2253-2256.

Daborn, P.J., Lumb, C., Boey, A., Wong, W. and Batterham, P., 2007. Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression. *Insect biochemistry and molecular biology*, *37*(5), pp.512-519.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., 2011. The variant call format and VCFtools. *Bioinformatics*, *27*(15), pp.2156-2158.

Darwin, C. R., 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, John Murray, London.

Dembeck, L.M., Huang, W., Magwire, M.M., Lawrence, F., Lyman, R.F. and Mackay, T.F., 2015. Genetic architecture of abdominal pigmentation in *Drosophila melanogaster*. *PLoS genetics*, *11*(5), p.e1005163.

Denecke, S., Fusetto, R., Martelli, F., Giang, A., Battlay, P., Fournier-Level, A., Richard, A.O. and Batterham, P., 2017. Multiple P450s and variation in neuronal genes underpins the response to the insecticide imidacloprid in a population of *Drosophila melanogaster*. *Scientific reports*, *7*(1), p.11338.

Dermauw, W. and Van Leeuwen, T., 2014. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. *Insect biochemistry and molecular biology*, *45*, pp.89-110.

Devonshire, A.L., 1977. The properties of a carboxylesterase from the peach-potato aphid, *Myzus persicae* (Sulz.), and its role in conferring insecticide resistance. *Biochemical Journal*, *167*(3), pp.675-683.

Devonshire, A.L. and Field, L.M., 1991. Gene amplification and insecticide resistance. *Annual review of entomology*, *36*(1), pp.1-21.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S. and Couto, A., 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature*, *448*(7150), p.151.

Dobzhansky, T., 1937. *Genetics and the Origin of Species*. Columbia University Press.

Dong, K., Du, Y., Rinkevich, F., Nomura, Y., Xu, P., Wang, L., Silver, K. and Zhorov, B.S., 2014. Molecular biology of insect sodium channels and pyrethroid resistance. *Insect biochemistry and molecular biology*, *50*, pp.1-17.

Duneau, D., Sun, H., Revah, J., San Miguel, K., Kunerth, H.D., Caldas, I.V., Messer, P.W., Scott, J.G. and Buchon, N., 2018. Signatures of insecticide selection in the genome of *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics*, *8*(11), pp.3469-3480.

Durham, M.F., Magwire, M.M., Stone, E.A. and Leips, J., 2014. Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. *Nature communications*, *5*, p.4338.

Eggler, A.L., Liu, G., Pezzuto, J.M., van Breemen, R.B. and Mesecar, A.D., 2005. Modifying specific cysteines of the electrophile-sensing human Keap1 protein is insufficient to disrupt binding to the Nrf2 domain Neh2. *Proceedings of the National Academy of Sciences*, *102*(29), pp.10070-10075.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T. and Nielsen, R., 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular biology and evolution*, *31*(5), pp.1275-1291.

Feyereisen, R., 1999. Insect P450 enzymes. *Annual review of entomology*, *44*(1), pp.507-533.

ffrench–Constant, R.H., Pittendrigh, B., Vaughan, A. and Anthony, N., 1998. Why are there so few resistance–associated mutations in insecticide target genes?. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1376), pp.1685-1693.

Field, L.M., Devonshire, A.L. and Forde, B.G., 1988. Molecular evidence that insecticide resistance in peach-potato aphids (*Myzus persicae* Sulz.) results from amplification of an esterase gene. *Biochemical journal*, *251*(1), pp.309-312.

Finney, D.J., 1971. Probit analysis, Cambridge University Press. *Cambridge, UK*.

Fisher, R.A., 1918. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, pp 399-433.

Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*. The Clarendon Press.

Fiston-Lavier, A.S., Singh, N.D., Lipatov, M. and Petrov, D.A., 2010. *Drosophila melanogaster* recombination rate calculator. *Gene*, *463*(1-2), pp.18-20.

Fusetto, R., Denecke, S., Perry, T., Richard, A.J. and Batterham, P., 2017. Partitioning the roles of CYP6G1 and gut microbes in the metabolism of the insecticide imidacloprid in *Drosophila melanogaster*. *Scientific reports*, *7*(1), p.11339.

Garlapow, M.E., Huang, W., Yarboro, M.T., Peterson, K.R. and Mackay, T.F., 2015. Quantitative genetics of food intake in *Drosophila melanogaster*. *PloS one*, *10*(9), p.e0138129.

Garud, N.R., Messer, P.W., Buzbas, E.O. and Petrov, D.A., 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS genetics*, *11*(2), p.e1005004.

Garud, N.R. and Petrov, D.A., 2016. Elevated linkage disequilibrium and signatures of soft sweeps are common in *Drosophila melanogaster*. *Genetics*, *203*(2), pp.863-880.

Gellatly, K.J., Yoon, K.S., Doherty, J.J., Sun, W., Pittendrigh, B.R. and Clark, J.M., 2015. RNAi validation of resistance genes and their interactions in the highly DDT-resistant *91-R* strain of *Drosophila melanogaster*. *Pesticide biochemistry and physiology*, *121*, pp.107-115.

Georghiou, G.P. and Taylor, C.E., 1977. Genetic and biological influences in the evolution of insecticide resistance. *Journal of economic entomology*, *70*(3), pp.319-323.

Good, R.T., Gramzow, L., Battlay, P., Sztal, T., Batterham, P. and Robin, C., 2014. The molecular evolution of cytochrome P450 genes within and between Drosophila species. *Genome biology and evolution*, *6*(5), pp.1118-1134.

Gotoh, O., 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry*, *267*(1), pp.83-90.

Green, L., Battlay, P., Fournier-Level, A., Good, R.T. and Robin, C., 2019. *Cis*- and *trans*-acting variants contribute to survivorship in a naïve *Drosophila melanogaster* population exposed to ryanoid insecticides. *Proceedings of the National Academy of Sciences*, *116*(21), pp.10424-10429.

Gregory, R., Darby, A.C., Irving, H., Coulibaly, M.B., Hughes, M., Koekemoer, L.L., Coetzee, M., Ranson, H., Hemingway, J., Hall, N. and Wondji, C.S., 2011. A de novo expression profiling of Anopheles funestus, malaria vector in Africa, using 454 pyrosequencing. *PLoS One*, *6*(2).

Grenier, J.K., Arguello, J.R., Moreira, M.C., Gottipati, S., Mohammed, J., Hackett, S.R., Boughton, R., Greenberg, A.J. and Clark, A.G., 2015. Global diversity lines–a five-continent reference panel of sequenced *Drosophila melanogaster* strains. *G3: Genes, Genomes, Genetics*, *5*(4), pp.593-603.

Guio, L., Barrón, M.G. and González, J., 2014. The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila. *Molecular ecology*, *23*(8), pp.2020-2030.

Hamblin, M.T. and Di Rienzo, A., 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *The American Journal of Human Genetics*, *66*(5), pp.1669-1679.

Harbison, S.T., McCoy, L.J. and Mackay, T.F., 2013. Genome-wide association study of sleep in *Drosophila melanogaster*. *BMC genomics*, *14*(1), p.281.

Harris, R.B., Sackman, A. and Jensen, J.D., 2018. On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. *PLoS genetics*, *14*(12), p.e1007859.

Heckel, D.G., 2012. Insecticide resistance after silent spring. *Science*, *337*(6102), pp.1612-1614.

Hemingway, J., Shretta, R., Wells, T.N., Bell, D., Djimdé, A.A., Achee, N. and Qi, G., 2016. Tools and strategies for malaria control and elimination: what do we need to achieve a grand convergence in malaria?. *PLoS biology*, *14*(3), p.e1002380.

Hermisson, J. and Pennings, P.S., 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, *169*(4), pp.2335-2352.

Highfill, C.A., Tran, J.H., Nguyen, S.K., Moldenhauer, T.R., Wang, X. and Macdonald, S.J., 2017. Naturally segregating variation at Ugt86Dd contributes to nicotine resistance in *Drosophila melanogaster*. *Genetics*, *207*(1), pp.311-325.

Hoi, K.K., Daborn, P.J., Battlay, P., Robin, C., Batterham, P., O'Hair, R.A. and Donald, W.A., 2014. Dissecting the insect metabolic machinery using twin ion mass spectrometry: a single P450 enzyme metabolizing the insecticide imidacloprid in vivo. *Analytical chemistry*, *86*(7), pp.3525-3532.

Hooven, L.A., Sherman, K.A., Butcher, S. and Giebultowicz, J.M., 2009. Does the clock make the poison? Circadian variation in response to pesticides. *PloS one*, *4*(7), p.e6469.

Huang, W., Richards, S., Carbone, M.A., Zhu, D., Anholt, R.R., Ayroles, J.F., Duncan, L., Jordan, K.W., Lawrence, F., Magwire, M.M. and Warner, C.B., 2012. Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proceedings of the National Academy of Sciences*, *109*(39), pp.15553-15559.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ràmia, M., Tarone, A.M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.F. and Magwire, M.M., 2014. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome research*, *24*(7), pp.1193-1208.

Huang, W., Carbone, M.A., Magwire, M.M., Peiffer, J.A., Lyman, R.F., Stone, E.A., Anholt, R.R. and Mackay, T.F., 2015. Genetic basis of transcriptome diversity in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *112*(44), pp.E6010-E6019.

Hunter, C.M., Huang, W., Mackay, T.F. and Singh, N.D., 2016. The genetic architecture of natural variation in recombination rate in *Drosophila melanogaster*. *PLoS genetics*, *12*(4), p.e1005951.

Ivanov, D.K., Escott-Price, V., Ziehm, M., Magwire, M.M., Mackay, T.F., Partridge, L. and Thornton, J.M., 2015. Longevity GWAS using the Drosophila genetic reference panel. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *70*(12), pp.1470-1478.

Jordan, K.W., Craver, K.L., Magwire, M.M., Cubilla, C.E., Mackay, T.F. and Anholt, R.R., 2012. Genome-wide association for sensitivity to chronic oxidative stress in *Drosophila melanogaster*. *PloS one*, *7*(6), p.e38722.

Joußen, N., Heckel, D.G., Haas, M., Schuphan, I. and Schmidt, B., 2008. Metabolism of imidacloprid and DDT by P450 CYP6G1 expressed in cell cultures of Nicotiana tabacum suggests detoxification of these insecticides in *Cyp6g1*-overexpressing strains of *Drosophila melanogaster*, leading to resistance. *Pest Management Science: formerly Pesticide Science*, *64*(1), pp.65-73.

Kang, J., Kim, J. and Choi, K.W., 2011. Novel cytochrome P450, *cyp6a17*, is required for temperature preference behavior in Drosophila. *PLoS One*, *6*(12), p.e29800.

Kao, J.Y., Zubair, A., Salomon, M.P., Nuzhdin, S.V. and Campo, D., 2015. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the south-eastern United States and Caribbean Islands. *Molecular ecology*, *24*(7), pp.1499-1509.

Kaplan, N.L., Hudson, R.R. and Langley, C.H., 1989. The" hitchhiking effect" revisited. *Genetics*, *123*(4), pp.887-899.

Kapun, M., Barrón, M.G., Staubach, F., Vieira, J., Obbard, D.J., Goubert, C., Rota-Stabelli, O., Kankare, M., Haudry, A., Wiberg, R.A.W. and Waidele, L., 2018. Genomic analysis of European *Drosophila melanogaster* populations on a dense spatial scale reveals longitudinal population structure and continent-wide selection. *Biorxiv*, p.313759.

Karasov, T., Messer, P.W. and Petrov, D.A., 2010. Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS genetics*, *6*(6), p.e1000924.

Katzenberger, R.J., Chtarbanova, S., Rimkus, S.A., Fischer, J.A., Kaur, G., Seppala, J.M., Swanson, L.C., Zajac, J.E., Ganetzky, B. and Wassarman, D.A., 2015. Death following traumatic brain injury in Drosophila is associated with intestinal barrier dysfunction. *Elife*, *4*, p.e04790.

Khan, H.A.A., Akram, W. and Shad, S.A., 2014. Genetics, cross-resistance and mechanism of resistance to spinosad in a field strain of *Musca domestica* L.(Diptera: Muscidae). *Acta tropica*, *130*, pp.148-154.

Kikkawa, H., 1961. Genetical studies on the resistance to parathion in *Drosophila melanogaster*. *Annual Reports on Scientific Works from Osaka University*, *9*, pp.1-20.

King, E.G., Macdonald, S.J. and Long, A.D., 2012. Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics*, *191*(3), pp.935-949.

Kinghorn, K.J., Castillo-Quan, J.I., Bartolome, F., Angelova, P.R., Li, L., Pope, S., Cocheme, H.M., Khan, S., Asghari, S., Bhatia, K.P. and Hardy, J., 2015. Loss of PLA2G6 leads to elevated mitochondrial lipid peroxidation and mitochondrial dysfunction. *Brain*, *138*(7), pp.1801-1816.

Kliot, A. and Ghanim, M., 2012. Fitness costs associated with insecticide resistance. *Pest management science*, *68*(11), pp.1431-1437.

Lack, J.B., Cardeno, C.M., Crepeau, M.W., Taylor, W., Corbett-Detig, R.B., Stevens, K.A., Langley, C.H. and Pool, J.E., 2015. The Drosophila genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*, *199*(4), pp.1229-1241.

Langley, C.H., Stevens, K., Cardeno, C., Lee, Y.C.G., Schrider, D.R., Pool, J.E., Langley, S.A., Suarez, C., Corbett-Detig, R.B., Kolaczkowski, B. and Fang, S., 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*, *192*(2), pp.533-598.

Lê, S., Josse, J. and Husson, F., 2008. FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, *25*(1), pp.1-18.

Le Goff, G., Boundy, S., Daborn, P.J., Yen, J.L., Sofer, L., Lind, R., Sabourault, C. and Madi-Ravazzi, L., 2003. Microarray analysis of cytochrome P450 mediated insecticide resistance in Drosophila. *Insect biochemistry and molecular biology*, *33*(7), pp.701-708.

Leinonen, R., Sugawara, H., Shumway, M. and International Nucleotide Sequence Database Collaboration, 2010. The sequence read archive. *Nucleic acids research*, *39*(suppl_1), pp.D19-D21.

Levine, M.T., Eckert, M.L. and Begun, D.J., 2010. Whole-genome expression plasticity across tropical and temperate *Drosophila melanogaster* populations from Eastern Australia. *Molecular biology and evolution*, *28*(1), pp.249-256.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, *25*(14), pp.1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), pp.2078-2079.

Li, X., Bai, S. and Cass, B.N., 2012. Accord insertion in the 5′ flanking region of CYP6G1 confers nicotine resistance in *Drosophila melanogaster*. *Gene*, *502*(1), pp.1-8.

Liu, X., Li, Y.I. and Pritchard, J.K., 2019. Trans effects on gene expression can drive omnigenic inheritance. *Cell*, *177*(4), pp.1022-1034.

Lobell, A.S., Kaspari, R.R., Negron, Y.L.S. and Harbison, S.T., 2017. The genetic architecture of ovariole number in *Drosophila melanogaster*: genes with major, quantitative, and pleiotropic effects. *G3: Genes, Genomes, Genetics*, *7*(7), pp.2391-2403.

Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., de Candia, T.R., Lee, S.H., Wray, N.R., Kendler, K.S. and O'Donovan, M.C., 2015. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature genetics*, *47*(12), p.1385.

Long, A.D., Macdonald, S.J. and King, E.G., 2014. Dissecting complex traits using the Drosophila synthetic population resource. *Trends in Genetics*, *30*(11), pp.488-495.

Lucas, E.R., Miles, A., Harding, N.J., Clarkson, C.S., Lawniczak, M.K., Kwiatkowski, D.P., Weetman, D., Donnelly, M.J. and Anopheles gambiae 1000 Genomes Consortium, 2019. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome research*, *29*(8), pp.1250-1261.

Mackay, T.F., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M. and Richardson, M.F., 2012. The *Drosophila melanogaster* genetic reference panel. *Nature*, *482*(7384), p.173.

Magwire, M.M., Bayer, F., Webster, C.L., Cao, C. and Jiggins, F.M., 2011. Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a duplication. *PLoS genetics*, *7*(10), p.e1002337.

Magwire, M.M., Fabian, D.K., Schweyen, H., Cao, C., Longdon, B., Bayer, F. and Jiggins, F.M., 2012. Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLoS genetics*, *8*(11), p.e1003057.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. and Cho, J.H., 2009. Finding the missing heritability of complex diseases. *Nature*, *461*(7265), p.747.

Mateo, L., Rech, G.E. and González, J., 2018. Genome-wide patterns of local adaptation in Western European *Drosophila melanogaster* natural populations. *Scientific reports*, *8*(1), p.16143.

McCart, C. and Buckling, A., 2005. ffrench-Constant RH. *DDT resistance in flies carries no cost. Current Biology*, *15*, pp.R587-R589.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), pp.1297-1303.

McKenzie, J.A., Parker, A.G. and Yen, J.L., 1992. Polygenic and single gene responses to selection for resistance to diazinon in *Lucilia cuprina*. *Genetics*, *130*(3), pp.613-620.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R. and Wittkopp, P.J., 2010. Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome research*, *20*(6), pp.816-825.

Mendel, G., 1866. Versuche über Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865. *Abhandlungen*, pp.3-47.

Menozzi, P., Shi, M.A., Lougarre, A., Tang, Z.H. and Fournier, D., 2004. Mutations of acetylcholinesterase which confer insecticide resistance in *Drosophila melanogaster* populations. *BMC evolutionary biology*, *4*(1), p.4.

Messer, P.W. and Petrov, D.A., 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*, *28*(11), pp.659-669.

Misra, J.R., Horner, M.A., Lam, G. and Thummel, C.S., 2011. Transcriptional regulation of xenobiotic detoxification in Drosophila. *Genes & development*, *25*(17), pp.1796-1806.

Misra, J.R., Lam, G. and Thummel, C.S., 2013. Constitutive activation of the Nrf2/Keap1 pathway in insecticide-resistant strains of Drosophila. *Insect biochemistry and molecular biology*, *43*(12), pp.1116-1124.

Montgomery, S.L., Vorojeikina, D., Huang, W., Mackay, T.F., Anholt, R.R. and Rand, M.D., 2014. Genome-wide association analysis of tolerance to methylmercury toxicity in Drosophila implicates myogenic and neuromuscular developmental pathways. *PloS one*, *9*(10), p.e110375.

Morozova, T.V., Huang, W., Pray, V.A., Whitham, T., Anholt, R.R. and Mackay, T.F., 2015. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult drosophila. *BMC genomics*, *16*(1), p.865.

Mota-Sanchez, D., Bills, P.S. and Whalon, M.E., 2002. Arthropod resistance to pesticides: status and overview. *Pesticides in Agriculture and the Environment*, pp.241-272.

Mouches, C., Pasteur, N., Berge, J.B., Hyrien, O., Raymond, M., de Saint Vincent, B.R., De Silvestri, M. and Georghiou, G.P., 1986. Amplification of an esterase gene is responsible for insecticide resistance in a California Culex mosquito. *Science*, *233*(4765), pp.778-780.

Mutero, A., Pralavorio, M., Bride, J.M. and Fournier, D., 1994. Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. *Proceedings of the national academy of sciences*, *91*(13), pp.5922-5926.

Najarro, M.A., Hackett, J.L., Smith, B.R., Highfill, C.A., King, E.G., Long, A.D. and Macdonald, S.J., 2015. Identifying loci contributing to natural variation in xenobiotic resistance in Drosophila. *PLoS genetics*, *11*(11), p.e1005663.

Najarro, M.A., Hackett, J.L. and Macdonald, S.J., 2017. Loci contributing to boric acid toxicity in two reference populations of *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics*, *7*(6), pp.1631-1641.

Newcomb, R.D., Campbell, P.M., Ollis, D.L., Cheah, E., Russell, R.J. and Oakeshott, J.G., 1997. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proceedings of the National Academy of Sciences*, *94*(14), pp.7464-7468.

Ogita, Z., 1958. The genetical relation between resistance to insecticides in general and that to phenylthiourea (PTU) and phenylurea (PU) in *Drosophila melanogaster*. *Botyu-Kagaku*, *23*(pt. 4).

Orr, H.A., 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, *52*(4), pp.935-949.

Orr, H.A., 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, *6*(2), p.119.

Pavlides, J.M.W., Zhu, Z., Gratten, J., McRae, A.F., Wray, N.R. and Yang, J., 2016. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome medicine*, *8*(1), p.84.

Pavlidis, P. and Alachiotis, N., 2017. A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki*, *24*(1), p.7.

Pedra, J.H.F., McIntyre, L.M., Scharf, M.E. and Pittendrigh, B.R., 2004. Genome-wide transcription profile of field-and laboratory-selected dichlorodiphenyltrichloroethane (DDT)-resistant Drosophila. *Proceedings of the National Academy of Sciences*, *101*(18), pp.7034-7039.

Pennings, P.S. and Hermisson, J., 2006. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Molecular biology and evolution*, *23*(5), pp.1076-1084.

Perry, T., Chan, J.Q., Batterham, P., Watson, G.B., Geng, C. and Sparks, T.C., 2012. Effects of mutations in Drosophila nicotinic acetylcholine receptor subunits on sensitivity to insecticides targeting nicotinic acetylcholine receptors. *Pesticide biochemistry and physiology*, *102*(1), pp.56-60.

Pittendrigh, B., Reenan, R. and Ganetzky, B., 1997. Point mutations in the Drosophila sodium channel gene para associated with resistance to DDT and pyrethroid insecticides. *Molecular and General Genetics MGG*, *256*(6), pp.602-610.

Plapp Jr, F.W., 1984. The genetic basis of insecticide resistance in the house fly: evidence that a single locus plays a major role in metabolic resistance to insecticides. *Pesticide Biochemistry and Physiology*, *22*(2), pp.194-201.

Pool, J.E., Corbett-Detig, R.B., Sugino, R.P., Stevens, K.A., Cardeno, C.M., Crepeau, M.W., Duchen, P., Emerson, J.J., Saelao, P., Begun, D.J. and Langley, C.H., 2012. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS genetics*, *8*(12), p.e1003080.

Pritchard, J.K., Pickrell, J.K. and Coop, G., 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology*, *20*(4), pp.R208-R215.

Puinean, A.M., Foster, S.P., Oliphant, L., Denholm, I., Field, L.M., Millar, N.S., Williamson, M.S. and Bass, C., 2010. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLOS genetics*, *6*(6), p.e1000999.

Pyke, F.M., Bogwitz, M.R., Perry, T., Monk, A., Batterham, P. and McKenzie, J.A., 2004. The genetic basis of resistance to diazinon in natural populations of *Drosophila melanogaster*. *Genetica*, *121*(1), pp.13-24.

Qiu, X., Sun, W., McDonnell, C.M., Li-Byarlay, H., Steele, L.D., Wu, J., Xie, J., Muir, W.M. and Pittendrigh, B.R., 2013. Genome-wide analysis of genes associated with moderate and high DDT resistance in *Drosophila melanogaster*. *Pest management science*, *69*(8), pp.930-937.

R Core Team, 2018. R: A language and environment for statistical computing.

Robin, C., Battlay, P. and Fournier-Level, A., 2018. What can genetic association panels tell us about evolutionary processes in insects?. *Current opinion in insect science*.

Rostant, W.G., Bowyer, J., Coupland, J., Facey, J., Hosken, D.J. and Wedell, N., 2017. Pleiotropic effects of DDT resistance on male size and behaviour. *Behavior genetics*, *47*(4), pp.449-458.

Roush, R.T. and McKenzie, J.A., 1987. Ecological genetics of insecticide and acaricide resistance. *Annual review of entomology*, *32*(1), pp.361-380.

Roush, R.T., 1989. Designing resistance management programs: how can you choose?. *Pesticide Science*, *26*(4), pp.423-441.

Roush, R.T., 1993. Occurrence, genetics and management of insecticide resistance. *Parasitology Today*, *9*(5), pp.174-179.

Sawicki, R.M. and Farnham, A.W., 1968. GENETICS OF RESISTANCE TO INSECTICIDES OF THE SKA STRAIN OF *MUSCA DOMESTICA*: III. Location and isolation of the factors of resistance to dieldrin. *Entomologia Experimentalis et Applicata*, *11*(2), pp.133-142.

Schadt, E.E., Monks, S.A., Drake, T.A., Lusis, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G. and Linsley, P.S., 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature*, *422*(6929), p.297.

Schlenke, T.A. and Begun, D.J., 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proceedings of the National Academy of Sciences*, *101*(6), pp.1626-1631.

Schmidt, J.M., Good, R.T., Appleton, B., Sherrard, J., Raymant, G.C., Bogwitz, M.R., Martin, J., Daborn, P.J., Goddard, M.E., Batterham, P. and Robin, C., 2010. Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS genetics*, *6*(6), p.e1000998.

Schmidt, J., 2014. *The Genetics of DDT Resistance in Drosophila melanogaster: New Insights into an Old Debate.* PhD thesis, The University of Melbourne.

Schmidt, J.M., Battlay, P., Gledhill-Smith, R.S., Good, R.T., Lumb, C., Fournier-Level, A. and Robin, C., 2017. Insights into DDT resistance from the *Drosophila melanogaster* genetic reference panel. *Genetics*, *207*(3), pp.1181-1193.

Schrider, D.R., Hahn, M.W. and Begun, D.J., 2016. Parallel evolution of copy-number variation across continents in *Drosophila melanogaster*. *Molecular biology and evolution*, *33*(5), pp.1308-1316.

Shi, M.A., Lougarre, A., Alies, C., Frémaux, I., Tang, Z.H., Stojan, J. and Fournier, D., 2004. Acetylcholinesterase alterations reveal the fitness cost of mutations conferring insecticide resistance. *BMC evolutionary biology*, *4*(1), p.5.

Shipp, E. and Otton, J., 1976. Orcadian rhythms of sensitivity to insecticides in *Musca domestica* (Diptera, Muscidae). *Entomologia Experimentalis et Applicata*, *19*(2), pp.163-171.

Shorter, J., Couch, C., Huang, W., Carbone, M.A., Peiffer, J., Anholt, R.R. and Mackay, T.F., 2015. Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. *Proceedings of the National Academy of Sciences*, *112*(27), pp.E3555-E3563.

Siegfried, B.D. and Scharf, M.E., 2001. Mechanisms of organophosphate resistance in insects. In *Biochemical sites of insecticide action and resistance* (pp.269-291). Springer, Berlin, Heidelberg.

Signor, S.A., New, F.N. and Nuzhdin, S., 2017. A large panel of *Drosophila simulans* reveals an abundance of common variants. *Genome biology and evolution*, *10*(1), pp.189-206.

Smith, J.M. and Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), pp.23-35.

Sparks, T.C., Lockwood, J.A., Byford, R.L., Graves, J.B. and Leonard, B.R., 1989. The role of behavior in insecticide resistance. *Pesticide Science*, *26*(4), pp.383-399.

Sparks, T.C. and Nauen, R., 2015. IRAC: Mode of action classification and insecticide resistance management. *Pesticide biochemistry and physiology*, *121*, pp.122-128.

Steele, L.D., Muir, W.M., Seong, K.M., Valero, M.C., Rangesa, M., Sun, W., Clark, J.M., Coates, B. and Pittendrigh, B.R., 2014. Genome-wide sequencing and an open reading frame analysis of dichlorodiphenyltrichloroethane (DDT) susceptible (*91-C*) and resistant (*91-R*) *Drosophila melanogaster* laboratory populations. *PloS one*, *9*(6), p.e98584.

Steinhauer, N., Kulhanek, K., Antúnez, K., Human, H., Chantawannakul, P. and Chauzat, M.P., 2018. Drivers of colony losses. *Current opinion in insect science*, *26*, pp.142-148.

Strycharz, J.P., Lao, A., Li, H., Qiu, X., Lee, S.H., Sun, W., Yoon, K.S., Doherty, J.J., Pittendrigh, B.R. and Clark, J.M., 2013. Resistance in the highly DDT-resistant *91-R* strain of *Drosophila melanogaster* involves decreased penetration, increased metabolism, and direct excretion. *Pesticide biochemistry and physiology*, *107*(2), pp.207-217.

Swarup, S., Huang, W., Mackay, T.F. and Anholt, R.R., 2013. Analysis of natural variation reveals neurogenetic networks for Drosophila olfactory behavior. *Proceedings of the National Academy of Sciences*, *110*(3), pp.1017-1022.

Sykiotis, G.P. and Bohmann, D., 2008. Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. *Developmental cell*, *14*(1), pp.76-85.

Szpiech, Z.A. and Hernandez, R.D., 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular biology and evolution*, *31*(10), pp.2824-2827.

Teets, N.M. and Hahn, D.A., 2018. Genetic variation in the shape of cold-survival curves in a single fly population suggests potential for selection from climate variability. *Journal of evolutionary biology*, *31*(4), pp.543-555.

Terhzaz, S., Cabrero, P., Brinzer, R.A., Halberg, K.A., Dow, J.A. and Davies, S.A., 2015. A novel role of Drosophila cytochrome P450-4e3 in permethrin insecticide tolerance. *Insect biochemistry and molecular biology*, *67*, pp.38-46.

Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, *14*(2), pp.178-192.

Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J. and Piro, A., 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*, *293*(5529), pp.455-462.

Toé, K.H., N'Falé, S., Dabiré, R.K., Ranson, H. and Jones, C.M., 2015. The recent escalation in strength of pyrethroid resistance in Anopheles coluzzi in West Africa is linked to increased expression of multiple gene families. *BMC genomics*, *16*(1), p.146.

Tomizawa, M. and Casida, J.E., 2005. Neonicotinoid insecticide toxicology: mechanisms of selective action. *Annu. Rev. Pharmacol. Toxicol.*, *45*, pp.247-268.

Turner, T.L., Levine, M.T., Eckert, M.L. and Begun, D.J., 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics*, *179*(1), pp.455-473.

Turner, T.L., Miller, P.M. and Cochrane, V.A., 2013. Combining genome-wide methods to investigate the genetic complexity of courtship song variation in *Drosophila melanogaster*. *Molecular biology and evolution*, *30*(9), pp.2113-2120.

Unckless, R.L., Rottschaefer, S.M. and Lazzaro, B.P., 2015. A genome-wide association study for nutritional indices in Drosophila. *G3: Genes, Genomes, Genetics*, *5*(3), pp.417-425.

Vaisnav, M., Xing, C., Ku, H.C., Hwang, D., Stojadinovic, S., Pertsemlidis, A. and Abrams, J.M., 2014. Genome-wide association analysis of radiation resistance in *Drosophila melanogaster*. *PloS one*, *9*(8), p.e104858.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. and Yang, J., 2017. 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, *101*(1), pp.5-22.

Visscher, P.M. and Goddard, M.E., 2019. From RA Fisher's 1918 Paper to GWAS a Century Later. *Genetics*, *211*(4), pp.1125-1130.

Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K., 2006. A map of recent positive selection in the human genome. *PLoS biology*, *4*(3), p.e72.

Vonesch, S.C., Lamparter, D., Mackay, T.F., Bergmann, S. and Hafen, E., 2016. Genome-wide analysis reveals novel regulators of growth in *Drosophila melanogaster*. *PLoS genetics*, *12*(1), p.e1005616.

Vontas, J., David, J.P., Nikou, D., Hemingway, J., Christophides, G.K., Louis, C. and Ranson, H., 2007. Transcriptional analysis of insecticide resistance in Anopheles stephensi using cross-species microarray hybridization. *Insect molecular biology*, *16*(3), pp.315-324.

Walsh, B. and Lynch, M., 2018. *Evolution and selection of quantitative traits*. Oxford University Press.

Wang, J.B., Lu, H.L. and Leger, R.J.S., 2017. The genetic basis for variation in resistance to infection in the *Drosophila melanogaster* genetic reference panel. *PLoS pathogens*, *13*(3), p.e1006260.

Weber, A.L., Khan, G.F., Magwire, M.M., Tabor, C.L., Mackay, T.F. and Anholt, R.R., 2012. Genome-wide association analysis of oxidative stress resistance in *Drosophila melanogaster*. *PloS one*, *7*(4), p.e34745.

Wilding, C.S., 2018. Regulating resistance: CncC: Maf, antioxidant response elements and the overexpression of detoxification genes in insecticide resistance. *Current opinion in insect science*, *27*, pp.89-96.

Willoughby, L., Chung, H., Lumb, C., Robin, C., Batterham, P. and Daborn, P.J., 2006. A comparison of *Drosophila melanogaster* detoxification gene induction responses for six insecticides, caffeine and phenobarbital. *Insect biochemistry and molecular biology*, *36*(12), pp.934-942.

Wilson, T.G., 2005. Drosophila: sentinels of environmental toxicants. *Integrative and comparative biology*, *45*(1), pp.127-136.

Wondji, C.S., Irving, H., Morgan, J., Lobo, N.F., Collins, F.H., Hunt, R.H., Coetzee, M., Hemingway, J. and Ranson, H., 2009. Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. *Genome research*, *19*(3), pp.452-459.

Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution, *Proceedings of the Sixth International Congress on Genetics* 1, pp.356–366.

Yadav, A.S., Singh, P.K. and Shrivastava, V.K., 2010. Pyrethroid resistance in Chagas disease vectors: The case of Triatoma infestans cuticle. *A Biannual Newsletter of the Center for Integrated Plant Systems (CIPS) in Cooperation with the Insecticide Resistance Action Committee (IRAC) and the Western Regional Coordinating Committee (WRCC-60)*, *19*(2), p.59.

Younis, H.M., Abo-El-Saad, M.M., Abdel-Razik, R.K. and Abo-Seda, S.A., 2002. Resolving the DDT target protein in insects as a subunit of the ATP synthase. *Biotechnology and applied biochemistry*, *35*(1), pp.9-17.

Zawaira, A., Ching, L.Y., Coulson, L., Blackburn, J. and Chun Wei, Y., 2011. An expanded, unified substrate recognition site map for mammalian cytochrome P450s: analysis of molecular interactions between 15 mammalian CYP450 isoforms and 868 substrates. *Current drug metabolism*, *12*(7), pp.684-700.

Zichner, T., Garfield, D.A., Rausch, T., Stütz, A.M., Cannavó, E., Braun, M., Furlong, E.E. and Korbel, J.O., 2013. Impact of genomic structural variation in *Drosophila melanogaster* based on population-scale sequencing. *Genome research*, *23*(3), pp.568-579.

Zhang, W., Reeves, R.G. and Tautz, D., 2019. Testing the omnigenic model for a behavioral trait in *Drosophila melanogaster*. *BioRxiv*, p.639955.

Appendices

Appendices are available from the following URLs:

Appendix 1

https://figshare.com/articles/Appendix_1/11302415

Appendix 2

https://figshare.com/articles/Appendix_2/11302418

Appendix 3

https://figshare.com/articles/Appendix_3/11302421