Transposable Elements and the Evolution of Virus Resistance in *Drosophila melanogaster*

Osama Brosh

Emmanuel College
Department of Genetics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

September 2021

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the School of Biology Degree Committee.

Osama Brosh

Summary:
Transposable Elements and the Evolution of Virus Resistance in *Drosophila melanogaster*

Viral infection results in a significant fitness cost on organisms, which driving evolutionary change in antiviral defences. Previous research in *Drosophila* has shown that there is large genetic variation in pathogen resistance, and that a significant amount of genetic variation to virus resistance can be explained by major-effect polymorphisms in a small number of genes.

In this thesis I will explore the role of transposable elements in the evolution of virus resistance. We discover a *Doc* retrotransposon insertion in the *Drosophila* gene *Veneno (Ven; CG9684),* a gain of function mutation which confers resistance to Drosophila A Virus (DAV). We show that the insertion acts by creating a truncated transcript *Ven_tra*, which is sufficient to induce DAV resistance in *Drosophila* cell culture. Although this transcript includes sequences of *Doc* element origin, we show that it doesn't require these sequences for its resistant phenotype. We also show that resistance doesn't rely on the functioning of the key protein domains *Ven_tra* encodes: an MYND Zinc finger and a Tudor domain. We work on further characterizing this allele by narrowing down the pathways it interacts and find it does not require a functional siRNA pathway to cause resistance, and fail to show its reliance on other immune pathways, leaving us with a limited understanding of its mechanism. This adds it to a growing list of poorly understood restriction factors, suggesting that virus resistance in *Drosophila* relies largely on these idiosyncratic restriction factors targeting specific viruses.

Furthermore, we examine another instance of the *Doc* element insertion into the sequence of the gene *CHKov1* which confers resistance to the *Drosophila melanogaster* sigma virus (DMelSV). We show that the *Doc* insertion in *CHKov1* leads to the expression of a transcript that includes sequences from both *CHKov1* and the *Doc* element's ORF1, encoding a gene-TE chimera which is sufficient to cause resistance.

Finally, we explore ways in which virus resistance can be artificially created in arthropods through genetic engineering. We attempt to use CRISPR-associated gene 13 (Cas13) to generate DAV resistance in *Drosophila melanogaster* in the hope that the system could be extended to confer resistance to other viruses (such as Zika) in other arthropods (such as *Aedes aegypti)*.

Acknowledgements

I would like to thank everyone who has made this possible. My family, and friends: you have made my life worth living. To Jon Day, Daniel Fabian, Gaspar Bruner, Alex Leitão, Carlos Luque, and everyone in the Jiggins Group: thank you for your invaluable support. To Frank Jiggins: thank you for your time, patience, invaluable insight, and excellent mentorship.

A final thank you to my mother Lina Zubdeh Brosh, my father Fadi Brosh, and my sister Jana Brosh. I dedicate this work to you.

Acknowledgements of Collaborations

The following results were done in part or fully by collaborators:

Fully done by collaborators (Rodrigo Cogni, Daniel Fabian, Ignacio Tolosana)

1. Results in Figure 1.2 (Dominance Assay, QTL Mapping)

2. Results in Figure 1.3 (Fine-scale recombinant mapping)

3. Results in Figure 1.6B (RNAi knockdown of candidate genes)

Partially done by collaborators:

1. Results in Figure 1.1 (DGRP Correlation of Resistance with *Ven* genotype):

    - Resistance assay done by Rodrigo Cogni
    - Genotyping DGRP lines done by me

2. Results in Figure 1.7 (Knocking Down Doc element sites using RNAi)

    - Preparation of Plasmids done by Jonathan Day
    - Rest of experiment done by me

3. Results in Figures 1.10-1.11 (Knocking Out *Veneno* with CRISPR/Cas9

    - Design of CRISPR/Cas9 targets done by Daniel Fabian
    - Preparation of Plasmids done by Jonathan Day
    - Rest of experiment done by me

Other:

crRNA sequences for Cas13 experiments in Chapter 4 were designed by Ian Patterson. Several primers were designed by collaborators (including those in appendices 1-2 which were designed by Daniel Fabian). Various plasmids, fly stocks, and other materials are acknowledged in the text.

Table of Contents

Declaration i

Summary: iii

Acknowledgements v

Acknowledgements of Collaborations vii

Introduction ***1***

Viruses **2**

Classification of Viruses with the Baltimore Scheme 3

Infection 3

Viruses in Drosophila 4

DAV 5

DMelSV 6

Antiviral Immunity in *Drosophila* **7**

1) RNAi pathway: 7

2) Phagocytosis 9

3) Apoptosis 10

4) Autophagy 10

5) Signalling Pathways 11

i) JAK/STAT pathway: 11

ii) IMD pathway 11

iii) Toll Pathway 12

6) Other Restriction Factors 13

Resistance to DAV and DMelSV 13

Genetic Variation in Virus Resistance **13**

Origins of Variation 14

Genetic Architecture of Variation 14

Evolutionary Models of Variation 15

Viral Resistance Through Genetic Engineering **16**

RNAi and other small RNA pathways 17

Upregulating Immune Pathways 18

CRISPR-Cas 18

System 19

Mechanism and Classification 19

Engineered antiviral uses 20

Transposable elements: **22**

Classification: by replication pattern 22

1) DNA Transposon (class II): 22

2) Retrotransposons (class I): 23

3) Non-autonomous transposons: 24

Transposition Is a Parasitic Trait 25

Impact of TE on evolution 26

Passive rearrangements: 26

Active rearrangements: 27

Effect on genome size 28

Mutagenesis: 28

TE Role in Altering Gene Expression 28

Transposable Elements Domestication: 30

TE role in telomeres: 30
RAG1 and RAG2 in vertebrate immunity 30
Chimeric TE-gene proteins 31
Bioinformatic searches 31
Examples of chimeric genes 34
Insertion Position Preference 37
Transposition Rate 37
Mechanisms of TE Resistance 40
piRNA pathway: 40
Endogenous siRNAs: 42
Methylation 42
Other: 42

***Doc* Elements 43**
In *white*: 44
Upstream of *Cyp6g1* 45
In *CHKov1* 45

***Methods* 47**
DGRP 47
Fly Care 47
DNA Extraction from flies 47
Quantifying DNA and Assessing Purity 47
TouchDown PCR Taq 48
Agarose Gels 48
ExoSap PCR Clean-up 49
Sanger Sequencing 49
Infecting flies with DAV 49
Infecting flies with DmelSV 50
TRIzol RNA Extraction 50
Reverse Transcription 51
Quantifying RNA and Assessing Purity 51
Measuring Viral Titre 51
Creating plasmids with inserts (NEBuilder) 53
Preparing Inserts 53
Preparing Vectors, Assembly and Transformation 54
Selection 55
Plasmid preparation (mini, midi) 55
Cell Culture Care 56
Transfecting cells 56
Infecting cells with DAV 57
Flow Cytometry 57
ϕC31 58

***Chapter 1 - Doc Element Insertion in Veneno Causes Resistance to DAV* 59**

Introduction 59

Results 60
A *Doc* element insertion in *Veneno* is associated with resistance to DAV. 60
Transcription of *Veneno* is necessary for resistance phenotype 67
No evidence that the transcription of the *Doc* Element is involved in DAV resistance 69
Doc element insertion leads to Truncated transcript of *Veneno* 73
Veneno Sequence upstream of Insertion Necessary for Resistance 75
Ven_tra causes DAV resistance in cells 80

Discussion 82

Methods **84**
Characterising genetic variation in resistance to DAV 84
Mapping to chromosome 85
QTL mapping 86
Fine-scale mapping 86
RNAi knockdown of candidate genes 89
Locating *Doc* Element Insertion 89
Sequencing *Doc* Element 90
Genotyping *Veneno* 91
Veneno RNAi 91
Targeting *Doc* element with RNAi 92
Mutagenesis with CRISPR/Cas9 94
Sequencing CRISPR target sites with Illumina 96
Nanopore (direct RNA) 97
Nanopore (PCR cDNA) 97
Bioinformatics 98
Testing DAV infection in cells 98
Determining TCID50 (Median Tissue Culture Infectious Dose) 99
Expressing Ven^{trunc} in cells 100
Infecting Cells Expressing Ven^{trunc} 102

Chapter 2 – Mechanism of DAV Resistance Caused by Ven^{Doc} Allele 103

Introduction **103**

Results **105**
Ven^{Doc} has lower expression level than Ven^{+} in female abdomens 105
Veneno Sequences of *Doc* element origin are not necessary for antiviral function 107
Neither the MYND Zn Finger Domain nor the Tudor Domain are Necessary for Antiviral Function 108
RNAi Pathway is Not Required for *Veneno's* Antiviral Function 111
Immune Pathway Knockdown Test (ImPaKT) 114
Proteomics 117

Discussion **118**

Methods **123**
Measuring Expression of *Veneno* 123
Q5 site directed mutagenesis 123
Mutating Plasmids 123
Infecting Cell Lines expressing mutated Ven^{trunc} 126
siRNA pathway mutants 126
Immune Pathway Knockdown Test (ImPaKT) 127
Proteomics 128

Chapter 3- Doc element insertion in CHKov1 Causes Resistance to DMelSV by Encoding Novel Chimeric Transcript 132

Introduction **132**

Results **133**
$CHKov1^{Doc1420}$ Produces 4 New truncated Transcripts 133
Novel *CHKov1* Transcript Causes Increased Resistance to DMelSV 134
Knockdown of *CHKov1* with RNAi 136
Infecting Cells with DMelSV 137

Discussion **139**

Methods **142**

Nanopore (PCR cDNA) ... 142
Bioinformatics ... 142
CHKov1 in DL2 cells ... 143
CHKov1 RNAi ... 144
CHKov1 truncated transcript in flies ... 144
Fly Lines ... 146
Experiment ... 147

Chapter 4 – Genetic Engineering of Virus Resistance Using CRISPR/Cas13 ... 148

Introduction ... 148

Results ... 149
Cas13 targeting DAV ... 149
Cas13 targeting Zika ... 153

Discussion ... 153

Methods ... 155
Cas13 ... 155
crRNA ... 156
Fly Stocks ... 159

Conclusion ... 161

References ... 164

Appendices ... 182
Appendix 1 - Primers used in locating *Doc* element ... 182
Appendix 2 - Primers for sequencing *Doc* insertion and flanking sequences ... 183
Appendix 3 - Oligos Used to make plasmid expressing RNAi targeting *Doc* Element ... 186
Appendix 4 - Oligos used for CRISPR/Cas9 gRNAs ... 187
Appendix 5 - Illumina primers for Sequencing CRISPR target sites: ... 188
Appendix 6 – Primers for pMT-*Ven^{trunc}*–puro plasmid ... 189
Appendix 7 – Primers for pMT-*Ven^{trunc}*-V5–puro plasmid ... 189
Appendix 8 – Primers for Q5 Site Directed Mutatagenesis ... 189
Appendix 9 - Primers for Proteomics Experiment Plasmids ... 190
Appendix 10 Primers for pMT-CHKov1_tr0-V5-puro ... 191
Appendix 11 - Primers for pUb-*CHKov1_tr0* ... 191
Appendix 12 – Primers for pUAST-CHKov1 ... 192
Appendix 13 - Oligo used for RNAi on *CHKov1* ... 192
Appendix 14 Ultramers for Cas13 crRNA ... 193
Appendix 15 – Primers to Amplify Cas13 crRNA Ultramers ... 194
Appendix 16 – Primers for Cas13dAdm plasmid ... 194
Appendix 17 - Primers targeting junction for sequencing plasmids and screening colonies: ... 195
pcfd5 ... 195
pUB ... 195
Appendix 18 – Primers surrounding insert for sequencing plasmids and screening colonies: ... 195
Appendix 19 - qPCR Primers and Probes ... 196
Appendix 20 – Synonymous mutation (Premature Control) and Cloning Does Not Abolish Virus Resistance ... 197
Appendix 21 - Flow cytometry data showing ***Ven^{trunc}*** expression in Cells from Fig 2.2 ... 198
Appendix 22- qPCR data showing ***Ven^{trunc}*** expression in Cells from Fig 2.2 ... 199
Appendix 23- Flow cytometry data showing ***Ven^{trunc}*** expression in cells from Fig 2.3 ... 200
Appendix 24- qPCR data showing ***Ven^{trunc}*** expression in cells from Fig 2.3 ... 201

Introduction

Viruses are a very diverse range of parasites which infect many organisms (see Virus section). The adverse effects of viral infection on fitness have led host species such as *Drosophila melanogaster* to develop a wide range of antiviral defences (see Antiviral Immunity section). As with many traits, viral resistance is subject to a lot of variation between individuals, some of which can be explained through a genetic basis. The genetic basis to the variation of virus resistance is important to study since it provides insight into the evolution of virus resistance, an adaptation of great importance for increasing fitness. Since viruses evolve their own methods of circumventing viral resistance pathways (for example see viral adaptation to the RNAi pathway in Antiviral Immunity section), understanding the evolutionary response on the side of host organisms also sheds light into the evolutionary arms race between host and virus.

Variation in virus resistance in *Drosophila melanogaster* often occurs at the level of a few major effect genes (see: Genetic Architecture of Variation). Polymorphisms within these genes result in some alleles having increased viral resistance, which then either go on to spread in the population as a result of positive selection, or become maintained at some frequency as a consequence of balancing selection (see Evolutionary Models of Variation).

Transposable elements are an important source of polymorphisms, since they result in a range of mutations in the host genome including the insertion of the transposable elements themselves (see: Impact of TE on evolution). Often these mutations are detrimental to the host's fitness (see: Transposable elements decrease host fitness), but on occasion, they can be beneficial (see transposable element domestication). Insertions into the coding sequence of genes can result in new phenotypes either by creating a truncated version of the gene, or by incorporating the sequence of the transposable element into the sequence of host gene creating a 'chimera' gene which uses the sequences of its two constituents to result in a novel function (see Chimeric TE-gene proteins).

In this thesis we will explore the role played by transposable elements in the evolution of virus resistance. We will begin by looking at a *Doc* insertion (see *Doc* insertion section) in the coding sequence of the gene *Veneno* causing resistance to *Drosophila* A Virus (DAV) (Chapter 1) and the mechanism of this resistant allele (Chapter 2). Then we will move on to another *Doc* insertion in the coding sequence of *CHKov1* causing resistance to *Drosophila melanogaster* Sigma Virus (DMelSV) (Chapter 3).

The insight into how antiviral defences evolve also allows us to think about creating novel antiviral defences using genetic engineering (see Viral Resistance Through Genetic Engineering). We will explore using Cas13 to generate a new defence pathway (Chapter 4).

Viruses

We have learned a lot about viruses since the famous 1892 Iwanowski experiment showing that the cause of mosaic disease in tobacco plants can pass through a bacteria proof-filter, and that it must therefore be caused by an even smaller pathogen, which we now know as the virus.

A virus is a selfish parasitic pathogen, known to infect organisms from every known group (Dimmock et al., 2007), and capable of reproducing solely from its genetic materials (Dimmock et al., 2007) by hijacking the molecular machinery of its host, particularly ribosomes (Dimmock et al., 2007).

Viruses are able to spread through populations by moving between organisms. In general there are two ways this can occur: Viruses can be transmitted horizontally between different organisms through contact. This can happen in through the respiratory route, conjunctival route, faecal route, sexual route, via urine, or through the mechanical route (e.g. mosquito bites). Viruses can also be transmitted vertically from parent (usually mother) to offspring (Dimmock et al., 2007).

Viruses store their genetic material in one type of nucleic acid (RNA or DNA)(Dimmock et al., 2007), and use them to synthesize different viral components, which are later synthesized into virus

particles (Dimmock et al., 2007). Although their classification as living or non-living things is a matter of controversy, viruses are subject to evolutionary pressures just like any organism, which results in a large diversity of viruses (Dimmock et al., 2007) that have evolved a plethora of (often ingenious) mechanisms to infect their hosts.

Classification of Viruses with the Baltimore Scheme

Viruses can be classified based on the type of nucleic acid they use as genetic material, as they can either have single-stranded or double stranded RNA, or single stranded or double stranded DNA (Dimmock et al., 2007). Since viruses lack ribosomes, they cannot translate their messenger RNA, and rely on host ribosomes to do so. They must therefore produce mRNA (or genomic viral RNA) which is recognized by the host cell ribosomes (Dimmock et al., 2007). Therefore, we can use the Baltimore scheme to classify viruses according to their genetic material, and their mechanism of mRNA synthesis.

Infection

In general, viruses multiply by first attaching to a host cell; in animals this occurs through binding to receptor molecules on the cell surface (Dimmock et al., 2007). Then the virus penetrates the cell; in animals this occurs either through the fusion of the virus' lipid bilayer with the cell membrane or, in the case of unenveloped viruses, by entry into a vesicle through receptor-mediated endocytosis (Dimmock et al., 2007). Next, the virus uncoats its genetic material, which then uses the cellular machinery to replicate in a process known as biosynthesis. The replicated genetic material is then assembled into new viral particles and released out of the host cell to seek out new cells (Dimmock et al., 2007). This release either occurs through the lysis of the cell or through cytoplasmic vesicles which release the virus (Dimmock et al., 2007). The viral particles also require maturation at some point either before or after release.

Class	Description
1	**double stranded DNA virus.** mRNA is transcribed from either DNA strand
2	**single stranded DNA virus**. DNA either +ive or -ive sense and must be converted to dsDNA before transcription
3	**double stranded RNA virus**. This class of virus carries genes that code for enzymes involved in transcription from dsRNA.
4	**single stranded positive sense RNA**. In this class of virus, proteins are translated directly from the RNA carried by the virus. So the virus's genetic material acts also as the mRNA. During replication, a dsRNA intermediate is produced using enzymes coded by the virus genome, with the negative sense strand acting as the template
5	**single stranded negative sense RNA**. In this class of virus, mRNA is produced using RNA-synthesizing proteins (RNA Replicase/ RNA-dependent RNA polymerase RdRp) coded by the viral genome. During replication, a double stranded RNA intermediate is produced, with the positive sense strand used as a template strand.
6	Viruses of this class carry **single stranded RNA** as their genetic material, and generate a double stranded DNA intermediate during replication.
7	Viruses of this class carry **double stranded DNA** as their genetic material, and generate a single stranded RNA intermediate during replication.

Table i.1. Baltimore Scheme Classification of Viruses

Viruses in Drosophila

Some viruses can naturally infect *Drosophila* such as Drosophila C virus (DCV), Drosophila A Virus (DAV), Nora virus, and Sigma virus; these are thought to have narrow ranges (Webster et al., 2015a; J. Xu & Cherry, 2014). Other viruses have a broad range making it possible to use them to infect *Drosophila* experimentally, such as Cricket Paralysis virus (CrPV), Flock House virus (FHV), and viruses in the arbovirus families such as *Alphaviridae, Flaviviridae,* and *Bunyaviridae* (J. Xu & Cherry, 2014). Here we go into further detail on two *Drosophila melanogaster* viruses: DAV and DMelSV.

In assaying natural *Drosophila melanogaster* populations world-wide for virus prevalence, Webster et al. found that some viruses such as DCV and Newfield Virus, which are highly prevalent in lab strains and in cell culture, appear to be less prevalent in natural populations (C. L. Webster et al., 2015). This would mean that a lot of studies on virus resistance which use common lab viruses such as DCV will not provide a complete picture of natural infection. On the other hand, they found that other viruses, such as La Jolla Virus, Nora Virus, DAV, Sigma Virus (DMelSV), Kallithea Virus, and Galbut Virus to mention a few, appear to be more frequent in natural populations (C. L. Webster et al., 2015).

La Jolla Virus (LJV), which belongs to the *Iflaviridae* family of positive strand RNA viruses, has been recently discovered by Webster et al. (2015). It has also been identified in *Drosophila Suzukii* where it preferentially infects the central nervous system (Carrau et al., 2018). Nora virus, which is a picorna-like positive strand RNA virus, has been shown to primarily infect the intestinal tract, and to be horizontally transmitted via the fecal-oral route (Habayeb et al., 2009). It seems to exhibit no pathological effects (Habayeb et al., 2009). Kallithea Virus (KV), the first naturally occurring DNA virus to be discovered in *Drosophila*, belongs to the *Nudiviridae* family (Palmer, Medd, et al., 2018). It has been shown to reduce survival in infected flies, with an especially large reduction in male fly survival after 10 days of infection (Palmer, Medd, et al., 2018). Furthermore, females infected with KV exhibit less movement, and lay significantly fewer eggs in late infection(Palmer, Medd, et al., 2018). Galbut virus belongs to the family *Partitiviridae* of double stranded RNA viruses has been shown to be vertically transmitted from either parent with 100% efficiency, with no evidence of horizontal transmission (Cross et al., 2020). In the next section, we go into further detail on DAV and DMelSV.

DAV

Drosophila A virus (DAV) is a non-enveloped positive sense RNA virus T=3 icosahedral core with a genome size of around 4.8 kb (Ambrose et al., 2009). It is present in around 5% of wild *D melanogaster* flies worldwide, and as much as 56% of flies in some populations (such as Athens, Georgia) (Webster et al., 2015a). Its genome encodes two ORFs with sequence similarities to an

RNA-dependent RNA polymerase (RdRP) and a capsid protein respectively (Ambrose et al., 2009). Ambrose et al. found that although DAV has a permuted RdRP which indicates its evolutionary relation to *Tetraviridae* and *Birnaviridae*, its capsid protein is most highly related to *Nodaviridae,* which makes it an unusual virus that is difficult to classify (Ambrose et al., 2009). It is now thought to be most closely related to *Permutotetraviridae* (Webster et al., 2015a), which is a family of viruses newly created in 2011.

Wu et al. used small RNA sequencing to assemble viral genomes from siRNAs and piRNAs (Q. Wu et al., 2010). They detected a viral RdRP with similar sequences to *Tetraviridae* RdRPs, which they called Drosophila tetravirus (DTrV) (Q. Wu et al., 2010). They later found that this sequence is identical to that of DAV (Q. Wu et al., 2010).

In *Drosophila suzukii*, DAV replicates mostly in the head followed by the thorax, and much less so in the abdomen (Carrau et al., 2018). However, DAV is much more virulent in *D. suzukii* than in *D. melanogaster*, causing 50% of infected flies to die by the third day of infection (Carrau et al., 2018), whereas it is much less pathogenic in *D melanogaster,* where infected flies have similar survival rates to uninfected flies 3 days post infection (Fabian, Cogni, unpublished data). This could mean that the replication pattern might be different in *D melanogaster*, and further investigation into this is required.

DMelSV

Drosophila melanogaster Sigma Virus (DMelSV) is a negative sense RNA virus in the family *Rhabdoviridae* (Longdon et al., 2012; J. Xu & Cherry, 2014). It is around 12.5 kb (Longdon et al., 2012). It was first discovered in 1937 by L'Heritier and Teissier due to the fact that CO_2 anesthetized flies infected with DMelSV become irreversibly paralyzed and die (Longdon et al., 2012).

DMelSV has 6 genes in the order 3'-N-P-M-G-X-L-5' where N is a nucleoprotein gene, P is the polymerase-associated protein gene, M is the matrix protein gene, G is the glycoprotein gene, and L is the polymerase protein gene (Teninges et al., 1993). The X gene which is of unknown function, has a similar structure to viral RNA polymerases (Longdon et al., 2012).

DMelSV is transmitted vertically from either parent to offspring (Brun & Plus, 1980). Prevalence of DMelSV in wild population has been estimated to be between 0 and 15% (Carpenter et al., 2007), around 10% (Wilfert & Jiggins, 2010) and around 30% (Wayne et al., 2011). The mode of transmission might appear to suggest that the cost of infection could be low; however, Yampolsky et al. showed that under intense competitive conditions, flies infected with DmelSV are outcompeted by uninfected flies, and exhibit at least a 20-30% loss in fitness (Yampolsky et al., 1999).

Antiviral Immunity in *Drosophila*

Drosophila melanogaster has diverse range of antiviral defence mechanisms. In this section, we will highlight some of these defences, and discuss the role they play in virus resistance; several of these processes are conserved in other species (Palmer, Varghese, et al., 2018).

1) RNAi pathway:

The RNAi (or siRNA) pathway is thought to be the main mechanism of antiviral defense in *Drosophila melanogaster*. The pathway relies on the RNaseIII enzyme Dicer-2 (Dcr-2) which cuts double stranded RNA into double stranded small interfering RNA (siRNA) (21-25 nt) (Huszar & Imler, 2008). With the help of double stranded RNA binding protein R2D2, the template strand of the siRNA is loaded into an RNA-induced silencing complex (RISC) (often referred to as the effector complex) containing (among others) the protein Argonaute-2 (AGO2) (Huszar & Imler, 2008). Subsequently, the end of the siRNA is methylated by protein Hen-1 (Bronkhorst & van Rij, 2014). The RISC carrying the siRNA then binds to RNA sequences targeted by the template

siRNA and cleaves them using the RNAseH domain of AGO2, thereby eliminating target double stranded RNA (Huszar & Imler, 2008).

Since RNA viruses generate dsRNA when they replicate, the siRNA pathway is therefore an effective way to degrade viral RNA and not host RNA. As a result, flies deficient for the genes *dicer-2*, *r2d2*, or *Ago2*, (which code for the proteins Dicer-2, R2D2, and AGO2 respectively), have been shown to be more susceptible to infection with certain viruses such as Drosophila C Virus (DCV), and Cricket Paralysis Virus (CrPV), Drosophila X Virus (DXV), Flock house virus (FHV), and Sindbis Virus (SINV). (Huszar & Imler, 2008)

To give some examples, Galiana-Arnoux et al. showed that a mutation in *dicer-2* causes increased susceptibility to Flock house virus (FHV), Drosophila C virus (DCV), and Sindbis virus (SINV) (Galiana-Arnoux et al., 2006). Similarly, Van Rij et al. showed that mutating Ago-2 led to increased susceptibility to DCV, and cricket paralysis virus (CrPV), and that mutating *Dcr-2* led to increased susceptibility to CrPV (van Rij et al., 2006). Wang et al. showed that *dcr-2* and *r2d2* mutants are more susceptible to CrPV and FHV (X.-H. Wang et al., 2006). Additionally, other genes involved in the RNAi pathway have been implicated in virus resistance such as *dFOXO* (Spellberg & Marr, 2015).

Furthermore, siRNAs targeting viruses can be detected in infected flies. For instance, Wang et al. detected anti-CrPV siRNAs in S2 cells within 24 hours of infection with CrPV (X.-H. Wang et al., 2006), and Galiana-Arnoux et al. detected anti-FHV siRNAs in flies 5 days post infection with FHV (Galiana-Arnoux et al., 2006).

Since the siRNA pathway is the main antiviral pathway in *Drosophila* it makes sense from a fitness perspective for viruses to adapt mechanisms for RNAi suppression. Indeed, evidence has been found that several viruses encode for a viral suppressor of RNAi (VSR), which inhibits RNAi function and therefore allows the proliferation of the virus. The next paragraph will highlight such examples.

Li et al. found that the B2 protein which is encoded by the FHV genome, has an RNAi suppression function (Li et al., 2002). FHV with mutations in B2 cannot replicate in S2 cells, but are able to replicate in S2 cells with AGO2 mutations, indicating B2 has a role in overcoming siRNA defences (Li et al., 2002). Chao et al. found that B2 operates by preventing the double stranded RNA from being cleaved by Dicer-2, and from being loaded into the RISC complex (Chao et al., 2005). Van Rij et al. found that transgenic flies expressing B2 are more susceptible to DCV infection (van Rij et al., 2006). They also found that DCV codes for its own RNAi pathway silencing, which is ineffectual against flies with artificially injected 21nt siRNAs, and therefore likely acts by blocking the processing of dsRNA into 21nt siRNAs (van Rij et al., 2006). Wang et al. also found evidence that CrPV codes for its own viral suppressor of RNAi (X.-H. Wang et al., 2006).

2) Phagocytosis

Phagocytosis is a cellular defense mechanism in which *Drosophila* hemoctyes first attach to a target particle then modify their cytoskeleton to engulf the particle, internalize it, and subsequently destroy it (Lemaitre & Hoffmann, 2007). Elrod-Erickson et al. showed that phagocytosis can be inhibited through the injection of polystyrene beads into the hemocoel (Elrod-Erickson et al., 2000). When paired with the inactivation of the Imd pathway this inhibition results in increased susceptibility to *E. coli*, revealing the immune role phagocytosis plays in concert with humoral pathways (Elrod-Erickson et al., 2000).

Costa et al. used Elrod-Erickson et al.'s method to find that flies with inhibited phagocytosis are more susceptible to CrPV showing that phagocytosis has a role in CrPV resistance (Costa et al., 2009). Furthermore, Lamiable et al. found that inhibiting phagocytosis by injecting latex beads led to an increased susceptibility to CrPV, FHV and VSV, but not to DCV SINV or IIV6 (Lamiable et al., 2016). They found that same increase in susceptibility to CrPV, FHV, and VSV when genetically depleting hemocytes in transgenic flies (Lamiable et al., 2016). These findings show that phagocytosis has an antiviral role in *Drosophila*.

3) Apoptosis

Apoptosis, or programmed cell death, occurs when a particular signal, activates cellular death receptors and proteins such as p53 leading to downstream effects including growth arrest, and the commencement of apoptosis (Roulston et al., 1999). During that process, DNA becomes fragmented, blebs form on the membrane, and the cell shrinks and eventually fragments (Roulston et al., 1999). Although apoptosis in virus infected cells could, in certain cases, give a fitness advantage to the virus (apoptotic bodies produced by the cell fragmentation can contain virus particles providing the virus easy entry into other cells) (Roulston et al., 1999), the untimely death of an infected cell can also limit the amount of replication a virus is able to do within that cell (Mondotte & Saleh, 2018; Roulston et al., 1999).

Liu et al. found that infection of *Drosophila* larvae with FHV results in apoptosis through the expression of pro-apoptotic genes, and that a null mutation in gene *p53* prevents this from occurring (B. Liu et al., 2013). The *p53* mutation, they found, results in an increased FHV titre, suggesting a role played by apoptosis in resisting the viral infection (B. Liu et al., 2013).

4) Autophagy

Autophagy, or "self-eating" is a process by which the cell breaks down its own components in which autophagosomes surround structures to be degraded, before delivering them to the lysosome for degradation (Juhasz & Neufeld, 2006; Mondotte & Saleh, 2018). At least 27 genes are involved in autophagy (Juhasz & Neufeld, 2006).

Lamiable et al. found that mutations in autophagy gene *Atg7* lead to increased susceptibility to VSV, and increased resistance to FHV, suggesting that the autophagy pathway can have an antiviral or proviral role depending on the virus (Lamiable et al., 2016). Similarly, Shelly et al. found that targeting autophagy genes (including *Atg5, Atg7, Atg18,* as well as others) with RNAi

led to increased susceptibility to VSV (Shelly et al., 2009). Furthermore, Moy et al. found increased susceptibility to Rift Valley Fever Virus (RVFV) upon silencing or mutating autophagy genes such as *Atg18* (Moy et al., 2014).

5) Signalling Pathways

i) JAK/STAT pathway:

The JAK/STAT pathway is a cellular signalling pathway which responds to extracellular ligands such as cytokines (Levy & Darnell, 2002). Ligands bind to cell-surface receptors, which activates Janus Kinases (JAKs) which in turn leads to the activation of signal transducers and activators of transcription (STAT) proteins (Levy & Darnell, 2002). STATs go on to alter transcription in the nucleus (Levy & Darnell, 2002).

In *Drosophila* there is one JAK kinase called Hopscotch (*hop*), and knocking it down results in increased susceptibility to DCV and CrPV (Kemp et al., 2013). Furthermore, infection with DCV and CrPV results in an increase in expression of genes involved with the JAK/STAT pathway (Kemp et al., 2013). Avadhanula et al. also found that mutating *stat92E* which codes for a STAT protein results in an increased replication of a SINV replicon produced endogenously in the tested fly line (Avadhanula et al., 2009). Taken together, these data show that the JAK/STAT pathway plays a role in virus resistance.

ii) IMD pathway

The IMD pathway is an immune signalling pathway which responds, primarily, to Gram-negative bacteria (Hoffmann, 2003). The IMD protein responds to the activation of the transmembrane protein PGRP-LC, causing downstream effects which lead to the activation of transcription factor

Relish, and eventually the transcription of genes some of which encoding antimicrobial peptides (AMPs) such as Diptericin B (DptB) (Hoffmann, 2003).

Avadhanula et al. showed that mutating *relish* results in an increased replication of a SINV replicon produced endogenously in the tested fly line (Avadhanula et al., 2009). Furthermore, they found that SIN replication led to 4.8 fold increase in Diptericin and a 9.2 fold increase in Metchnikowin, both of which are AMPs regulated by the IMD pathway (Avadhanula et al., 2009). The IMD pathway likely upregulates these AMPs to combat viral infection; indeed Huang et al. showed that knockdown of DptB results in increased susceptibility to SINV (Huang et al., 2013). These data suggest that the IMD pathway is involved in SINV resistance, and that it increases SINV resistance by upregulating AMPs such as DptB.

Costa et al. showed that flies with mutations in IMD pathway genes such as *PGRP-LC* and *relish* were more susceptible to CrPV infection (Costa et al., 2009). However, CrPV infection did not lead to an increase in AMPs (Costa et al., 2009), indicating that the IMD pathway also plays a role in CrPV resistance which does not involve AMP upregulation.

iii) Toll Pathway

The Toll pathway is an immune signalling pathway in which the transmembrane Toll receptor is activated by an extracellular cytokine Spätzle, primarily in response to infection by fungi or Gram-positive bacteria (Hoffmann, 2003). This leads to the activation of downstream proteins including Pelle, which eventually activate transcription factors Dorsal and Dif, leading to the transcription of genes such as ones coding for antimicrobial or antifungal peptides (Hoffmann, 2003).

Ferreira et al. found that mutants for *spätzle, dorsal, pelle,* and *toll*, but not *dif* are more susceptible to DCV through oral infection (Ferreira et al., 2014). They also found that mutations in *pelle* result in increased susceptibility to CrPV, Nora virus, and FHV (Ferreira et al., 2014). These data show that the Toll pathway plays an antiviral role in *Drosophila*.

6) Other Restriction Factors

Other molecules, known as restriction factors, have been implicated in antiviral roles, but their function remains poorly understood. Examples of these include *Vago* (Deddouche et al., 2008), *ref(2)P* (Bangham, Kim, et al., 2008), *CHKov1* (Magwire et al., 2011), *pastrel* (Magwire et al., 2012), and *Ge-1* (Cao et al., 2016). In Chapters 1 and 2 we will discuss a recently discovered restriction factor with an antiviral role: *Veneno*.

Resistance to DAV and DMelSV

Very little is known about the roles of the classical antiviral pathways highlighted in this section in DAV and DMelSV immunity. Webster et al. showed using RNA sequencing that flies produce siRNA sequences targeting both DAV and DMelSV, without showing that the siRNA pathway effectively reduces infection of either of these viruses (C. L. Webster et al., 2015). This shows that the siRNA pathway targets DAV and DMelSV infections, but it is unclear whether it is an effective resistance mechanism. Additionally, autophagy, could potentially be linked to DMelSV resistance since the restriction factor ref(2)p, which was found by Cogni et al. (2016) to explain 24% of variance in DMelSV resistance, is an orthologue of p62 and plays a key role in regulating autophagy in *Drosophila* (Bitto et al., 2014).

Genetic Variation in Virus Resistance

There is a large amount of genetic variation in pathogen resistance in fruit flies (Lazzaro et al., 2004; Tinsley et al., 2006). In the next few paragraphs we will discuss the variation in virus resistance: the conditions which lead to its origin, the genetic architecture it manifests in, and the evolutionary models which explain it.

Origins of Variation

Two mechanisms appear as likely culprits driving the variation in virus resistance: adaptation to local conditions, and selection as a consequence of viral infection. These are not mutually exclusive, and it is likely they both play roles in driving variation. Variation in pathogen and virus resistance has been detected between populations from different climate regions, which supports the idea that adaptation to local conditions drives variation (Thomas-Orillard et al., 1995; Tinsley et al., 2006). Evidence for selection as a result of viral infection driving variation is laid out in the next paragraph.

Magwire et al. used genome wide association studies to find variation in resistance to infection by DCV, FHV, DMelSV, and DAffSV (*Drosophila affinis* Sigma Virus) in *Drosophila melanogaster* (Magwire et al., 2012). They found that of the four viruses they checked, variation in resistance was greater for viruses which naturally infect *Drosophila melanogaster* (DMelSV and DCV) than for the ones which don't (DaffSV and FHV) (Magwire et al., 2012). Duxbury et al. cross infected populations of different *Drosophila* species with sigma virus species that naturally infect them and sigma virus species that do not naturally infect them (Duxbury et al., 2019). They found that when the virus naturally infects the species, that led to a highly significant increase in natural variation in susceptibility to that virus within populations of that species, further supporting the finding of Magwire et al. (Duxbury et al., 2019). These findings support the notion that variation in virus resistance is driven in a large part by selection as a consequence of virus infection.

Genetic Architecture of Variation

There is plenty of evidence that virus resistance in *Drosophila* exhibits a simple genetic architecture: Major effect polymorphisms play significant role in genetic variation to virus resistance in Drosophila (Bangham, Knott, et al., 2008; Cogni et al., 2016; Magwire et al., 2012). Magwire et al. found that 29% of heritability in resistance to DMelSV in the DGRP is due to a polymorphism in gene *CHKov1* (Magwire et al., 2012). The resistant allele contains *Doc* element

insertion, which is the likely cause of the resistant phenotype (Magwire et al., 2011). Another 8% of the heritability in resistance to DMelSV is due to polymorphisms in *ref(2)P* (Magwire et al., 2012). In DCV resistance, a polymorphism in one gene (*pastrel*) explains 47% of heritability (Magwire et al., 2012). Cogni et al. used the DSPR panel to also find a large proportion of variance in virus resistance caused by major effect genes, with *pastrel* explaining 78% of variance in resistance to DCV, and ref(2)p explaining 24% of variance in resistance to DMelSV (the susceptible allele of *CHKov1* is absent in the DSPR panel) (Cogni et al., 2016). Additionally, they found only one epistatic interaction between polymorphisms controlling resistance for DMelSV, and none between polymorphisms controlling resistance to DCV indicating that for the most part, epistatic interactions do not play a large role in the genetic architecture of DCV and DMelSV resistance (Cogni et al., 2016).

Duxbury et al. found that when comparing variation in resistance to viruses which naturally infect the population to variation in resistance to viruses which don't, there is a greater number of major effect polymorphisms affecting resistance to naturally infecting viruses (Duxbury et al., 2019). This is consistent with the idea that selection due to infection by viruses increases the frequency of polymorphisms with major effect on virus resistance in the population.

Evolutionary Models of Variation

Two evolutionary models can explain the variation in virus resistance and its simple genetic basis: major effect polymorphisms can be under positive selection (wherein they sweep rapidly through a population), or alternatively be under balancing selection (where they are maintained at moderate frequencies for long periods of time within the population) such as negative frequency-dependent selection (Bangham et al., 2007). In the case of the aforementioned polymorphisms in *ref(2)p*, Bangham et al. found evidence that a mutation conferring resistance to DMelSV spread through the population due to positive selection since the resistant haplotype contained less variation than

expected by chance (Bangham et al., 2007). In the case of the polymorphism in *CHKov1* which causes resistance to DMelSV, linkage disequilibrium between the *Doc* insertion causing the resistant allele and the surrounding sites indicate that positive selection has led to an increase in frequency of the resistant allele (Aminetzach et al., 2005; Magwire et al., 2011).

However, although the previous evidence points to a selective sweep scenario, it does not fully preclude the possibility that the polymorphisms were maintained in the population by frequency-dependent selection, perhaps due to a negative fitness cost associated with the resistant allele which offsets the positive fitness costs of DMelSV resistance (Bangham et al., 2007). Even assuming that in the case of both *ref(2)p* and *CHKov1*, positive selection has led to a sweep of the resistant allele through the population, that would not imply that this is the case for all major effect polymorphisms causing virus resistance. Indeed, in the case of the polymorphism in *pastrel* which causes DCV resistance, Cao et al. found no evidence of selection driving up the frequency of the resistant allele, which is old and maintained at low frequency in populations worldwide (Cao et al., 2017). This low frequency could possibly be maintained by balancing selection, or by the neutral evolution of this polymorphism, and Cao et al. were unable to find evidence supporting one hypothesis over the other (Cao et al., 2017).

The previous section highlights the importance of finding such major effect polymorphisms in virus resistance: they not only help us understand the still poorly understood viral resistance pathways of fruit flies, but can also shed insight into the evolution of resistance to infectious disease.

Viral Resistance Through Genetic Engineering

Given what we know about virus biology and antiviral immunity, it is possible to artificially generate viral resistance in organisms through genetic engineering. In this section we will discuss a few examples of this.

RNAi and other small RNA pathways

The endogenous RNAi pathway is the main antiviral pathway in Drosophila melanogaster (see virus resistance pathways section). It is therefore a sensible pathway to utilize in order to manipulate viral resistance in Drosophila, and arthropods more generally.

By engineering the expression of RNAi templates targeted at a particular virus, it is possible to generate resistance to that virus. For example, Olson et al. were able to reduce the replication of dengue type 2 (DENV-2) virus in the salivary glands of *Aedes aegypti* mosquitos by transducing female mosquitos with antisense RNA targeting the genome of the virus (Olson et al., 1996). They were therefore able to limit the transmission of the virus to humans (Olson et al., 1996).

Franz et al. were then able to develop this further into a heritable resistance, by creating transgenic *A. aegypti* mosquitos which express an inverted repeat RNA derived from the genome of DENV-2 (Franz et al., 2006). The inverted repeat forms a double stranded RNA which is then used by the endogenous RNAi machinery to target the virus, thereby resulting in increased resistance (Franz et al., 2006). Since the inverted repeats are expressed from the DNA of the mosquitos, they are capable of passing down this resistance to their offspring.

RNAi has also been used to produce virus resistance in mammals. For instance, McCaffrey et al. induced an RNAi response in mice cells and in immunocompetent and immunodeficient mice by expressing short hairpin RNAs targeting Hepatitis B Virus (HBV) mRNA (McCaffrey et al., 2003). This resulted in reduced levels of HBV replication (McCaffrey et al., 2003).

Boden et al. were also able to reduce viral replication using RNAi (Boden et al., 2004). They incorporated RNAi sequences targeting the HIV-1 genome into a human miR-30 pre-microRNA backbone, which allowed them to express the siRNA in cells, and reduce viral replication more effectively than with using hairpin RNA (Boden et al., 2004). The introduced sequences were 100% complementary to the target, which allowed processing via the RNAi pathway (Boden et al.,

2004). They were therefore able to make use of two different short RNA pathways (RNAi and miRNA) to engineer resistance.

Small RNA pathways have also been used in plants to engineer virus resistance. By expressing a hairpin RNA targeting the genome of the potato virus Y (PVY), Smith et al. were able to use the siRNA pathway to create resistance to the virus in tobacco plants (N. A. Smith et al., 2000). Additionally, Qu et al. expressed artificial miRNA in tobacco plants targeting *Cucumber mosaic virus,* and that resulted in increased resistance to the virus (Qu et al., 2007).

One problem with using the RNAi pathway to target viruses is that RNAi is a natural defence used by some organisms against viruses, which means that viruses have had time to evolve mechanisms to overcome RNAi defences. We discuss evolved RNAi suppression in viruses in the Antiviral Pathways section.

Upregulating Immune Pathways

Another avenue of using genetic engineering to generate virus resistance is through upregulating endogenous immune pathways. For example, Jupatanakul et al. upregulated the JAK/STAT pathway in the fat body tissue of *Aedes aegypti* by using bloodmeal-inducible promoters to overexpress either *Dome* or *Hop* (2 genes in the pathway)(Jupatanakul et al., 2017). Mosquitos with the upregulated pathway had increased resistance to DENV but not Zika virus (ZIKV) or chikungunya virus (CHIKV) (Jupatanakul et al., 2017).

CRISPR-Cas

An alternative approach to engineering virus resistance is to use pathways from other species which target nucleic acid sequences. One such pathway is CRISPR-Cas which we describe in the next few paragraphs.

System

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) are a genomic feature present in prokaryotes, which, together with Cas (CRISPR associated) genes comprise a system of adaptive anti-viral immunity in prokaryotes (Barrangou et al., 2007; Y. Zhu et al., 2018). By showing that bacterial strains with CRISPR spacer sequences that are identical to bacteriophage sequences were resistant to those bacteriophages, and that experimentally altering the spacer sequences leads to a loss of specificity, Barrangou et al. determined that CRISPR-Cas can provide specific antiviral resistance based on genomic sequence (Barrangou et al., 2007).

In 2012, Jinek et al. found that Cas9 in the presence of the appropriate guide RNA sequences targeted at a plasmid sequence cleaves the plasmid DNA specifically at the target site (Jinek et al., 2012). They showed that CRISPR/Cas9 functions by acting as an RNA guided nuclease, introducing double stranded breaks in the DNA matching with the guide RNA (Jinek et al., 2012). This opened up the potential for the CRISPR/Cas9 to be exploited in genome editing by introducing engineered guide RNAs targeting specific DNA sequences, and indeed the system was later successfully applied to edit the genomes of organisms including fruit flies, human cells, mice, and plants (Doudna & Charpentier, 2014).

Mechanism and Classification

In general, CRISPR-Cas systems interfere with the target sequence by first forming an effector complex, which is a ribonucleoprotein (RNP) complex composed of the Cas protein(s) and the CRISPR RNA (crRNA). Then this complex recognizes a sequence targeted by the crRNA (or a highly similar sequence), and uses the Cas nuclease to cleave it and inactivate it (Koonin & Makarova, 2019; Y. Zhu et al., 2018).

In addition to CRISPR-Cas9, there are several other CRISPR-Cas systems. These systems have been classified into several types (I through VI) based on the Cas protein(s) they involve (Koonin & Makarova, 2019). The CRISPR-Cas types are grouped into two classes based on the type of their ribonucleoprotein (RNP) complex (Zhu et al., 2018):

- Class 1 systems have multisubunit RNP complexes which include multiple Cas proteins. They include types I, III, and IV. (Koonin & Makarova, 2019; Y. Zhu et al., 2018)
- Class 2 systems have single protein RNP complexes, which have one large multidomain protein. They include types II (CRISPR/Cas9), V, and VI (CRISPR/Cas13). (Koonin & Makarova, 2019; Y. Zhu et al., 2018)

Although a lot of research has focused on the role of CRISPR-Cas systems in targeting DNA, not all CRISPR-Cas systems have that function (Y. Zhu et al., 2018). In fact, both CRISPR type III and type VI target RNA (Y. Zhu et al., 2018).

Engineered antiviral uses

DNA

One way for a CRISPR-Cas system to be used in engineering virus resistance, is by using it to target viral DNA. For this strategy to work, the target virus must either be a DNA virus, or have a DNA intermediate in its life cycle (Kennedy & Cullen, 2017). This has been done successfully in cells as well in living organisms using CRISPR-Cas9 (Kennedy & Cullen, 2017; Price et al., 2016).

For example, Liu et al. designed crRNAs targeting the conserved regions of HBV (which is a DNA virus) (X. Liu et al., 2015). They found that using gRNA/Cas9 systems to target HBV in cells resulted in up to a 100 fold inhibition in replication with no cytotoxicity or off-target effects (X. Liu et al., 2015). Additionally, by injecting these gRNA/Cas9 systems into mice, they were able to cause substantial decreases in viral titre *in vivo* (X. Liu et al., 2015).

In another example, Zhu et al. selected 10 different targets for CRISPR/Cas9 in the genome of HIV-1, which is a retrovirus with a DNA provirus stage (W. Zhu et al., 2015). They engineered the crRNA-Cas9 systems for these targets, and introduced the systems into Jurkat cells (line of human T lymphocyte cells), and were thereby able to reduce viral production by up to 20-fold (W. Zhu et al., 2015). They therefore showed that such systems could be used to target HIV-1 in human cells, particularly at its provirus phase which is the primary obstacle to curing latent HIV infections (Price et al., 2016).

RNA

To target RNA viruses, Cas proteins should be able to target RNA. CRISPR-Cas9 has been shown capable of targeting RNA by O'Connell et al: Cas9 does not natively bind to RNA because RNA lacks the double stranded protospacer adjacent motif (PAM) sequence which must present near the target for Cas9 binding (O'Connell et al., 2014). O'Connell et al. found that by artificially introducing theses PAM sequences in the form of PAM oligonucleotides (PAMmers), they were able to activate Cas9-RNA binding, and the ability of Cas9 to specifically target RNA sequences for cleavage (O'Connell et al., 2014).

Furthermore, some strains of bacteria have a version of Cas9 which targets RNA natively, such as *Franscisella novicada* whose Cas9 (FnCas9) which targets bacterial mRNA and thereby represses genes (Price et al., 2015). Price et al. were able to use FnCas9 to engineer resistance to hepatitis C virus, a positive sense RNA virus, by expressing both FnCas9 and a guide RNA targeting the virus in a strain of human cells (Price et al., 2015). They found that cells expressing both these components had a lower level of viral luciferase than controls, indicating the inhibition of viral replication (Price et al., 2015).

Finally, Cas9 is not the only Cas protein which can be used to target RNA. Cas13 (type VI) is capable of targeting RNA. In Chapter 4, we will explore using Cas13 to engineer virus resistance.

Transposable elements:

Transposable elements (TEs) are selfish genomic elements that replicate throughout the genome. They are the most widespread selfish genetic element, and have the capacity not only to make copies of themselves within an organism's genome, but to also move between different species of organisms (BURT & Trivers, 2009). They comprise large percentages of some species' genome e.g. maize (50%) (García Guerreiro, 2012), humans (45%) (Beck et al., 2011; García Guerreiro, 2012) , *Drosophila* (12%) (García Guerreiro, 2012).

Classification: by replication pattern

Transposable elements can replicate in three distinct ways.

1) DNA Transposon (class II):

1a) **Subclass 1** (e.g. *P-element*): These are typically 1-10kb and typically have terminal inverted repeats (TIRs) (BURT & Trivers, 2009); they replicate via cut-and-paste mechanisms. These transposons code for Transposases, which bind to both ends of the transposon at the TIRs, excise the transposon from its position in the genome, and subsequently paste it at a different site in the genome.

Despite the cut-and-paste mechanism removing the original copy of the transposon, subclass 1 transposons can still increase in number. One way they can do so is by utilizing host DNA repair mechanisms. Nassif et al. found that the original copy of the *P-element* in *Drosophila melanogaster* following transposition can pick up insertions found in homologous sequences elsewhere in the genome (Nassif et al., 1994), suggesting that after the excision of the transposable element the double-stranded break generated is repaired using a homologous sequence (which may or may not be the sister chromatid) as a template. This allows the *P-element* to be maintained at its original site

despite being excised. Another mechanism through which subclass 1 transposons can increase in number is by transposing during DNA replication from a locus that has already been replicated to one that hasn't; this means that the transposon gets replicated twice, which creates a sister chromatid with two copies of the transposon. A transposon that has already been replicated can also transpose to a locus on the sister chromatid, which results in the original chromatid having no copies of the transposon, and the sister chromatid having two (BURT & Trivers, 2009; FEDOROFF, 1983).

1b) **Subclass 2** (eg. *Helitrons*): These elements transpose via a copy-and-paste mechanism wherein only one strand is displaced during transposition (Wicker et al., 2007).

2) Retrotransposons (class I):

These replicate by coding for an RNA intermediate, which is then reverse transcribed and inserted into the genome at a different site (preserving the original TE). There are two types of retrotransposon:

2a) **LTR retrotransposon**: flanked by long terminal repeats (LTR). LTRs contain a promoter recognized by RNA Polymerase II of the host. LTR retrotransposons also code for Gag protein (which creates a structure containing the RNA intermediate and the TE proteins), reverse transcriptase which reverse transcribes the sequence into complimentary DNA, and integrase for inserting into the genome. (BURT & Trivers, 2009; Levin & Moran, 2011)

After being transcribed, an LTR retrotransposon is either translated ('somatic fate'), or reverse transcribed ('germline fate') and inserted into the genome (but never both) (BURT & Trivers, 2009). Once the structure created by the Gag protein is assembled, reverse transcription of the 'germline' RNA into double stranded cDNA occurs, after which the structure falls apart and the cDNA is inserted into the host genome by the integrase (BURT & Trivers, 2009). LTR retrotransposon proteins can act in trans, which allows transposition of elements lacking functional proteins; for example *'mini-Ty1'* elements consisting of the 3' LTR, the transcribed

portions of the 5' LTR, and a 285bp internal region is capable of being transposed without coding for its own functional proteins (H. Xu & Boeke, 1990).

Since they are identical, the LTRs in LTR retrotransposons can recombine, effectively excising the LTR transposon out of the genome.

2b) **Non-LTR retrotransposons** (aka LINEs): comprise ~17% of human DNA (Beck et al., 2011), and include the *Doc* Element in *Drosophila melanogaster*. They are not flanked by long terminal repeats, and resemble an integrated mRNA (Han, 2010). Non-LTR retrotransposons code for a multifunctional enzyme (ORF2) with a Reverse Transcriptase domain, and typically an endonuclease domain with DNA binding and cleavage capabilities; many non-LTR transposons also code for a second protein (ORF1) which has RNA binding activity (BURT & Trivers, 2009; Han, 2010). Transposition begins with transcription, often regulated by internal promoters, and nuclear export, followed by translation of the open reading frames, and the assembly into a ribonucleoprotein particle (RNP) which is then imported back into the nucleus, and integrated via target-primed reverse transcription (Han, 2010). That involves creating an opening in the genomic DNA at the insertion site, which is likely done by the endonuclease, and which allows the RNA intermediate of the TE to bind to the DNA and act as a template for the minus strand synthesis (Han, 2010). The minus strand then most likely acts as a template for the plus strand synthesis, which is to be inserted on the complementary strand of the target site, resulting, based on the transposon, in a target site duplication, target site deletion, or blunt insertion (Han, 2010).

3) Non-autonomous transposons:

Some transposons are not capable of producing their own transposition machinery, and end up relying on other transposons' replication mechanisms. For example:

Short Interspersed Nuclear Element (SINE): family of non-autonomous non-LTR retrotransposons. Often, non-LTR retrotransposons lose the ability to produce their own

Endonuclease and Reverse Transcriptase, and become non-autonomous, relying on the functional proteins of other active and autonomous non-LTR retrotransposons in order to replicate.

Processed pseudogenes: mRNA reverse transcribed and inserted into genome by LINE machinery (see section below on active rearrangements of retrotransposons).

Transposition Is a Parasitic Trait

Due to their disruption of the host genome through insertional mutations and chromosomal rearrangements (see section below), TEs can be detrimental to host fitness. For example, Pasyukova et al. compared isogenic *Drosophila melanogaster* lines which have varying numbers of TEs and found that TE number is negatively correlated with fitness and egg hatchability, and estimated the fitness cost of a single insertion to be 0.4% (Pasyukova et al., 2004). Furthermore, Houle and Nuzhdin established *Drosophila* mutation accumulation lines (lines on which natural selection is partially suspended to allow for the accumulation of mutations) through brother-sister mating of lines derived from an inbred line, then assayed their fitness and measured the number of TE *copia* insertions they had using in situ hybridization (Houle & Nuzhdin, 2004). They found the number of *copia* element insertions to be correlated with decreases in fitness, with an average of 0.76% loss of fitness per insertion (Houle & Nuzhdin, 2004). Additionally, Nikitin et al. found that transposition can lead to lower fitness: *Drosophila simulans* males that had somatic movement of TE *mariner* (as indicated by mosaic eyes) had shorter lifespans than those with stable *mariner* insertions (though they did not find the same effect in *Drosophila melanogaster*) (Nikitin & Woodruff, 1995).

As a result, TEs are often selected against at the level of the host. Petrov et al. analysed population frequencies of individual TEs in different *Drosophila melanogaster* populations, and found strong evidence for purifying selection against insertions (Petrov et al., 2011). Strength of selection increases with copy number of polymorphic TE, TE length, and recombination rate of insertion site, which is consistent with the hypothesis that selection happens at the level of ectopic

recombination (Petrov et al., 2011). This negative fitness effect of transposition causes host genomes to encode mechanisms to silence TEs (see section below). Perhaps as a result, transposable elements have a low rate of transposition per element compared to the activity of other selfish genes (BURT & Trivers, 2009). One final piece of evidence reflecting the negative fitness effect of transposition is that TEs always encode their own (or, in the case of non-autonomous TEs, rely on another TE's) functional proteins for transposition, and never the hosts' (BURT & Trivers, 2009).

The strong evidence for the negative fitness effects of transposable elements confirms that transposition is a parasitic trait, which is beneficial for the TEs but detrimental to host fitness. However, this does not mean that every single transposable element is itself a parasite; indeed, there are examples of TEs being neutral or even advantageous in specific cases (see below). Since purifying selection will select against most insertions, very rare advantageous insertions are overrepresented in the genome, since they are not subject to this selection, and therefore more likely to be maintained than detrimental insertions. The existence of these examples of advantageous insertions is therefore not inconsistent with the fact that most insertions are harmful, and that transposition is parasitic.

Impact of TE on evolution

Passive rearrangements:

TEs can cause passive chromosomal rearrangements. Since multiple copies of the same element exist in the genome, homologous recombination can occur between two identical TE sequences in different locations in the genome. This could lead to ectopic recombination between sequences near TEs, which could result in deletions, duplications, inversions, and translocations (BURT & Trivers, 2009). For example, transposable element *hobo* was found on both breakpoints of three inversions sampled from natural *Drosophila melanogaster* populations in Hawaii, with a fourth inversion having *hobo* on one of its breakpoints, indicating that the inversions likely occurred as a

result of ectopic recombination involving *hobo* (Lyttle & Haymer, 1992). Rearrangements are not always necessarily detrimental or neutral, but can provide a fitness benefit; for example, inversions can link complimentary alleles to one another.

Active rearrangements:

DNA transposons can cause active rearrangements in several ways. One way is through alternative transposition: instead of excising both ends of a TE, a transposase sometimes excises two ends of two different copies of the TE, resulting in varying recombination based on the relative positions of the two copies (Gray, 2000). For example, alternative transposition of *P-elements* in *Drosophila melanogaster* can lead to recombination in the male germ line (in which homologous recombination does not usually occur) (Svoboda et al., 1995). DNA transposons can also cause alternative end joining, that is the joining of different fragments of DNA resulting from the simultaneous excision of multiple TEs in an order or orientation different from the original, which can produce inversions and translocations (Engels & Preston, 1984). Finally, DNA transposons can cause active rearrangements through misrepair of sequences after an excision event (BURT & Trivers, 2009). When the transposon has been excised, host repair mechanisms meant to fix the flanking host sequence can, in some cases, lead to duplications, deletions, or inversions in that sequence.

Retrotransposons can also cause active rearrangements. The reverse transcription proteins of retrotransposons occasionally insert mRNA, instead of transposon RNA, into the genome, resulting in a duplicated gene which lacks introns and untranscribed promoters, and has a poly-A tail since it is derived from mRNA (BURT & Trivers, 2009). These duplicated genes can also be recognized by the direct repeats of the target site, and are known as processed pseudogenes if they are non-functional (BURT & Trivers, 2009). There are examples of these duplicated genes retaining their function, such as *PGK-2* in humans which codes for a functional version of phosphoglycerate kinase that lacks introns and has the characteristics of a pseudogene (McCarrey & Thomas, 1987). Further, retrotransposons can pick up host sequences and genes on the 3′ end of the retrotransposon during its transcription, which is known as 3′ transduction. These sequences

are then inserted along with the retrotransposon at the insertion site, resulting in a duplication. This happens frequently in human *LINE-1* elements (Goodier et al., 2000). Finally, retrotransposon proteins can reverse transcribe mRNA into cDNA which then recombines with the genomic copy of the gene, thereby removing the introns of the original from the genome (Derr & Strathern, 1993).

Effect on genome size

The transposition of TEs (especially retrotransposons) clearly causes an increase in genome size, and the effects mentioned in the previous sections resulting in insertions and deletions also have obvious effects on genome size. As a result, transposable elements can have drastic effect on genome size. For example, Lynch and Conery found a positive correlation in number of TEs (in all classes), and their total length relative to the genome, with genome size, which suggests that TEs can drive increases in genome size (although another possible explanation of those data could be that larger genomes are more capable of supporting a higher number of slightly deleterious mutations, or that both TE number/proportion and genome size are being driven by a third evolutionary force such as effective population size) (Lynch & Conery, 2003).

Mutagenesis:

TEs can also be a source of mutations, which are responsible for some of the negative fitness effects, but can, in rare instances, be beneficial (see section below on domestication).

TE Role in Altering Gene Expression

By inserting near genes and promoter sequences, TEs can influence the gene expression of their neighbouring genes. For example, several miniature inverted-repeat transposable elements (MITEs) are hypothesized to play regulatory roles in nearby genes such as *Ditto-Os2* which is

hypothesized to be the TATA box of the gene *Kn-1* in rice, *Gaijin-So1* which most likely supplies the polyadenylation signal and site for a sugarcane gene *SGT2*, and *Tourist-Zm11* which provides the core promoter for maize gene *Abp1* (Bureau et al., 1996). This effect can have a significant impact on the transcriptome.

In the human genome, one study showed that almost 25% of analysed promoter sequences were of transposable element origin (Jordan et al., 2003), suggesting that TEs are one of the main drivers of gene regulation. Van de Lagemaat et al. analysed untranslated regions of mouse and human mRNA, and found a similar prevalence TE sequences (van de Lagemaat et al., 2003). They found associations between TE insertions and different expression patterns of human and mice orthologues, and associations between certain insertions and tissue specific promoters, indicating that these insertions can alter expression patterns, and drive tissue specific expression (van de Lagemaat et al., 2003). They also found that TEs are disproportionally present in promoters/transcripts of rapidly evolving genes, suggesting a role in the rapid evolution and functional diversification of gene families (van de Lagemaat et al., 2003).

Another way TEs alter expression is through antisense promoters. The 5′ untranslated region of some TEs, such as *L1* retrotransposons in humans contains an antisense promoter, which drives transcription in an antisense direction into the region upstream of the insertion, thereby creating a chimeric transcript (Nigumann et al., 2002). This can also occur in non-autonomous SINEs (Ferrigno et al., 2001).

This capacity to alter expression could have very negative fitness effects. For instance, genome reprogramming of germ-line cells in mice involves TE silencing, and relies on proteins Miwi2 and Dnmtl3. Mice deficient in these proteins are not fully protected against transposition. Vasiliauskaitė et al. found that in such mice, transposable element *IAP* is active and results in the upregulation of many genes, which contributes to a deregulated transcriptome, potentially underlying spermatogonial dysfunction (Vasiliauskaitė et al., 2018).

Altering gene expression can also occasionally produce positive fitness effects. The involvement of TE insertions in insecticide resistance is well documented (Rostant et al., 2012). This involvement is often a result of the transposable element altering expression patterns of the nearby genes. One such example is the *Doc* element insertion in the region of the *Cyp6g1* gene, which we discuss in the section on *Doc* elements.

Transposable Elements Domestication:

In rare cases, transposable elements can be co-opted by the host to play an active role in a host's biology, and produce a positive fitness effect on the host. We discuss a few such domestications:

TE role in telomeres:

In *Drosophila*, which lacks telomerase, telomeres are formed of tandem repeats of the two non-LTR retrotransposons *Het-A* and *TART* which fall in the *jockey* clade (Pardue & DeBaryshe, 2003). Abad et al. identified a third telomeric retrotransposon *TAHRE* which shares a common ancestor with the other two, and whose structure suggests that existing non-LTR retrotransposons were co-opted by *Drosophila* to fulfil the function of telomere maintenance (Abad et al., 2004b). piRNA pathway proteins are used to regulate this (see piRNA pathway section).

RAG1 and RAG2 in vertebrate immunity

To recognize antigens, vertebrate lymphocytes rely on a diverse set of immunoglobulins and T-Cell Receptors which are encoded by modular genes assembled from smaller segments (Agrawal et al., 1998; Tonegawa, 1983). The assembly of these segments, known as V(D)J recombination, which generates the combinatorial diversity of the resultant genes, relies on the genes *RAG1* and *RAG2*, which code for proteins that act as transposases excising the segments and inserting it into the target DNA (Agrawal et al., 1998; Hiom et al., 1998). The facts that *RAG1* and *RAG2* have a

compact genomic organization similar to that of TE genes, that the V(D)J process is highly similar to DNA transposition, and that the end product contains a duplication at the insertion site suggest that *RAG1* and *RAG2* are domesticated TE genes which have been adapted to a different function by the host (Agrawal et al., 1998).

Chimeric TE-gene proteins

Transposable element insertions can also lead to the coding of TE-gene chimeric proteins (that is, proteins derived from both TE and gene sequences), which could lead to novel functions in the host. Examples of chimeric genes have been found are at various levels of domestication, and several studies have looked for TE sequences in transcripts and protein coding sequences of genes.

Bioinformatic searches

Several bioinformatic studies have been done to look for TE sequences within transcripts, and protein coding sequences, usually using the software REPEATMASKER or BLASTN and BLASTP(Almeida et al., 2007; Britten, 2006; Ha et al., 2012; Lipatov et al., 2005; Lockton & Gaut, 2009; Lorenc & Makałowski, 2003; Nekrutenko & Li, 2001; Sela et al., 2010; M. Wu et al., 2007). Generally speaking, there is a higher percentage of such sequences in transcripts than in proteins, most likely due to quality control by host machinery at the level of translation, though also possibly due to biases in available sequences and annotations (Almeida et al., 2007; Gotea & Makałowski, 2006; M. Wu et al., 2007). In the next few paragraphs we will discuss examples of these bioinformatic searches.

Vertebrates:

Several studies have looked for TEs in transcripts and protein sequences of vertebrates. Lorenc and Makalowski found that 0.38% of protein sequences in vertebrates contained transposable element insertions, and that the most frequent function these proteins have is nucleic acid binding (Lorenc & Makałowski, 2003). That function makes sense since TEs contain DNA binding machinery. They argue that most insertions into open reading frames originate either from transposition into an exon or the exonization of an intronic TE (Lorenc & Makałowski, 2003). The latter, which is more common, can occur through either the creation of a whole exon originating from TE sequence, or the silencing of a splice mechanism and the extension of an already existing exon with TE sequence (Lorenc & Makałowski, 2003).

Within vertebrates, humans have been the focus of many such studies. Nekrutenko and Li looked for TEs in coding regions in humans, and found 533 genes (around 4%) had TE insertions into protein coding sequences (Nekrutenko & Li, 2001). They note that the majority of these insertions were insertions into introns which were later exonized as opposed to insertions into exons which are more disruptive (Nekrutenko & Li, 2001); this is consistent with the findings of Lorenc and Makalowski. Britten compared TE sequence translations with human coding sequences and found even more (4653) instances of human protein coding sequences matching with transposable element sequences (Britten, 2006). Britten notes these are likely the result of much fewer insertion events followed by duplications of these sequences (Britten, 2006). Wu et al. found that 4.4% of human genes contain a TE segment in their transcript, and that these segments are usually alternatively spliced, with patterns indicating that they originate as TEs which are turned to exons through alternative splicing (M. Wu et al., 2007). In manually reviewed protein sequences, they found at least 0.35% of human proteins had TE-encoded segments, which is lower than the percentage of transcripts possibly due to mechanisms controlling translation (M. Wu et al., 2007).

Other vertebrates have also been the subject of similar studies. Almeida et al. found that 2.37% of genes in *Bos taurus* have TEs in their exon sequence, such as LOC538046 which is similar to zinc finger 452 and has an exon closely matching *Charlie 10* DNA transposon (Almeida et al., 2007).

They hypothesize that zinc finger protein 452 went through an exaptation event possibly following a duplication after the TE insertion of *Charlie*, acquiring a new function as a result of the insertion (Almeida et al., 2007). Another vertebrate studied is the pig: Ha et al. found that 9.7% of the porcine reference mRNA sequences were chimeric with TE sequences (Ha et al., 2012).

invertebrates:

Sela et al. found that exonization is also present in invertebrates, albeit to a lesser extent than vertebrates, and that it is usually accompanied by alternative splicing which maintains the original transcript (Sela et al., 2010). In *Drosophila*, Lipatov et al. conducted an in silica search for chimeric TE-gene transcripts (supported by expressed sequence tags), and found that fewer than 1% of *Drosophila* genes produce such mRNAs (Lipatov et al., 2005). Lipatov et al. also found that the number of chimeric TEs is lower than expected relative to the total number of TEs, suggesting that chimeric TEs in general are more deleterious than non-chimeric TEs, and that ~80% of chimeric TEs are immediately purged from the genome due to purifying selection (Lipatov et al., 2005). Their population genomic analysis concludes that the majority of observed chimeric TE insertions are deleterious, but that a small proportion, such as a *Doc* insertion in gene *CHKOV1*, might lead to novel gene sequences with adaptive effects (Lipatov et al., 2005).

plants:

Chimeric genes also exist in plants. Lockton and Gaut found that 7.8% of expressed genes of *Arabidopsis thaliana* had chimeric TE sequences in their protein coding regions (Lockton & Gaut, 2009). They used phylogenetic analysis to determine whether these sequences are due to TE insertions affecting the gene or TEs incorporating gene sequences, and estimated that in as much as 1.2% of expressed annotated genes, TEs have contributed protein segments (Lockton & Gaut, 2009).

A potential problem with bioinformatic searches for chimeric genes or exonization is that it is often not enough to detect TE sequences in transcripts or proteins, because that could simply be a

TE which has taken up a gene sequence rather than the other way around. One way some authors such as Lockton and Gaut are able to resolve this is by looking at the phylogenetic trees of the TE and the gene, to see if the sequence is novel in either of them; if it is novel in the host and conserved in the TE, that usually indicates that it is derived from the TE (Lockton & Gaut, 2009).

Examples of chimeric genes

transposase

Several interesting examples of chimeric genes have been found which contain entire transposase genes domesticated from DNA elements. For example, Cordaux et al. found that the primate chimeric gene *SETMAR* (aka *metnase*), which is derived from *SET* a primate histone methyltransferase gene and the transposase gene of *Hsmar1* transposon, was derived following the insertion of *Hsmar1* 40-58 million years ago when the transposase-coding region of *Hsmar1* was captured as an exon and fused to the *SET* transcript through a deletion of the stop codon of *SET* and the creation of a new intron (Cordaux et al., 2006). This was not a single event, but rather a stepwise conversion over millions of years (Cordaux et al., 2006). Using an in vitro assay, Lee et al. found that *SETMAR* is a histone methyl transferase gene, suggesting that its *SET* domain retains its function (Lee et al., 2005). Liu et al. used a different in vitro assay to determine that *SETMAR* retains several of the transposition functions of *Hsmar1*, such as site specific DNA binding, and DNA cleavage (D. Liu et al., 2007). Tellier and Chalmers overexpressed wild-type and methyltransferase deficient *SETMAR* in cells, then determined through immunoprecipitation and RNA sequencing that it targets *Hsmar1* terminal repeats and regulates the expression of the genes containing them using the methyltransferase function of *SET*, thereby combining the biochemical functions of both the gene and the transposable element insertion sequences into a new cellular function (Tellier & Chalmers, 2019).

Another example was found in a computational genomics analysis of human genome and expressed sequence tag data by Sarkar et al. who identified 5 genes which contain a domesticated *piggyBac* element, including *PGBD1,* which is a chimeric gene containing a SCAN domain followed by *piggyBac* (Sarkar et al., 2003). Belbin et al. found that *PGBD1* is associated with age-at-onset of late-onset Alzheimer's disease (Belbin et al., 2011). Tipney et al. identify and characterize a third example: the gene *GTF2IRD2* which contains a *CHARLIE8* transposase-like domain which likely came about due to a TE insertion resulting in a chimeric gene-TE protein (Tipney et al., 2004). They predict that the protein retains some transposase functionality, and suggest it might be involved in the pathology of Williams-Beuren Syndrome (Tipney et al., 2004).

DNA-binding Domain

There are other examples of chimeric genes arising through the recruitment of DNA binding domains from TEs (Sinzelle et al., 2009). For example, Belenkaya et al. describe a chimeric *P-PH* protein formed from the DNA binding domain of the *P-element* and a partial sequence of the *Polyhomeotic* (*PH*) protein in *Drosophila melanogaster* (Belenkaya et al., 1998). Using immunostaining, they found that this lead to the localization of *PH* and other downstream proteins at *P-element* sites; they also found that this lead to the repression of the expression of *P-element* induced alleles of gene yellow (Belenkaya et al., 1998). This suggests that the insertion led to the localization of the protein at a different site, resulting in a novel function.

The *P-element* DNA binding domain is involved in another domestication event described by Roussigne et al. who identified a protein domain they termed THAP which shares a similarity with the *Drosophila P-element* DNA binding domain, and is conserved in vertebrates such as humans and mice, and invertebrates such as worms and *Drosophila* (Roussigne et al., 2003). Quesneville et al. observe that THAP domains are similar to DNA-binding domains in *P-elemen*t neogenes, which are recurrent molecular domestications of *P-elements*; they hypothesize that THAP proteins

acquired this domain as a result of recurrent molecular domestications of *P-elements* (Quesneville et al., 2005). Proteins having the THAP domain have roles in cell proliferation, cell-cycle control, and apoptosis (Sinzelle et al., 2009).

Another example of a domesticated DNA binding domain is the BED finger, which was characterized by Aravind in insulator proteins BEAF and DREF (Aravind, 2000). Aravind found evidence that this domain was recruited from TE transposases on two or more independent occasions for a cellular function (Aravind, 2000). Specific DNA binding is necessary for the insulator function of BEAF, which targets sites on chromosomes and affects chromatin structure (Gilbert et al., 2006).

retrotransposon

Examples of TE-gene chimeras and exonization are not limited to DNA transposons, but can also be seen in retrotransposons. For example, Bae et al. found that the first exon of horse gene *BMAL1*, which is involved in regulating circadian rhythm, is produced by the exonization of *LINE3* (*CR1*) and SINE (*MIR*) sequences (Bae et al., 2011).

detection problem

A potential problem to keep in mind with detecting examples of such genes is that once the TE is taken up by the host genome, and a novel function is introduced, the selective pressures on the sequence can lead to changes in the TE sequence, causing it to diverge from its original version. As a result, studies looking for chimeric genes will be biased towards newer insertions. However, Gotea and Makałowski argue that young TE cassettes are not likely to be translated into proteins, even if they are transcribed into mRNA, due to quality control mechanisms before or after translation, and that even if these TEs were translated, the proteins they produce would be non-functional or even deleterious (Gotea & Makałowski, 2006). According to Gotea and Makałowski, most exapted TE sequences are from old TE insertions that have had plenty of evolutionary time

to adapt (Gotea & Makałowski, 2006); however, other studies such as that of Wu et al. find many functional proteins with translated sequences from young TE cassettes (M. Wu et al., 2007). Either way, we must seriously consider the possibility that chimeric TE-gene insertions are underrepresented in the literature.

Insertion Position Preference

One way TEs can mitigate their fitness cost to the host is by having preference for insertion sites which are less harmful to the host; indeed some transposable elements have preferences for insertion sites (Levin & Moran, 2011). Some transposable elements will target gene rich regions, such as *Ty1* in yeast (Devine & Boeke, 1996). However, it might be advantageous for transposable elements to avoid ORFs of genes since inserting into ORFs might disrupt the gene's function, thereby decreasing the fitness of the host, and making it less likely the TE is passed to another generation. For example, *Ty1* inserts into a specific insertion site upstream of the transcription start site of tRNA genes (Devine & Boeke, 1996), and *P-element* avoids open reading frames, inserting instead near or within promoters, upstream of genes (Bellen et al., 2011).

Alternatively, several TEs preferentially target the heterochromatin. For example, Zou et al. found that *Ty5* preferentially targets silent chromatin in yeast, which is analogous to heterochromatin in other eukaryotes (Zou et al., 1996).

Transposition Rate

The rate at which TEs transpose is influenced by a complex array of factors, including selection at the levels of both host and TE which exert different pressures which ultimately determine the rate. Hosts will evolve mechanisms to alter transposition rates (i.e. defences against transpositions), and selection on TEs can affect their transposition rates by varying their transposition mechanisms (to either drive transposition up or down) or by potentially generating 'self-regulation' mechanisms which lower transposition rates under certain conditions. Charlesworth and Langley have shown

that given a high enough rate of dominant lethal transpositions, there is a theoretical advantage in TEs regulating down their transposition rates in diploid sexual organisms (Charlesworth & Langley, 1986). However, those conditions are quite stringent, and may or may not be satisfied in natural populations, which could make self-regulation a difficult trait for TEs to evolve.

But why might selection on TEs drive transposition rates down? The intuitive view on transposition rate is that a higher transposition will lead to more copies for the TE, and therefore a fitness advantage is conferred onto the TE. However, natural selection on the TE need not necessarily favour an increased transposition rate. Increased transposition can decrease the fitness of the host organism (too many transposition events could overwhelm DNA repair mechanisms for example), which decreases the fitness of the TE since it's less likely to be passed down. As a result, there may be selection on the TE to have a lowered transposition rate.

There are some data which are consistent with the possibility that lowered transposition rates are naturally favoured (though they are by no means definitive proof that that is the case). Lampe et al. created *Himar1 mariner* mutants with around 12x more transposition activity (Lampe et al., 1999), showing that wild type transposases are not at their maximum level of transposition. Similarly, Mates et al. were able to create a mutant *SB100X* of the vertebrate transposon *Sleeping Beauty* which has a hyperactive transposase leading to 100-fold increase in transposition efficiency (Mátés et al., 2009).

Not only is there evidence that transposable elements transposing at a sub-maximal rate, there is also direct evidence that TEs sometimes regulate down their own transposition. For example, Jiang found that in the yeast *Saccharomyces cerevisiae* (which lacks the RNA interference pathway involved in TE silencing), TE *Ty1* can switch to a co-suppressed state in which all the copies of the TE in the genome become inactive as a form copy number control (Jiang, 2002). This co-suppression is dependent on copy number, and relies on a transcription product of *Ty1* (Jiang, 2002). Saha et al. found that this co-suppression is due to a truncated form of the Gag protein, which is encoded by internal transcripts of *Ty1*, and is trans-dominant, that is, it can silence all copies of *Ty1* once it is formed (Saha et al., 2015). Another example of a TE regulating its own

transposition rate might be the *Mos1 mariner*-like element, which Lohe and Hartl found exhibits autoregulation in *Drosophila mauritiana* in the form of overproduction inhibition, in which increased expression of the transposon results in less excision (Lohe & Hartl, 1996).

In general, it is in the TE's evolutionary interest to minimize its harm to the host, and balance that with its transposition rate, as indicated by Burt and Trivers in the equation R=t/s, where R is the reproductive rate of the TE to be maximized, t is the transposition rate, and s is the fitness cost to the host (BURT & Trivers, 2009). To accommodate these opposing evolutionary forces, TEs must maintain a balance between expression and repression (Bourque et al., 2018).

This balance is further complicated when we factor in the host defences, which will affect the actual transposition rate and make it different from the one encoded by the TE. After all, though selective pressures on the TE will push for an 'ideal' rate that will maximize TE fitness, selective pressures on the host will push for a different rate (most likely 0), which is more ideal for the host fitness. Hosts defences will therefore drive down transposition rates. For example, in *Drosophila melanogaster*, P-M hybrid dysgenesis (in which the male parent has a *P-element*, and the female parent does not) results in an increased transposition rate of *P-elements* in the offspring when compared to non-dysgenic lines (in which both parents have the *P-element*) (Eggleston et al., 1988). This is due to some piRNAs involved in TE silencing of *P-elements* being maternally passed down through the cytoplasm (Brennecke et al., 2008), and being therefore absent in the offspring of females without *P-elements*. We will discuss host defences in depth in the section below.

The 'ideal' rate from the TE's perspective is not constant, and will vary based on whether the TE is in the germline or the soma, which can cause TEs to have variable expression between the two tissue types. In so far as TE activity is a mechanism to create more copies of the TE, and propagate its genes, it is only important to TE fitness when it occurs in the germline, whereas somatic expression can be detrimental to host fitness (as seen in Nikitin & Woodruff, 1995), and therefore TE fitness. As a result, many TEs evolve to be active in germline cells and be silent in somatic cells (Haig, 2016). For example, Chain et al. found a regulatory region in the *P-element* of *Drosophila*

melanogaster which splices an intron out of the transcript in the germline, thereby allowing for transposition to only occur in the germline (Chain et al., 1991). Other TEs contain binding sites for transcription factors that activate them in germline cells (such as Nanog or Oct4) or inactivate them in somatic cells (such as p53) (Haig, 2016). Despite all that, there is evidence that anti-TE piRNA pathways are active in somatic cells in most arthropod species (Lewis et al., 2018) suggesting transposition occurring in somatic cells, and several examples of TEs expressing in somatic cells in mammals (Haig, 2016) and even *Drosophila* (Nikitin & Woodruff, 1995), which means that this variable transposition is different in different host species and TE types.

Transposable elements can also alter their transposition rates based on other factors such as habitat expansion (Kofler et al., 2015). This minimizes fitness costs while boosting the copy number propagated, conferring a fitness advantage to transposable elements, allowing them to spread rapidly through a population. Coupled with the capacity to move horizontally between species, that can allow a transposable element to rapidly spread through an entire species undergoing habitat expansion. For example, the *P-element* is not native to *Drosophila melanogaster*, but rather has been introduced to it horizontally from a distant relative *Drosophila willistoni* (Engels, 1992) as the former colonized the territory of the latter (Kofler 2015), and has become ubiquitous in wild populations (Anxolabéhère et al., 1988; Ozata et al., 2019).

Mechanisms of TE Resistance

piRNA pathway: PIWI-interacting RNAs (piRNAs) are small (21-35 nt) animal specific class of RNA (Ozata et al., 2019), which bind to Piwi proteins, a subfamily of Argonaute proteins involved in germ-line specific events (Girard et al., 2006). piRNAs arise mainly from single-stranded RNA precursor (Vagin et al., 2006), which is transcribed from genomic piRNA clusters (Ozata et al., 2019). In *Drosophila* (as in other arthropods), these piRNA clusters include sequences derived from transposable elements, and allow the piRNA pathway to regulate TE expression in the germline (Brennecke et al., 2007). *Drosophila* piwi proteins are mostly limited to germline cells, the exception being certain somatic lineages such as ovarian somatic sheath (OSS) cells (Ross et al., 2014). The majority of the arthropods, including their most recent common

ancestor, have somatic piRNAs involved in TE defence; *Drosophila*, however, are one of few arthropods to mostly lack somatic piRNAs (with the exception of specific somatic lineages) (Lewis et al., 2018).

In the germline, antisense piRNA strands are processed from piRNA clusters, and bind to Piwi proteins such as Aubergine (Aub) to form a complex that targets transposable element transcripts (or other target transcripts) (Brennecke et al., 2007). The subsequent cleavage generates a sense piRNA strand which then binds to Ago3, forming a complex that targets piRNA cluster transcripts, thereby regenerating an antisense piRNA, and amplifying the signal (Brennecke et al., 2007). This process is known as ping-pong amplification. In flies, Aub and Ago3 target TE transcripts for cleavage in the cytoplasm (Brennecke et al., 2007); Piwi is expressed in the nucleus and is involved in the targeting mechanism for heterochromatin formation, preventing the expression of TEs (S. H. Wang & Elgin, 2011).

In addition to its role in regulating TE expression, the piRNA pathway is used by some organisms such as *Aedes egypti* in fighting viruses (Morazzani et al., 2012; Schnettler et al., 2013). However, Petit et al. found that *Drosophila melanogaster* does not produce viral piRNAs, and that mutants for piRNA pathway genes are not more susceptible to viruses DCV, DXV, and SINV (Petit et al., 2016). This indicates that the piRNA pathway does not have an antiviral function in fruit flies (Petit et al., 2016).

The piRNA pathway evolves rapidly (Lewis et al., 2018), which explains the variety and diversity of secondary functions it can have (Sarkar et al., 2017), such as possibly regulating mRNAs, particularly in mammals (Ozata et al., 2019). Another secondary function was discovered by Kiuchi et al. in silkworms, which rely on piRNA precursor *Fem* that codes for a piRNA involved in silencing the gene *Masc* in females, leading to the production of an isoform of *Bmdsx* and subsequently female sex determination (Kiuchi et al., 2014). In *Drosophila melanogaster*, Khurana et al. used ChIP show that piRNA pathway proteins encoded by *armi* and *aub* are necessary for the telomere binding of HOAP, which is a part of the telomere protection complex, indicating a role for piRNA pathway proteins in telomere protection (Khurana et al., 2010). Interestingly,

genes *rhi* and *ago3*, which also code for piRNA pathway proteins do not seem to be involved, suggesting that this secondary function does not involve the piRNA pathway, but rather individual proteins from that pathway (Khurana et al., 2010). Savitsky et al. found that *Drosophila melanogaster* mutants for piRNA pathway genes *aub* and *spn-E* had increased telomeric element retrotransposition, suggesting that the piRNA pathway controls the process of telomere formation (Savitsky et al., 2006). Radion et al. found that telomeric piRNA clusters facilitate the assembly of telomeric chromatin via an interaction with the piRNA pathway protein Rhino, and are important in the nuclear positioning of the telomeres (Radion et al., 2018).

Endogenous siRNAs: Ghildiyal et al. found endogenous siRNAs (see section on siRNA) in somatic cells of *Drosophila melanogaster* which map to transposable elements (Ghildiyal et al., 2008), and Czech et al. found endogenous siRNAs mapping to TEs in both somatic and germ line cells (Czech et al., 2008), indicating that the siRNA pathway is used as TE defence in somatic and germ line *Drosophila* cells.

Methylation: One type of DNA modification which eukaryotic cells can do is cytosine methylation. A cytosine, usually followed by a guanine (BURT & Trivers, 2009) is methylated with the addition of a methyl group. Some organisms defend against transposable element insertions by methylating the TE sequence, which prevents its transcription, and eventually leads to the destruction of the transposon (Yoder et al., 1997).

Other: Cytosine deaminases such as APOBEC3 can inhibit transposition activity of TEs (Chiu & Greene, 2008; Muckenfuss et al., 2006). Furthermore, DNA repair factors can limit transposition. For example, proteins involved in NHEJ cause a truncated version of the transposon *Zfl2-2* to be inserted in DT40 cells, which lacks the coding sequences to be transposed again in subsequent generations (Suzuki et al., 2009).

Doc Elements

In this thesis, we will explore the role of certain transposable element insertions in the evolution of virus resistance in *Drosophila melanogaster*. Specifically, we will focus on an insertion in the gene *Veneno* and an insertion in the gene *CHKov1*. Both of these insertions are *Doc* transposable element insertions. In this section we describe what is known about *Doc* transposable elements.

Doc transposable elements are non-LTR retrotransposons found in *Drosophila* (Driver et al., 1989; O'Hare et al., 1991). They belong to the Jockey clade of non-LTR retrotransposons (Berezikov et al., 2000). They are around 4.7kb long, and contain two open reading frames (ORFs) which are ~1.7 kb (ORF1) and ~2.9 kb (ORF2) long (O'Hare et al., 1991). Consistent with other non-LTR retrotransposons (see Transposable Elements Section), ORF1 of *Doc* is predicted to code for a protein with a nucleic acid (possibly RNA) binding function, whereas ORF2 is predicted to code for a protein which has a reverse transcriptase function (O'Hare et al., 1991). The protein coded by ORF1 of the *Doc* element has been shown by Rashkova et al. to localize to irregular clusters in the cytoplasm without entering the nucleus (Rashkova et al., 2002).

Doc elements are similar in structure to transposable elements *F*, *G*, *Jockey*, and *I* factors, which are also non-LTR retrotransposons (Driver et al., 1989; O'Hare et al., 1991). They all lack terminal repeats, have an adenine rich tail (a poly-A tail in the case of *Doc*), insert with flanking duplications (6 to 13 bp in the case of *Doc*), have conserved 3' ends but sometimes truncated 5' ends, and a putative reverse transcriptase domain in their sequence (Driver et al., 1989; O'Hare et al., 1991; Vaury et al., 1994).

Like other non-LTR retrotransposons, *Doc* element are present in three forms: full length elements, truncated elements which have been recently transposed but are no longer transcribed due to the truncation, and elements which have long been immobilized by truncation (Vaury et al., 1994). The latter form is found in common between unrelated strains of *D. melanogaster*, and contains accumulated mutations including deletions which tend to occur more frequently in the 3'

end of the transposable element (Vaury et al., 1994). The presence of these old insertions suggests that the *Doc* element has colonized *D. melanogaster* long ago (Vaury et al., 1994).

Vaury et al. found that the *Doc* element is transcribed as poly-adenlyated full-length RNA ~5kb in length (Vaury et al., 1994). They found that the transcript was more abundant in adults than in larvae, and had no apparent tissue specificity (Vaury et al., 1994). However, Zhao and Bownes later found that the *Doc* transcript is more abundant in the germ-line and that in the oocyte it localizes to the region containing the cytoskeleton (Zhao & Bownes, 1998). Expression of the *Doc* element is regulated by a sequences internal to its transcript, ~20bp downstream of the transcription start site known as the B region (CONTURSI et al., 1995; MINCHIOTTI et al., 1997). In some *D. melanogaster* strains, the *Doc* element is unstable, and undergoes frequent transposition (Pasyukova & Nuzhdin, 1993).

Doc has been shown to insert in or near genes and generate novel phenotype often by affecting the transcript sequence (as in Ruiz-Vazquez & Silva, 1999) or affecting the levels of expression (as in Kronhamn & Rasmuson-Lestander, 1999; Page et al., 2007; van Beest et al., 1998). Here we detail three specific examples:

In *white*:

The *white-one* (*w^1*) mutation in eye pigmentation isolated by Morgan in 1910 is the very first mutation characterized in *D. melanogaster* (Morgan, 1910). It started off the field of *D. melanogaster* genetics, and led to the discovery of sex-linked traits (Morgan, 1910). 79 years later, Driver et al. identified the cause of this famous mutation: a *Doc* element insertion in the sequence of the *white* gene (Driver et al., 1989). The insertion occurs near the transcription start site of *white*, and likely acts either by inactivating the promoter (Driver et al., 1989; O'Hare et al., 1991), or terminating transcription (O'Hare et al., 1991). Since the *Doc* element is a non-LTR retrotransposon, it is not excised as a part of the transposition process, and therefore this mutation is stable (O'Hare et al., 1991). However, secondary insertions of other transposable elements *into*

the *Doc* insertion in w^1 result in further phenotypes such as *white-honey* and *white-eosin* (O'Hare et al., 1991).

Upstream of *Cyp6g1*

Schlenke and Begun found extremely decreased heterozygosity in a 100kb region present in a California population relative to an African population, which indicates a likely selective sweep in the California population, likely due to a beneficial mutation (Schlenke & Begun, 2004). They find that a *Doc* insertion upstream of cytochrome P450 *Cyp6g1* in *Drosophila simulans* is associated with this decreased heterozygosity (Schlenke & Begun, 2004). The insertion occurs ~200bp upstream of the putative transcription start site, and its presence is associated with at least a two-fold increase in *Cyp6g1* transcript in both *D. simulans* and *D. melanogaster*, suggesting that the insertion up-regulates the gene (Schlenke & Begun, 2004). This is particularly interesting since the upregulation of *Cyp6g1* is correlated to DDT resistance (Daborn et al., 2001, 2002), suggesting that the *Doc* insertion may have played an adaptive role in providing the DDT resistance resulting in a selective sweep across the population, which would be consistent with the large decrease in heterozygosity around the insertion (Schlenke & Begun, 2004).

In *CHKov1*

Aminetzach et al. found that the Doc1420 insertion in the coding sequence of *CHKov1* led to the generation of two new derived transcripts (Aminetzach et al., 2005). The data supporting these results are not described and the sequencing data was not made publicly available. Subsequent efforts to replicate these results have failed (Cao, unpublished data). This insertion was followed by a selective sweep which led to the allele with insertion to be present in ~80 of *D. melanogaster* strains worldwide (Aminetzach et al., 2005). Aminetzach et al. attributed the apparent positive selection to the *Doc1420* insertion, presenting evidence that the insertion is associated with resistance to the pesticide azinphos-methyl- phosphate (AZM) (Aminetzach et al., 2005). However, in a genome wide association study of AZM resistance, Battlay et al. found no

significant association between the Doc1420 insertion and AZM resistance (Battlay et al., 2016). This result could be different due to Battlay et al's use of larvae as opposed to adult flies which were used by Aminetzach et al. (Battlay et al., 2016). However, subsequent efforts to replicate Aminetzach's results have failed (Battlay, unpublished data). This casts some doubt on *Doc1420's* role in AZM resistance.

Another possible explanation for the selective sweep on the *Doc1420* allele is its association with DMelSV resistance: Magwire et al. found that the insertion is associated with a highly significant increase in resistance to the sigma virus (Magwire et al., 2011). Furthermore, another polymorphism involving *CHKov1* involves two duplications of parts of the *CHKov1* (including parts of the *Doc* element sequence) as well as the neighbouring and related gene *CHKov2* (Magwire et al., 2011). This pair of duplications was found to be associated with a massive increase of resistance to DMelSV, with flies carrying the duplications having a 79-fold lower viral titres than flies carrying the Doc1420 allele without the duplications (Magwire et al., 2011). Furthermore, the flies having the duplications were found to have higher expression of *CHKov2* (and not *CHKov1*) (Magwire et al., 2011) which might explain the emergence of this phenotype. However, the primers used to measure gene expression amplified a region of the transcript downstream of the *Doc1420* insertion, so they would not amplify any transcripts lacking this region of the gene. Alternatively, the duplicated parts *CHKov1* carrying parts of the Doc1420 insertion might be involved in this phenotype, particularly since the original insertion leads to increased DMelSV resistance.

Methods

DGRP

We used the Drosophila Genetic Reference Panel (DGRP) which is a panel of fly lines that have been collected from the wild and highly inbred (Mackay et al., 2012).

Fly Care

We kept flies either at 25°C or 18°C with cornmeal media as well as yeast. Flies were moved to new food every 20 days at 25°C or 40 days at 18°C.

DNA Extraction from flies

For sequencing, we extracted DNA from flies by homogenizing 5-12 flies using plastic pellet pestle in buffer ATL (Qiagen) then using the Qiagen DNeasy Blood & Tissue Kit according to the manufacturer's protocols.

For *Veneno* genotyping, we used Chelex extraction. We added 150µl of 5%(w/v) Chelex suspension prepared with Chelex 100 (Sigma) and 1mm Zirconia beads to 1 fly in a PCR tube, then homogenized using a TissueLyser for 4 minutes at 30Hz. We added 1.5µl proteinase K and incubated overnight at 56°C then centrifuged, removed 100µl of supernatant, and inactivated proteinase K by incubating at 94°C for 15 minutes.

Quantifying DNA and Assessing Purity

We quantified DNA using a Qubit™ 2.0 Fluorometer and the Qubit dsDNA HS Assay Kit (ThermoFisher) following manufacturer's protocols. We made working solution by diluting Qubit reagent 1:200 in Qubit Buffer. We made two standards to calibrate the machine by adding 10µl of standard 1 and 2 from the kit to 190µl of working solution. Then we prepared our DNA sample for measurement by diluting 1µl of DNA sample in 199µl of working solution. We then measured DNA concentration using the Qubit 2.0 fluorometer (Invitrogen).

We used a NanoDrop ND-1000 Spectrophotometer to assess the purity of DNA samples following manufacturer's protocols. We checked for normal spectrum shape with appropriate 260/280 and 260/230 ratios.

TouchDown PCR Taq

The general method we used for amplifying DNA was TouchDown. For a 20µl reaction we used 15.4µl nuclease free water (nfH_2O), 2µl 5x standard Taq reaction buffer (New England Biolabs), 0.5µl dNTP mix (0.5mM) (New England Biolabs), 1µl primers (forward and reverse both at 10µM in water) (Merck), 1µl template DNA, and 1µl Taq polymerase (New England Biolabs). We used the cycle settings in table ii.1, adjusting the annealing temperatures when necessary:

	Temperature	Time
1x	95°C	2 min
10x	95°C	30 s
	62°C – 1°C per cycle	15 s
	72°C	1min/kb
25x	95°C	30 s
	52°C	15 s
	72°C	1min/kb
1x	68°C	4 mins

Table ii.1 Cycling Conditions for TouchDown PCR with Taq polymerase

Agarose Gels

All gels we used were 1% (w/v) agarose TAE gels with Ethidium bromide (1µl/50µl of gel). Samples were mixed with 6x DNA Loading Dye (Thermo Scientific). We used HyperLadder™ 1kb (Bioline) as our ladder.

ExoSap PCR Clean-up

When sequencing PCR products, we cleaned up the DNA after the PCR reaction using Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (EXOI) (New England Biolabs) for enzymatic PCR clean-up. We ran 5µl of PCR product in 10µl reactions with 1µl SAP and 0.1µl EXOI (and nfH_2O to bring the volume up to 10µl). We incubated the sample at 37°C for one hour, then at 72°C for 15 minutes.

Sanger Sequencing

For Sanger sequencing, we used BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) to run our sequencing reaction on plasmids or clean PCR products. In a 10µl reaction, we used 2µl of the BigDye Terminator v1.1 &3.1 5X Sequencing Buffer, 5µl nfH_2O, 1 µl of 3.2µM primer, 1µl of BigDye Ready Reaction Mix, and 2µl of cleaned PCR product of plasmid DNA. We ran the sample using the conditions in ii.2, then sent them to Source Bioscience for reading the results.

	Temperature	Time
1x	96°C	1 min
25x	96°C	10 s
	50°C*	5 s
	60°C	4 min

Table ii.2 Cycling Conditions for BigDye Sequencing Reaction

Infecting flies with DAV

We used Drosophila A Virus which was kindly provided by Dr Karyn Johnson (The University of Queensland, Australia) (Ambrose et al., 2009), replicated in a virus-free isogenic line of Drosophila melanogaster (w1118), then extracted 10 days after infection by immersing the flies in liquid

nitrogen, keeping them frozen at -80˚C to rupture cells, and then homogenizing them in Ringer's solution (4 µl/fly) (as in Ambrose et al., 2009). We diluted the aliquot serially to 10^{-5} its concentration (~3000 x $TCID_{50}$) before infecting 7-15 (unless otherwise specified) mated 3-6 day old female flies per replicate by pricking them in the left pleural suture of the thorax with a 0.15 mm diameter anodized steel needle (Austerlitz Insect Pins) as in Martinez et al., 2016 while they were anesthetized on a CO2 pad. The infected flies were incubated for 72 ± 3 hours (unless otherwise specified) before being frozen with liquid nitrogen then homogenized in 250µl of TRIzol Reagent (ThermoFisher) and frozen. Unless otherwise specified, the DAV resistant DGRP line used is 362.

Infecting flies with DmelSV

We infected flies with sigma virus strain Hap23 prepared as in Magwire et al., 2011. We prepared needles by pulling 3.5 inch long microinjection capillaries made of borosilicate glass with an outer diameter of 1.14mm and an inner diameter of 0.53mm (Drummond), using a Narishige pulling machine. We cut the needles under a microscope using a scalpel to create a sharp long tip for ease of piercing the Drosophila exoskeleton. We then injected 69nl of virus suspension was into the abdomen of 7-15 mated 3-6 day old female flies per replicate using Nanoject II Auto-Nanolitre Injector (Drummond Scientific). We then left the flies to incubate for 12 days, moving them to new cornmeal food vials every three days, before freezing them with liquid nitrogen, and homogenizing them in TRIzol Reagent (ThermoFisher) and storing them at -80°C.

TRIzol RNA Extraction

When not otherwise specified, we extracted RNA using TRIzol Reagent (ThermoFisher) in the following method: We homogenized 5-10 flies in 250µl of TRIzol in TissueLyser at 30Hz for 2 minutes with ~12 1mm zirconia beads, then stored them at -80°C. Samples were defrosted and incubated at room temperature for 5 minutes, then centrifuged at 12000g at 4°C for 10 minutes. We transferred 160µl of supernatant to clean tube, then added 62.5µl of chloroform, mixed well by shaking the tube for 15 seconds, then incubated at room temperature for 3 minutes. We

centrifuged the samples a second time at 12000g at 4°C, then transferred 66µl of the aqueous phase to a clean tube. We added 156 µl of isopropanol, inverted tube twice, then incubated at room temperature for 10 minutes. Then we centrifuged a third time at 12000g at 4°C, removed supernatant, washed pellet with 70% ethanol, then dissolved sample in 15µl nfH_2O, and stored it at -80°C.

Reverse Transcription

To reverse transcribe RNA into complementary DNA, we used GoScript™ Reverse Transcriptase (Promega) and random hexamer primers (RanHex). In the first stage, we incubated 0.6µl of RanHex (50µM), 0.9µl of nfH_2O, and 1µl of RNA at 70°C for 5 minutes, then placed the samples on ice for 5 minutes. Next we added, 2µl of GoScript 5X Reaction Buffer, 1.9µl of $MgCl_2$, 0.5µl of dNTP mix (0.5µM), 0.25µM of RNasin® Plus Ribonuclease Inhibitor (Promega), 0.5µl of GoScript Reverse Transcriptase, and 2.35µl of nfH_2O. We incubated at 25°C for 5 minutes, then 42°C for 1 hour, then 70°C for 15 minutes. We then diluted the samples 1:10 using nfH_2O.

Quantifying RNA and Assessing Purity

When needed, we quantified RNA using a Qubit™ 2.0 Fluorometer and the Qubit RNA HS Assay Kit (ThermoFisher) following manufacturer's protocols. We made working solution by diluting Qubit reagent 1:200 in Qubit Buffer. We made two standards to calibrate the machine by adding 10µl of standards 1 and 2 from the kit to 190µl of working solution. Then we prepared our RNA sample for measurement by diluting 1µl of RNA sample in 199µl of working solution. We then measured RNA concentration using the Qubit 2.0 fluorometer (Invitrogen).

We used a NanoDrop ND-1000 Spectrophotometer to assess the purity of RNA samples following manufacturer's protocols. We checked for normal spectrum shape with appropriate 260/280 and 260/230 ratios.

Measuring Viral Titre

We used two methods to measure viral titre. When not specified, we used method 1.

1) From cDNA using SensiFAST™ SYBR Hi-ROX

To measure viral titre, we reverse transcribed RNA from Trizol extractions. We diluted the reverse transcribed cDNA by a factor of 10 using nfH_2O. We then performed qPCR on the cDNA using StepOnePlus Real Time PCR System (Applied Biosystems) and SensiFAST™ SYBR Hi-ROX Kit (Bioline). For a 10µl reaction, we used 5µl of SYBR, 2µl nfH_2O, (2.5mM), 2µl of 1:10 diluted cDNA, and 1µl of primers for either DAV or DMelSV in the virus reaction, and either Rpl32 or Act5C in the housekeeping gene reaction (see primers in Appendix 19). We followed the 2 step cycling protocol in ii.3.

	Temperature	Time
1x	95°C	2 min
40x	95°C	5 s
	60°C (data collect)	30 s
Melt Curve	95°C	30 s
	60°C+0.3°C until 95°C (data collect)	1 min
	95°C	15 sec

Table ii.3 Cycling Conditions for qPCR with SensiFAST™ SYBR Hi-ROX

We used nfH_2O as a negative control. We verified the detection/non-detection of virus/control in our samples by checking that melt curves have the appropriate peaks for the primer pair. We then calculated C_t values (that is the number of cycles at which the reaction fluorescence crosses the fluorescence threshold) using a threshold ΔR_n value of 1 (ΔR_n is the difference between the normalized experimental reporter value, which is the fluorescent signal generated by SYBR Green, and the normalized baseline reporter value generated from the Rox reference dye in the reaction). We then found ΔC_t by subtracting the C_t value of our housekeeping gene from the C_t value of our virus. Viral titre is proportional to $2^{-\Delta Ct}$ so we used $-\Delta C_t$ as an estimate of $\log_2$(relative viral titre).

2) One-step RT-qPCR

This method was used in Chapter 1 for results presented in Figures 1.1,1.2, and 1.3, and 1.6B. RNA copy number relative to mRNA from the reference gene *EF1α100E* in a one-step RT-qPCR reaction (10µl) using the QuantiTect Virus kit (Qiagen) on an Applied Biosystem Step One Plus Real-time PCR machine. We used nfH_2O as a negative control. Viral and fly cDNAs were amplified in a duplex reaction using virus and fly primers in association with dual-labeled fluorescent probes (Sigma) (See appendix 19).

Creating plasmids with inserts (NEBuilder)

To make plasmid constructs with inserts for transfection into cells or injecting into flies we followed the following procedure.

Preparing Inserts

We used Snapgene (Insightful Science) to design primers to amplify the inserts for NEBuilder: we added tails to the primers to create products which share overlapping bases with the surrounding sequences in the desired plasmid (that is either the surrounding vector sequence, or the adjacent inserts). We picked settings for 15 to 25 overlapping bases with a target T_m of 50°C. The overlaps are required for efficient assembly with NEBuilder. When applicable, we added or removed bases from the tails of primers to adjust the final sequence of the insert when adjustments were desired such as to place the sequence in frame or to generate a stop codon.

Inserts were amplified using KAPA HiFi Hot Start Ready Mix, following manufacturer protocol, and using the cycling conditions in table ii.4.

	Temperature	Time
1x	95°C	3 min
22x	98°C	20 s
	60°C* + 0.5°C per cycle	15 s
	72°C	1min/kb
10x	98°C	30 s
	70°C*	15 s
	72°C	1min/kb
1x	72°C	4 mins

*: adjusted based on primers melting temperature

Table ii.4 Cycling Conditions for insert amplification with KAPA HiFi Hot Start Ready Mix

We then run the PCR product through a 1%(w/v) agarose in TAE gel with 0.0002mg/ml Ethidium Bromide (Sigma), then purified it using QIAquick Gel Extraction Kit (Qiagen) according to manufacturer protocol: We cut out the correct sized fragment from the gel using a scalpel, weighed it, then dissolved it in 3 volumes of Buffer QG, added 1 volume isopropanol, and ran it through the QIAquick column using a centrifuge. We then washed the column with 500µl of Buffer QG, then 750µl of Buffer PE, then eluted the DNA using 30µl of Buffer EB.

Preparing Vectors, Assembly and Transformation

The vector was digested with either NotI-HF (NEB) or EcoRI-HF (NEB) (see chapter methods for details on each plasmid assembly) by incubating plasmid DNA in a 50µl reaction with 5µl CutSmart Buffer (NEB) and 1µl of the enzyme at 37°C for an hour.
We then purified the digested vector using MinElute PCR purification kit following the manufacturer's protocol:

We added 5 volumes of Buffer PB to 1 volume of digested plasmid and ran it through a MinElute column using a centrifuge. We then washed with 750µl Buffer PE and eluted the DNA using 30µl Buffer EB.

We quantified the vector and insert samples, then combined them for each reaction (see chapter methods) at a vector:insert molarity ratio of 1:2. We then added our DNA combined sample to an equal volume of NEBuilder HiFi DNA Assembly Master Mix and incubated at 50 °C for 15 minutes to assemble the inserts into the vector. We then transformed the product into NEB® 5-alpha Chemically Competent E. coli: We thawed the cells thawed from -80°C on ice, then added 2µl of the chilled assembled product to the cells and mixed gently. We left the mixture on ice for 30 minutes, then heat shocked them at 42°C in a water bath for 30 seconds. We then placed the cells on ice for 2 minutes before adding 950µl of room temperature SOC media (NEB) and incubating the mixture at 37°C in a shaking incubator at 150 RPM. We then transferred them to selection plates.

Selection

All of the plasmids we used had a gene for Ampicillin resistance. We plated the transformed bacteria onto Lysogeny broth agar (LB-agar) plates with 100µg/ml Ampicillin for selection of transformants carrying the plasmid, and incubated overnight at 37°C. We picked off colonies and verified the presence of an insert a PCR reaction using the bacteria directly as a template, and running the product through a gel (primers in Appendices 17-18). We then further verified colonies which had the correct band size by sanger sequencing the insert region of the plasmids (primers in Appendices 17-18). Samples with the correct insert were then grown further for freezing and for plasmid extraction.

Plasmid preparation (mini, midi)

We grew *E. coli* overnight in a 150RPM shaking incubator at 37 °C in 15ml (for minipreps) or 200ml (for midipreps) of Lysogeny broth (LB) and 100µg/ml Ampicillin. We then extracted the

DNA using QIAprep Spin Miniprep Kit (Qiagen). For larger preparations (such as for injection into flies) we used the QIAGEN Plasmid Plus Midi Kit.

Cell Culture Care

We kept Drosophila cells in either 75cm^2 Tissue Culture (TC) flasks with canted necks and vented caps (Corning) or 25cm^2 flasks with canted necks and vented caps (Falcon) at 25°C using Schneider's Drosophila medium (ThermoFisher) with added Penicillin-Streptomycin (for a final concentration of 100U/ml) and 10% (v/v) Heat Inactivated non-USA origin sterile-filtered Fetal Bovine Serum (FBS) (Sigma). Cells were split 1:5 and passaged to new medium every 7-10 days or when approaching 100% confluency as ascertained by microscope. Cell lines were frozen using 1ml or 2ml freezing medium (FBS with 20% (v/v) Dimethyl sulfoxide (DMSO)) at -190°C for preservation.

Transfecting cells

We transfected cells with plasmids using Effectene Transfection Reagent (Qiagen). We placed 1.6 ml of medium containing cells (at ~150,000 cells/ml) in a clear Costar 6-well TC plate (Corning) and incubated the plate overnight, then replaced the medium with 1.6ml of fresh medium. We diluted ~400ng of our plasmid DNA in Buffer EC (Qiagen) to 100µl, then added 3.2µl of Enhancer (Qiagen) to it, vortexed for 1 second, then incubated the sample at room temperature for 2-5 minutes. We added Effectene Transfection Reagent (Qiagen) to the mixture, vortexed for 5 seconds, then incubated the mixture again for 5-10 minutes. We then added 600µl of medium to the mixture, mixed by pipetting, then added the mixture drop-wise to our cells in the 6-well plate.

All of the plasmids we transfected carry a puromycin resistance gene. 24 hours after transfection, we replaced the medium with 2.5ml of fresh medium and added puromycin (Puromycin dihydrochloride, Sigma) at a final concentration of 10µg/ml to select the lines that were successfully transfected. We verified the successful transfection by checking for expression of the

gene using qPCR (primer VentruncExpFR in Appendix 19) or (in lines expressing a V5 fluorescent marker) using flow cytometry (see below).

Infecting cells with DAV

We added CuSO4 diluted 1:1000 in the cell medium to a final concentration of 5mM to induce the expression of the pMT driven insert or, in the case of the control cells which lacked the insert, to maintain similar conditions between our test samples and our control. 24-48 hours later, we diluted the cells to ~1.5x10^5 cells/ml or ~1.5x10^6 cells/ml in medium containing 5mM $CuSO_4$, counting the cells under the microscope using Fast-Read 102 cell counting chambers (Biosigma). We placed 90µl of our diluted cells in a 96 well plate (Cellstar). We infected the cells with DAV diluted with medium containing 5mM $CuSO_4$ at either a 10^{-8} dilution (using 10µl of 10^{-7} dilution DAV in 90µl of cells which is ~3.16x $TCID_{50}$ or 2.1 viral particles/well) or a 2x10^{-8} dilution (using 10µl of 2x10^{-7} dilution DAV in 90µl of cells which is ~6.32 $TCID_{50}$ or 4.2 viral particles/well). 72 ± 3 hours later, cells were spun down, medium was removed, and cells were dissolved with 30µl nfH_2O, then frozen in 200µl TRIzol Reagent (ThermoFisher) at -80°C until RNA extraction, or dissolved in 50µl nfH_2O then frozen in 150µl TRIzol LS Reagent (ThermoFisher) at -80°C until RNA extraction.

Flow Cytometry

Some of the cell lines we created had V5 tagged protein products which we could stain and detect using flow cytometry. To stain our cells for flow cytometry analysis we used the following protocol. Transcription of our inserts (which all had an MT promoter) was induced by incubation with 5mM of CuSO4 in 1ml of medium for 24 hours in a Costar 6 well TC plate (Corning). 300µl of cells with at least a 10^7 cells/ml density were transferred to a clean tube and centrifuged for 3 minutes at 400g. We removed the supernatant, and resuspended in 4%(w/v) formaldehyde for 20 minutes at room temperature for fixing. We then washed twice with Phosphate-buffered saline (PBS), and resuspended in PBS with 1%(v/v) Triton-X and 1%(v/v) normal goat serum (NGS)

(Merck) for 30 minutes. We centrifuged cells for 3 minutes at 400g and removed supernatant, then incubated overnight with 0.1μg V5 tag monoclonal antibody (Invitrogen), diluted in 300μl of PBS with 1% NGS. We washed twice with PBS, then incubated overnight with Alexa Fluor 488 Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody (ThermoFisher) diluted 1:1000. We then washed twice with PBS then incubated for 45 minutes with 6μM blue fluorescent DAPI (SelectFX Nuclear Labeling Kit, Invitrogen). Finally we washed twice with PBS, then resuspended with 300μl of PBS.

We analysed the stained cells using Attune NxT analyser at the flow cytometry facility at the University of Cambridge Department of Pathology. We used the blue laser 1 channel for Alexa Fluor 488, and the violet laser 1 channel for DAPI as well as forward scatter channel and the side scatter channel. We then gated the SSC area vs FSC area plot to exclude debris, then further gated the FSC width vs FSC area plot to exclude doublets. We then plotted Alexa Fluor 488 vs DAPI for analysis.

ϕC31

To transform fly lines, we used the ϕC31 Integrase system which uses the ϕC31 Integrase to recombine an attachment site attB present on our plasmid with an attachment site attP present in the genome of the fly line to be transformed (M. C. M. Smith et al., 2010)). The integrase can either be injected on a plasmid along with the plasmid carrying attB, or it can be expressed transgenically in the line which is injected with the plasmid.

Chapter 1 - *Doc* Element Insertion in *Veneno* Causes Resistance to DAV

Introduction

In this chapter, we will discuss a polymorphism in *Drosophila melanogaster* which causes resistance to Drosophila A Virus (DAV). As we've discussed above, DAV is a positive sense RNA virus which naturally infects wild *Drosophila melanogaster* flies. It would not be surprising, therefore, if flies have evolved mechanisms to resist DAV infection. The evolution of such mechanisms is important to study, since it gives us insight into the constant evolutionary arms race occurring between virus and host, and, in particular, how the hosts adapt to selective pressures created by viral infection in the wild.

We find that the polymorphism is due to a *Doc* retrotransposon insertion in the coding sequence of a gene *Veneno* (*CG9684*) of unknown function, predicted to have a role in the piRNA pathway. Transposable elements (which we discuss in the Introduction section of this thesis) typically have detrimental effects on the fitness of the host especially when they occur within coding sequences, since they can lead to harmful mutations. It is therefore very interesting to study the link between them and the evolution of virus resistance, an adaptive trait. We find that the *Doc* insertion results in a truncation in the *Veneno* transcript, leading to a novel restriction factor which leads to the gain of the antiviral function.

Results

A *Doc* element insertion in *Veneno* is associated with resistance to DAV.

We infected lines from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012) with Drosophila A Virus (DAV) and found a large variation in resistance to the virus, with two of the DGRP lines having a ~19000-fold lower DAV titre than the other lines (Figure 1.1). We determined that the DAV resistance which these lines have is a dominant phenotype by crossing the resistant lines to susceptible lines and infecting their heterozygous offspring with DAV, to find that the heterozygotes had a low viral titre 3 days post infection, similar to that of resistant flies (Figure 1.2A).

We mapped the resistance to chromosome 3 by combining 1 chromosome from a resistant line with 2 chromosomes from a susceptible line and infecting the resultant flies. Only flies with chromosome 3 from the resistant line were also resistant (Figure 1.2B).

To perform quantitative trait locus (QTL) mapping of DAV resistance, we infected flies with a recombined third chromosome containing sequences from both resistant and susceptible lines. We found a QTL peak between 47 and 49 cM (95% CI) in chromosome 3 of resistant line 239 and between 46 and 50 cM (95% CI) in chromosome 3 of resistant line 362 (Figure 1.2C-D). We then used fine-scale recombinant mapping (Figure 1.3) to narrow down the polymorphism responsible for the increased resistance first to a region on chromosome 3 between 3.75 Mb and 4.26 Mb (Figure 1.3B-C), then further down to a chromosome 3 region between 4.16 Mb and 4.19 Mb (Figure 1.3D-E).

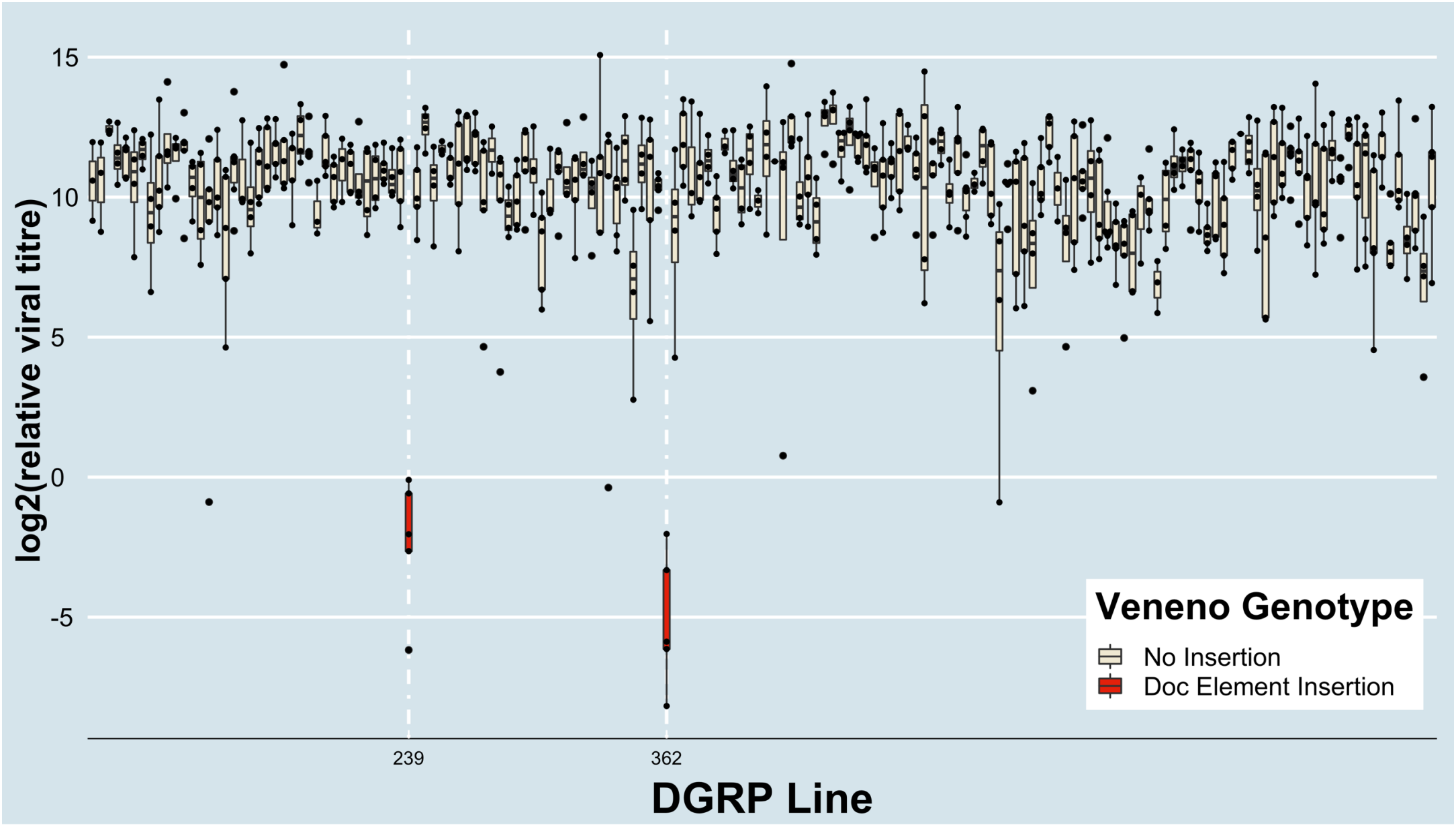


Figure 1.1 *Doc* element insertion correlated with lower DAV titre. Boxplot showing DAV titre of DGRP lines 3 days post infection: DGRP lines were infected with DAV. After 3 days, viral titre was measured relative to EF1 using qPCR. Each dot represents 15 flies, boxes represent IQR, whiskers 1.5*IQR. Lines with *Doc* element in *Veneno* (Ven^{Doc}) are highlighted in red. The association between resistant lines and the *Doc* element insertion is highly significant (Fisher's exact test : $p = 7.7*10^{-5}$.

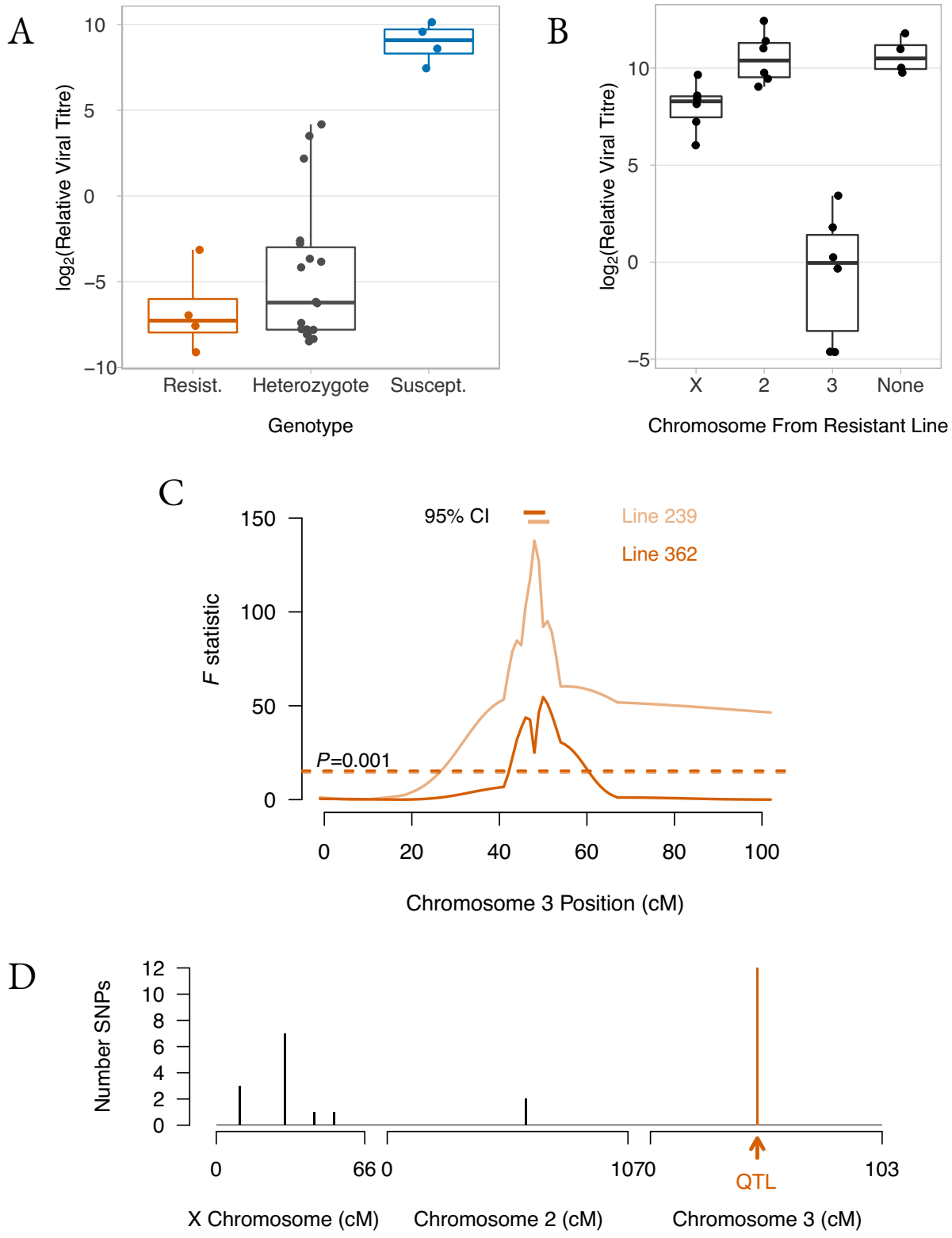


Figure 1.2. Genetic mapping of DAV resistance. A) Dominance was estimated by measuring viral titre in resistant flies (DGRP lines 239 and 362), susceptible flies (DGRP lines 48 and 91) and F1 progeny of crosses between these lines. The pairs of lines were combined for plotting as they did not differ significantly in viral titre. Each point is a single RNA extraction from 20 female flies. B) Resistance was mapped to chromosome by combining a single resistant chromosome (DGRP lines 239 and 362) with two chromosomes from susceptible flies (DGRP lines 48 and 91). Each point is a single RNA extraction. The crosses involving the two DGRP lines are combined as they yielded similar results. C) Resistance QTL were mapped by crossing DGRP lines 239 x 373 and 362 x 306, to generate 84 and 72 recombinant lines respectively. DAV titre was measured in a single female from each line. Ten molecular markers were genotyped across Chromosome 3, and genotypes between these markers inferred using the Hayley-Knott method. The Y axis shows an *F* statistic which measures the strength of association between genotype and viral titre. The dashed line is a significance threshold from a null distribution of the *F* statistic that was obtained by permuting the viral titre estimates across the lines 1000 times and recording the maximum value of *F* across the chromosome. The horizontal bars are 95% confidence intervals on the location of the QTL peak obtained by resampling the lines with replacement 1000 times, recording the location of the maximum *F* statistic, and plotting the central 95 percentiles. In panels A-C flies were infected by intrathoracic inoculation. Viral titre was estimated 3 dpi using quantitative PCR to measure the concentration of viral RNA relative to mRNA from the *Drosophila* gene *EF1α100E*. Boxes show median and interquartile range. D) The genomic location and number of SNPs where the resistant lines 239 and 362 had the same allele and all the susceptible lines had a different allele.

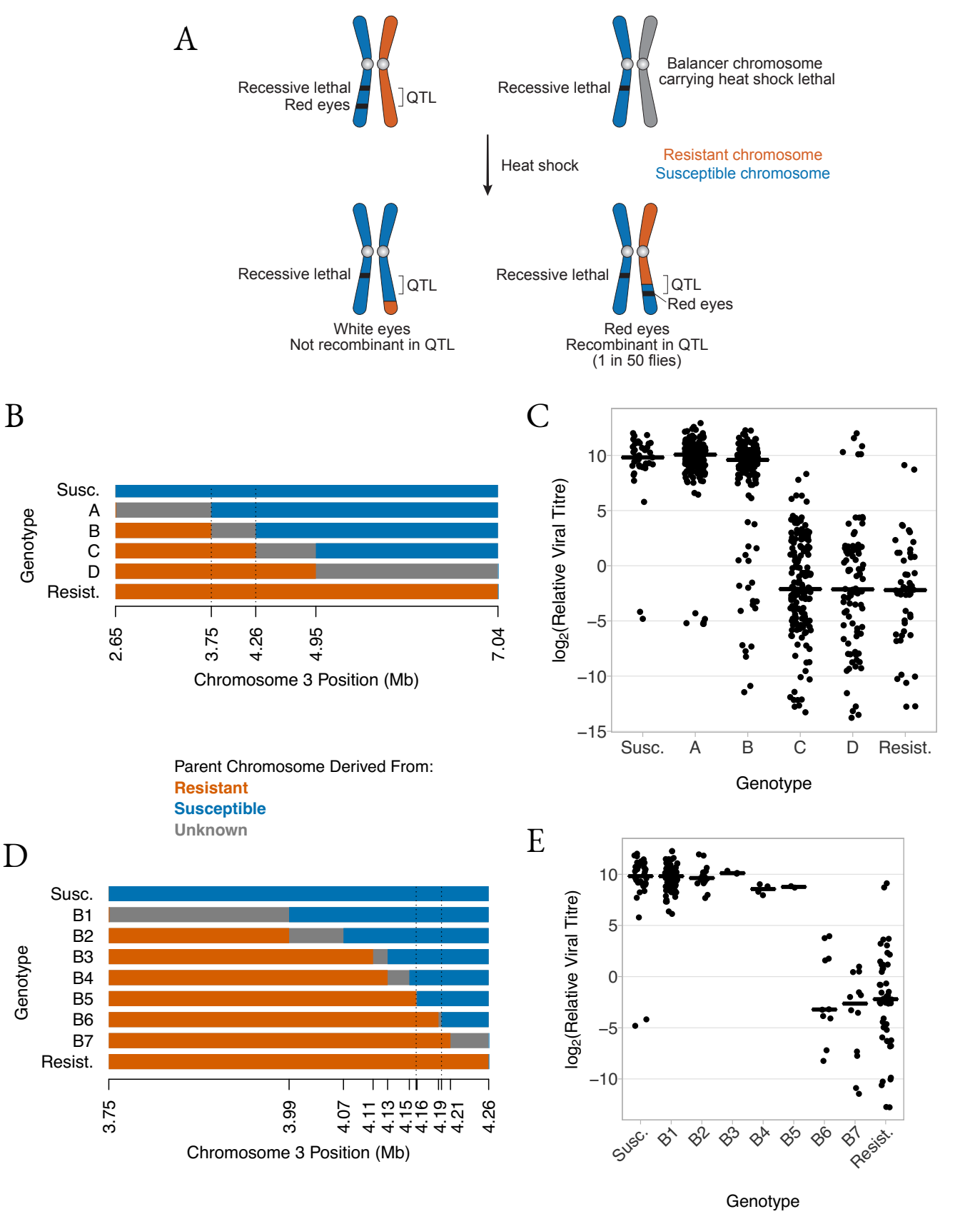
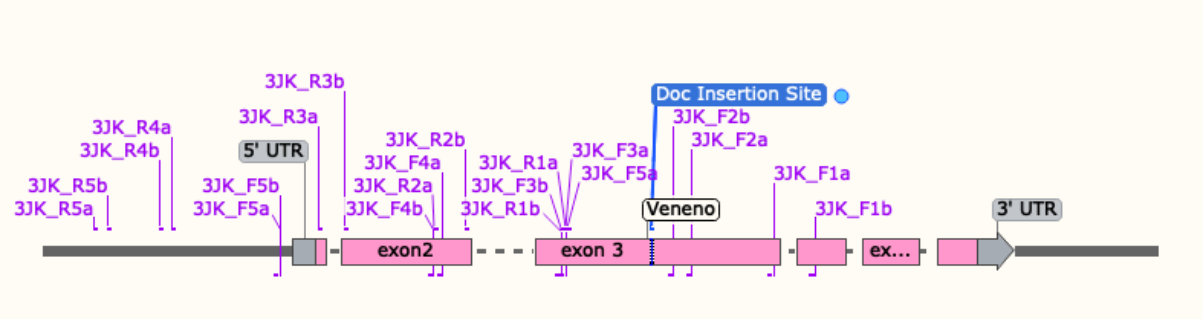


Figure 1.3. High resolution genetic mapping of DAV resistance. A) Schematic of the genetic cross used to select recombinants within the QTL controlling susceptibility to DAV. (B) The recombinant genotypes generated in the first phase of mapping. The X axis tick marks show the location of five molecular markers used to genotype the recombinants. The dashed lines mark the region containing the gene controlling susceptibility inferred from the data in Panel C. (C) The viral titre of the six genotypes in Panel B. Each point is an independent recombinant line. In most cases this represents a single RNA extraction from 5-20 flies (mean, 16.6 flies). (D) Recombinants within the region between the dotted lines in Panel B. The X axis tick marks show the location of 12 molecular markers used to genotype the recombinants. The dashed lines mark the region containing the gene controlling susceptibility inferred from the data in Panel E. (E) The viral titre of the genotypes shown in Panel D. Viral loads of recombinant lines that define the location of the resistant allele were measured in at least 4 biological replicates (RNA extractions from 5-20 flies); the points are the mean of these replicates. In panels C and E flies were infected by intrathoracic inoculation. Viral titre was estimated 3 dpi using quantitative PCR to measure the concentration of viral RNA relative to mRNA from the *Drosophila* gene *EF1α100E*. Note the titre measurements in panel E are also shown in panel C.

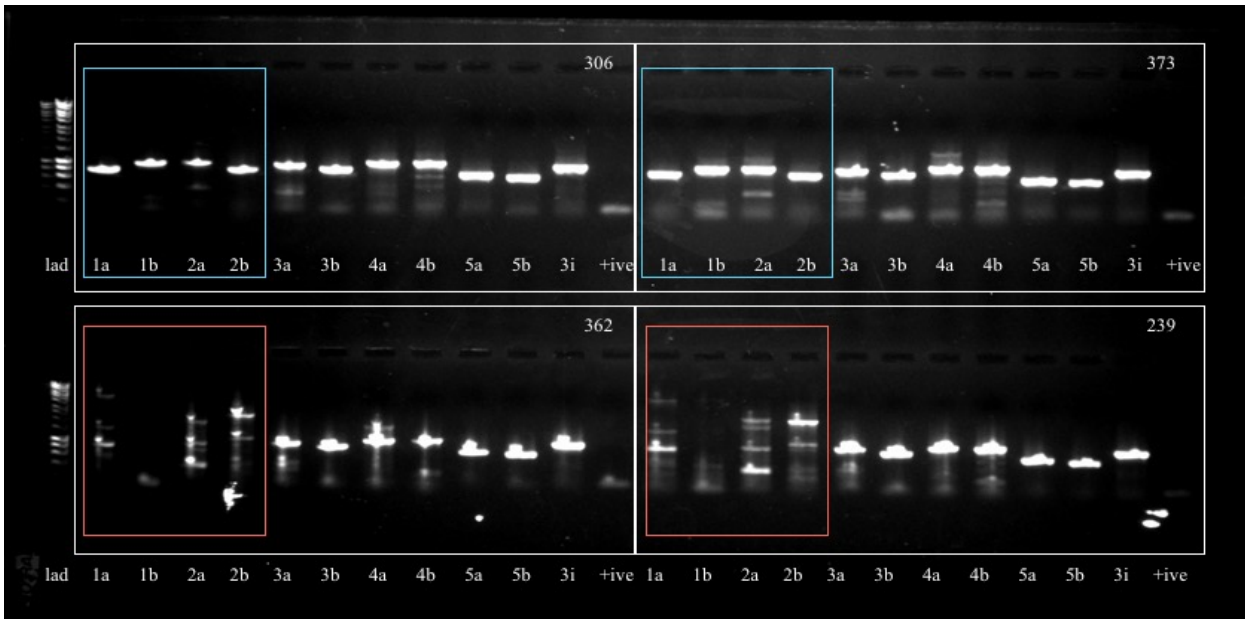
The region we mapped resistance to contains 12 genes. Using RNAi, we knocked down some of these genes and found that knocking down gene *CG9684* resulted in a loss of the resistant phenotype (Figure 1.6B). The orthologue of this gene in *Aedes aegypti* called *Veneno* has been shown to be involved in the antiviral piRNA pathway (Joosten et al., 2018). We will refer to *CG9684* as *Veneno (Ven).*

Using polymerase chain reactions, we amplified different regions of *Ven* in susceptible and resistant lines (Figure 1.4A, primers in Appendix 1), and ran the product through agarose gel to check if the amplified DNA was of an expected size given the reference sequence (dmel_r6.28_FB2019_3; Larkin et al., 2021). Regions corresponding to primer pairs 3JK_1a, 3JK_1b, 3JK_2a, and 3JK_2b (which all amplify regions in exon 3) showed a normal product in the susceptible lines, but had a larger product with multiple bands in the resistant lines (Figure 1.4B). We hypothesized that this is due to a structural variant of *Veneno* in the resistant lines.

For the products formed by amplification of the resistant lines with primers 3JK_1a, 3JK_1b, 3JK_2a, and 3JK_2b, we cleaned the reaction using ExoSap (see main methods) then Sanger sequenced (see main methods) each reaction with the forward and reverse primers used to amplify that product. This sequencing showed evidence of a transposable element insertion: Sequences were run through BLAST (ALTSCHUL et al., 1990) and returned matches to the *Doc* element, a non-LTR retrotransposon. We hypothesized that the polymorphism in the resistant lines is a *Doc* element insertion in *Veneno* (Figure 1.5A). To confirm this, we designed primer pairs spanning the gene-TE junction assuming that the *Doc* element is indeed inserted at the position hypothesized (3' Junction and 5' Junction Primers in Appendix 2). Using these primers, we were able to amplify sequences in the resistant lines, and not the susceptible lines, which is consistent with our *Doc* element insertion hypothesis. We also designed primers spanning the point in the sequence hypothesized to contain the insertion (see Appendix 2: Spanning Primers), and were able to amplify the entire *Doc* element sequence, which we then sequenced, confirming indeed the presence of a 4685bp *Doc* element insertion in the coding sequence of exon 3 of *Veneno* (GenBank MZ047782).

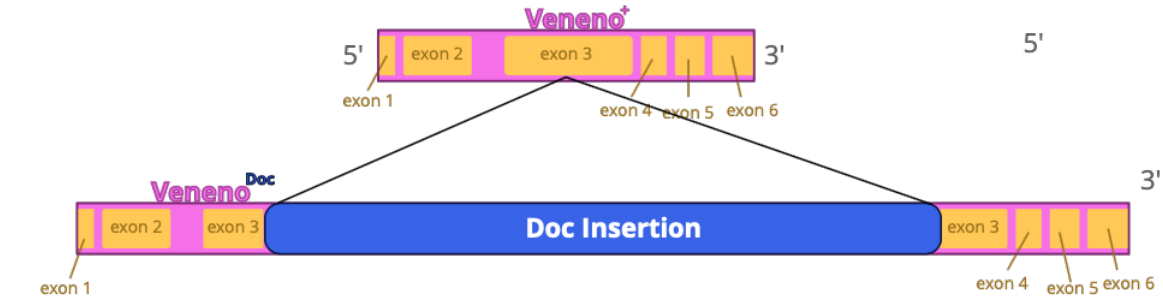


A



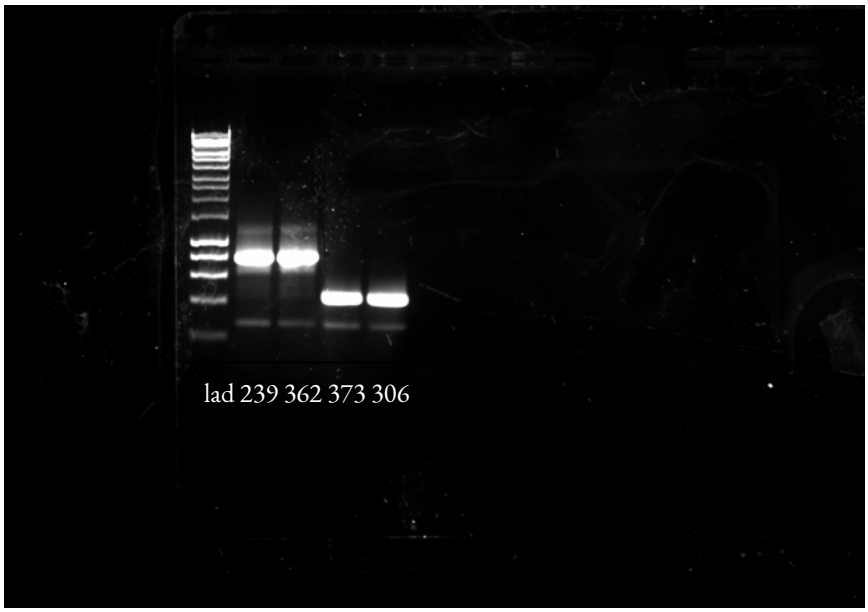
B

Figure 1.4. Resistant alleles of *Veneno* have structural variant.
A) Map of gene *Veneno* showing primer locations and *Doc* element insertion site
B) Gel image comparing PCR amplification of the *Veneno* gene in two susceptible lines 306 and 373 (top), and in two resistant lines 362 and 239 (bottom). The ladder used is Bioline 1kb ladder. The PCR reactions yielded similar bands in the 4 lines except for the 4 highlighted lanes. The susceptible (blue) have the expected sized bands, whereas the resistant (orange), show multiple bands including larger than expected bands in the 1a, 2a and 2b lanes, and no amplification in the 1b lane.



A

B



C

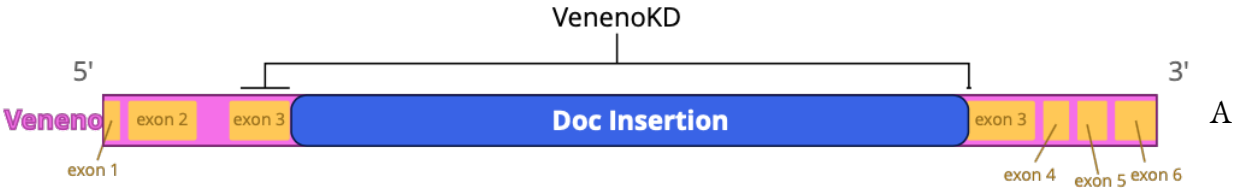
Figure 1.5. A) *Ven*^{Doc} allele of *Veneno* contains a *Doc* element insertion. Map shows alleles *Ven*$^{+}$ and *Ven*^{Doc}, the latter containing a *Doc* Insertion (blue) occurring in exon 3.
B) Veneno Genotyping We used a primer pair targeting sequences flanking the *Doc* element insertion (CGthenDocF5 and DocThenCGR2) with an expected product size ~390 bp (no insertion present) and a third primer targeting a sequence within the *Doc* element insertion (DocThenCGF2) that along with (DocThenCGR2) has an expected product size of ~730bp (insertion present) (See Appendix 2 for primers).
C) Gel image comparing PCR amplification products in the *Veneno* gene in lines 239 (resistant), 362(resistant), 373 (susceptible) and 306 (susceptible). The ladder used is Bioline 1kb ladder. The gel result is consistent with the insertion being present in 362 and 239, but not 306 and 373.

We hypothesized that the *Doc* element insertion is driving the increase in resistance. To test the association, we checked for the presence of the *Doc* element insertion in 162 DGRP lines. We used two primers on either side of the *Doc* element insertion, and one within, to create a three primer mix (Primers DocThenCGR2, DocThenCGF2, CGThenDocF5, Appendix 2, Figure 1.5B) which we used to genotype the DGRP lines as either having the allele of *Ven* with the *Doc* element (Ven^{Doc}) or the wild type allele without the *Doc* element (Ven^{+}) (Figure 1.5C). Using this method, we were able to confirm that the *Doc* element insertion correlates perfectly with the increase in resistance: the two lines having Ven^{Doc} had on average ~19000 lower DAV titre than flies with Ven^{+} (Figure 1.1; Fisher's Exact Test: $p=7.7*10^{-5}$).

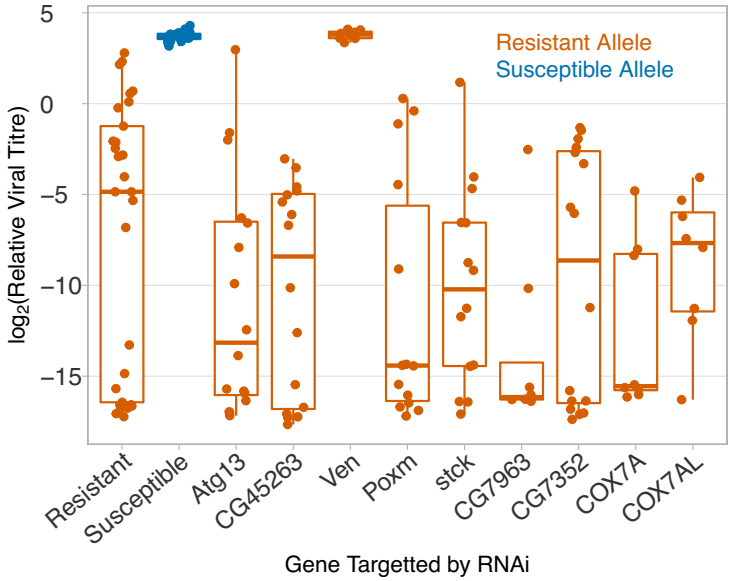
Transcription of *Veneno* is necessary for resistance phenotype

Given our results showing that knocking down *Veneno* results in a loss of the resistant phenotype (Figure 1.6B), we hypothesized that transcription is necessary for that phenotype. To confirm this, we repeated the knock down on the *Veneno* transcript. We used VDRC fly line 24094 carrying the UAS driven RNA hairpin construct (Fig 1.6A, UAS-VenenoKD) (Dietzl et al., 2007; provided kindly by the Vienna Drosophila Resource Center) on the second chromosome, which we combined with Ven^{Doc} on the third chromosome and crossed with a second chromosome daughterless::Gal4 line, resulting in a UAS- VenenoKD /dagal4 heterozygote ubiquitously expressing the hairpin construct (See crossing scheme Figure 1.14).

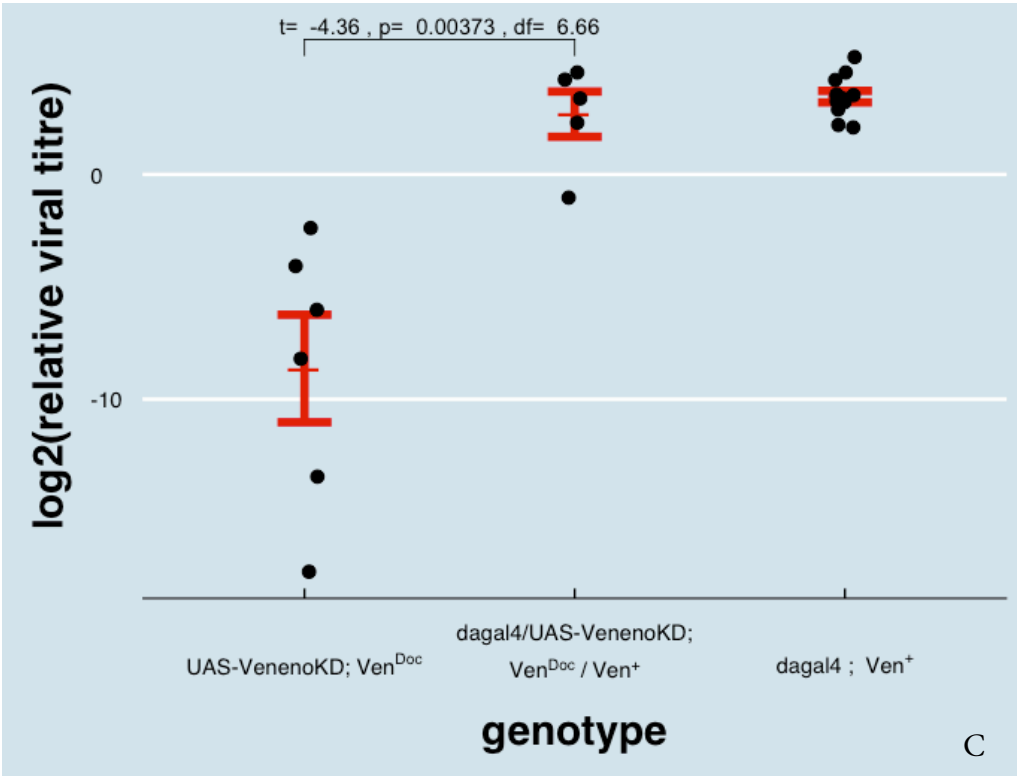
The region targeted by the hairpin construct overlaps with the insertion site, so it targeted potential transcripts both upstream and downstream of the insertion site (figure 4A). We found that the knockdown resulted in increased viral titre in the line heterozygous for Ven^{Doc}/Ven^{+} which expresses both UAS-VenenoKD and dagal4 compared to the Ven^{Doc} line which only expresses UAS-VenenoKD (Figure 1.6C, Welch's t-test, t=-4.36, p= 0.00373, df=6.66). The resultant viral titres were similar to those in the susceptible control (expressing dagal4 and Ven^{+}), indicating that the resistant phenotype is completely lost by knocking down *Veneno.* This confirms that the transcription of *Veneno* is necessary for its role in causing resistance.



A



B



C

Figure 1.6. Knocking down Veneno results in loss of resistant phenotype. A) *Veneno* Knockdown site is on either side of the *Doc* element insertion.

B) Genes in the region to which DAV resistance was mapped were knocked down using RNAi then infected with DAV. Y-axis shows $\log_2$ of viral titre relative to *EF1α100E*. Boxes show median and interquartile range. Knockdown of *Ven* resulted in an increase in viral titre to the level of the susceptible control whereas knockdown of the other genes did not. Resistant line used is DGRP_362.

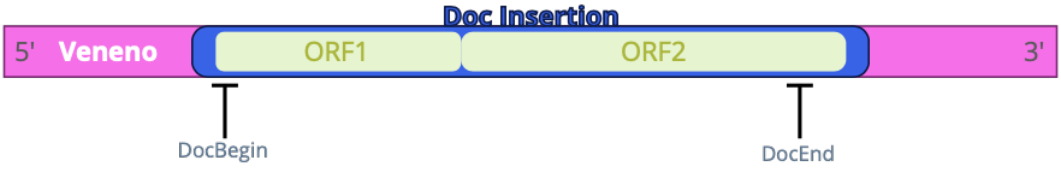
C) Flies with knocked down Veneno in a resistant background (dagal4/UAS-VenenoKD; Ven^{Doc}/Ven^{+}) as well as resistant controls (UAS-VenenoKD ; Ven^{Doc}) and susceptible controls (dagal4; Ven^{+}) were infected with DAV. DAV titre relative to *Actin5C* of knockdown and control lines was measured 3 days post infection with qPCR. Each dot represents 6-15 infected flies. y-axis shows the -ΔCT values, which is the log of the relative viral titre. Red coloured error bars show standard error centred around the mean. P-values calculated using Welch's t-test. Ven^{Doc} line is DGRP_362

No evidence that the transcription of the *Doc* Element is involved in DAV resistance

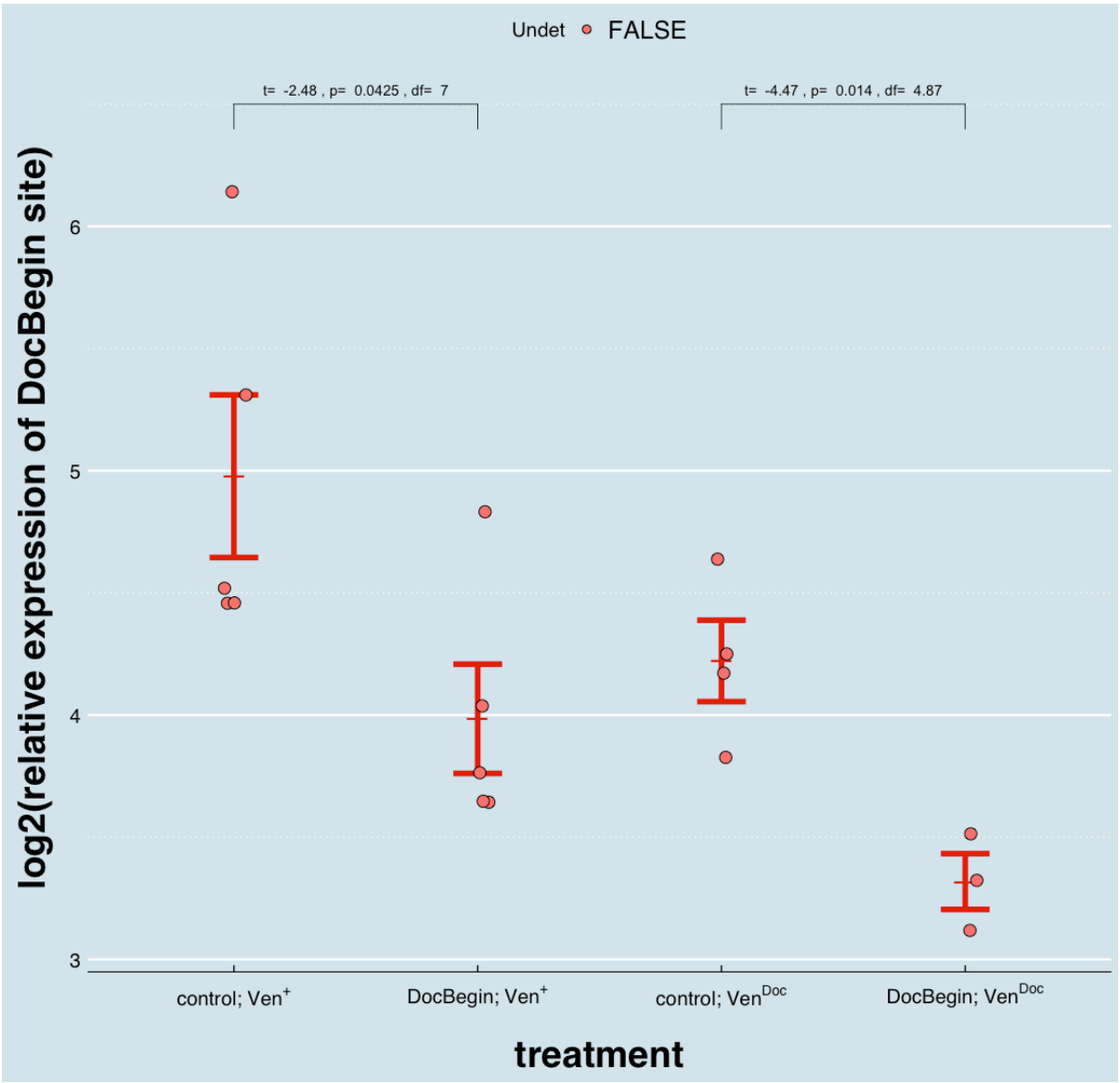
The presence of the *Doc* element insertion in the DAV resistant allele of *Veneno* was particularly interesting because a *Doc* element insertion found in gene *CHKov1* was found to cause resistance to DMelSV (See Chapter 3). We speculated that the *Doc* element itself might have a viral resistance function which is being activated by these insertions. Since resistance to DAV relies on transcribing *Veneno,* we hypothesized that the *Doc* element insertion in Ven^{Doc} might itself be transcribed and thereby cause the resistant phenotype.

To test whether transcription of the *Doc* element itself is necessary for the resistant phenotype, we performed in vivo RNAi to knockdown two regions in the *Doc* element. The two regions, DocBegin and DocEnd (see Figure 1.7A and appendix 3), are respectively in open reading frames 1 and 2 of the *Doc* element.

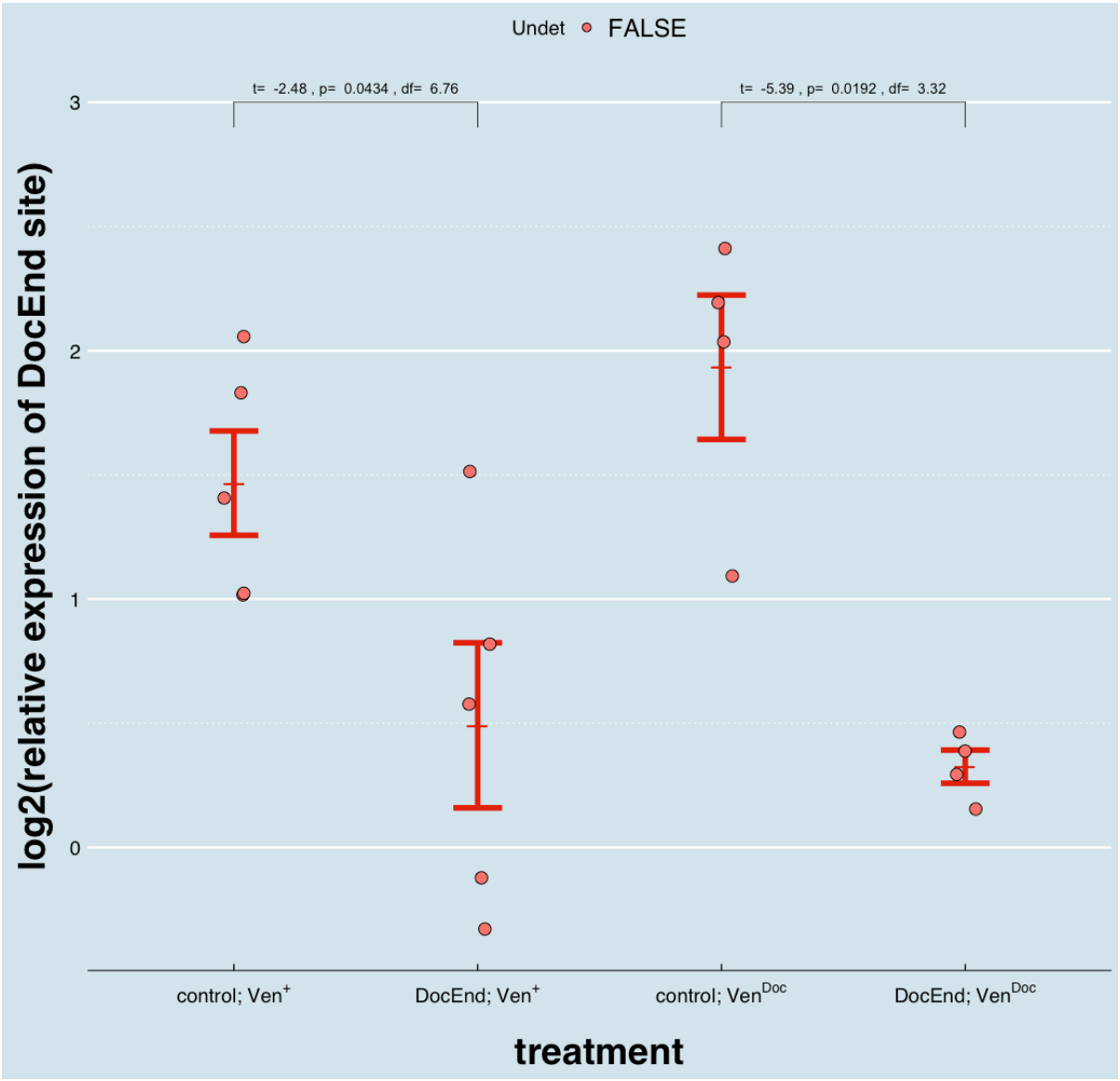
We made use of the transgenic RNAi project (TRiP) (Zirin et al., 2020) which has created plasmids for the generation of RNAi flies. We generated TRiP plasmids by inserting the UAS driven TRiP constructs into the WALIUM 20 vector, then transformed the plasmids into the second chromosome of a fly line. The UAS-TRiP target was crossed into either resistant or susceptible background (on the third chromosome) and then further crossed at 29C with lines of the same background carrying ubiquitously expressed gal4 drivers(see crossing scheme in figure 1.15). The gal4 activates the UAS resulting in the expression of the TRiP hairpin RNA, which are then processed by endogenous siRNA machinery, and use as templates for the breakdown of the target RNA. To verify the effect of the knockdown, we checked the expression of the target sites indeed dropped (see figure 1.7B). In all cases, there were significant decreases in expression between control and knockdown lines ($p<0.05$, Welch's t-test, FDR adjustment, n=2), indicating that the knockdown was effective.


Doc Insertion
5'
Veneno
ORF1
ORF2
3'
DocBegin
DocEnd

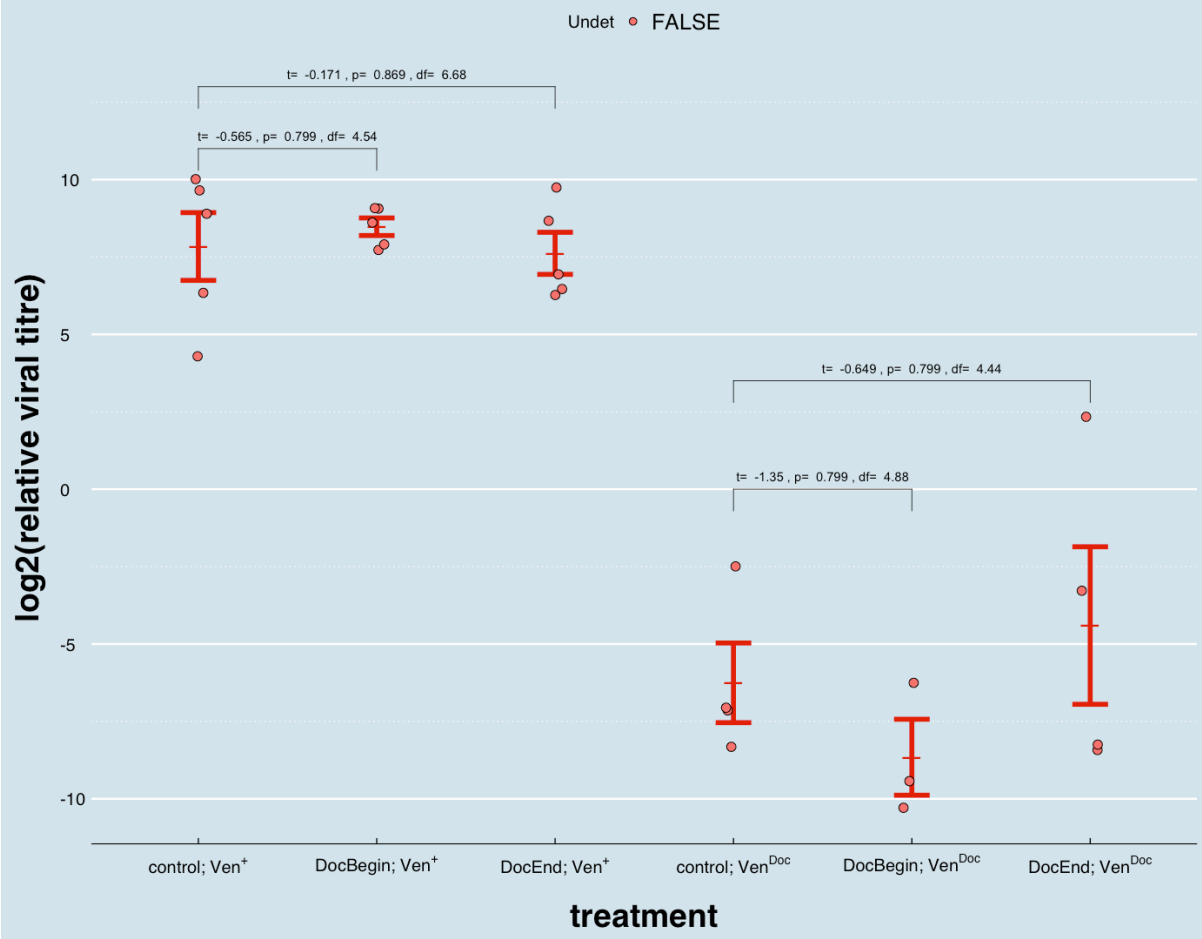
A


Undet
FALSE
t= -2.48 , p= 0.0425 , df= 7
t= -4.47 , p= 0.014 , df= 4.87
log2(relative expression of DocBegin site)
6
5
4
3
control; Ven+
DocBegin; Ven+
control; VenDoc
DocBegin; VenDoc
treatment

B i


Undet FALSE
t= -2.48 , p= 0.0434 , df= 6.76
t= -5.39 , p= 0.0192 , df= 3.32
log2(relative expression of DocEnd site)
3
2
1
0
control; Ven+
DocEnd; Ven+
control; VenDoc
DocEnd; VenDoc
treatment

B ii



C

Figure 1.7. A) Knocking down *Doc* element sites using RNAi. We selected two sites, DocBegin and DocEnd in Open Reading Frames 1 and 2 respectively of the *Doc* element insertion in Ven^{Doc} to target with RNAi.

B) Expression of sites in *Doc* element targeted by RNA decreased in the RNAi flies compared to the control flies. Flies with RNAi constructs targeting DocBegin (i) and DocEnd (ii), as well as control flies with no knockdown targets, all having either Ven^{+} or Ven^{Doc} were kept (for their entire lifecycle) at 29 degrees and infected with DAV. We measured *Doc* element expression at the target sites. Each dot represents 6-15 flies. Y-axis shows qPCR expression relative to Rpl32. Red bars represent standard error of -ΔCT, centred around the sample mean. P-values calculated using Welch's t-test, and adjusted for multiple tests using false discovery rate (FDR) method.

C) No significant effect of *Doc* element knockdown on DAV resistance. Flies with RNAi targeting DocBegin and DocEnd, as well as control flies with no RNAi targets, all having either Ven^{+} or Ven^{Doc} were infected with DAV and kept at 29 degrees. Y-axis shows viral titre relative to *Rpl32* 3 days post infection. Each dot represents 6-15 flies. Red bars represent 95% standard error of -ΔCT, centred around the sample mean. Regardless of knockdown target, viral titre was determined entirely by *Doc* element status in *Veneno*. P-values calculated using Welch's t-test, and adjusted for multiple tests using false discovery rate (FDR) method.

Subsequently, knockdown and control flies were infected with DAV, and viral titre was measured three days later. Viral titres of the knockdown flies were not significantly different from those of the controls (see figure 1.7C). Therefore, we find no evidence that knocking down either site in the *Doc* element causes any changes in resistance, or that *Doc* element transcription is involved in the resistant phenotype.

To verify this result and test whether the sequence of the *Doc* element itself is necessary for the phenotype, we used CRISPR/Cas9 to target the sequence for mutations at two different sites, DocRT and DocRNAGene, which fall within the two ORFs in the *Doc* element (Figure 1.8). Flies expressing cas9 ubiquitously in both resistant and susceptible backgrounds were crossed with flies expressing guide RNA molecules ubiquitously in the respective background. This should result in the F1 generation having both cas9 and gRNA expressed in all of its cells, leading to mutations occurring in every cell with intact *Doc* elements. However, the cross did not result in any viable mutant offspring, possibly since the *Doc* element is found in many copies in the genome, which possibly led to many double stranded breaks resulting in non-viable flies.

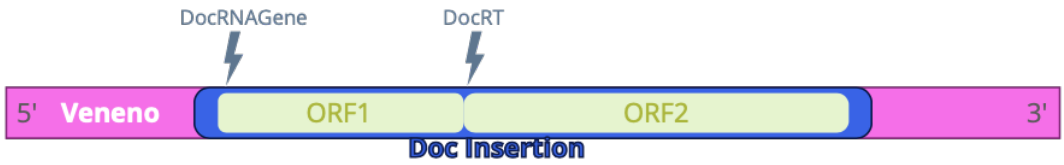


Figure 1.8. Knocking out *Doc* element using CRISPR/Cas9. We selected two sites, DocRNAGene and DocRT in Open Reading Frames 1 and 2 respectively of the *Doc* element insertion in Ven^{Doc} to target with CRISPR/Cas9 for mutations.

Doc element insertion leads to Truncated transcript of *Veneno*

The data presented so far seems to suggest the possibility that the region of Ven^{Doc} of non-TE origin is involved in DAV resistance, but that the *Doc* sequence itself may not be involved. One way this could occur is if the *Doc* element insertion results in a truncated transcript of *Veneno* which, due to the truncation, gains an antiviral function. To test whether the transcript of Ven^{Doc} is indeed a truncated *Veneno*, we tried to find the sequence of the transcript. Since the *Doc* element is found in multiple locations throughout the genome, it is not possible to find the transcript sequence by assembling short reads (since the short reads mapping to the *Doc* element would not be unique to Ven^{Doc}). We therefore opted to use Nanopore sequencing, which outputs long reads that we can uniquely map to *Ven*.

We extracted RNA from flies carrying the resistant allele of *Veneno* and the resistant allele of *CHKov1* using Trizol-chloroform extraction then purified it by binding it to columns and treating it with DNAse. We then performed both direct RNA and PCR-cDNA nanopore sequencing on the entire transcriptome of DGRP line 362 which carries the Ven^{Doc} allele.

Whereas the direct RNA sequencing yielded 1.9 million reads, the PCR-cDNA had a much higher throughput of 8.8 million mapped reads (SRA: SRR15541957). Furthermore, direct RNA sequencing provided very few sequences mapping to *Ven.* We therefore performed our analysis on the PCR-cDNA dataset. Using minimap2, we mapped the sequences to the *Drosophila melanogaster* genome which we modified to include the *Doc* element insertion in *Ven*. We then used the pinfish pipeline to find likely *Veneno* transcripts from the data. (Figure 1.9)

We identified three transcripts: transcript a (4 reads), transcript b (39 reads), and transcript c (4 reads) (Figure 1.9). Of these three transcripts, only Transcript a contained regions which were targeted by the *Veneno* RNAi knockdown, which had successfully eliminated the resistant phenotype, and had therefore targeted regions contained in whichever transcript was involved in resistance. Furthermore, transcript a was the only one which was consistent in its intron regions with the reference sequence data for *Veneno* (dmel_r6.28_FB2019_3; Larkin et al., 2021).

Transcript a (to be referred to as *Ven_tra*) therefore stood out as the likely transcript involved in resistance. *Ven_tra* contains the *Veneno* transcribed sequence upstream of the *Doc* element, and the first 137bp of the *Doc* element including a stop codon 18 base pairs downstream from the site of insertion (Sequences can be found on benchling: https://benchling.com/obrosh/f_/Zoz9jidS-thesis-sequences/). *Ven_tra* is predicted to encode an MYND Zn Finger domain and a Tudor Domain.

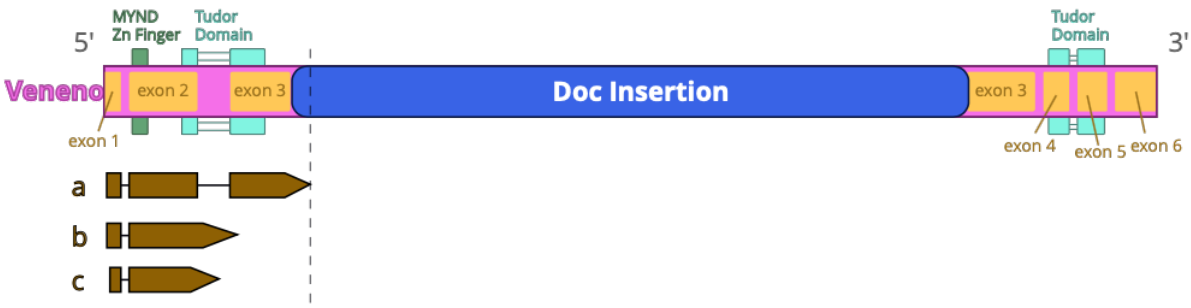


Figure 1.9. Nanopore sequencing shows three transcripts mapping to *Veneno*. Using pinfish we were able to narrow down the *Ven*^{Doc} transcript to three possibilities (a, b, and c). Of the three, transcript a (*Ven_tra*) seems to be the most likely candidate. It contains exons 1 and 2, the region of exon 3 upstream of the *Doc* element, and 137bp of the *Doc* element sequence, and it encodes an MYND Zn Finger domain and a Tudor Domain. Sequences can be found on benchling https://benchling.com/obrosh/f_/Zoz9jidS-thesis-sequences/ .

Veneno Sequence upstream of Insertion Necessary for Resistance

We hypothesized that the sequence of *Veneno* upstream of the *Doc* element, which is expressed by *Ven_tra*, is involved in the resistant phenotype. To test this, we knocked out two different sites in the DNA sequence of *Veneno*, both upstream (updoc) and downstream (downdoc) of the *Doc* element insertion (1.10A). Using CRISPR/Cas9, we targeted these regions for mutation (see target sites and oligos in appendix 4). We generated flies that express cas9 carrying either *Ven*$^{+}$ or *Ven*^{Doc}. We also generated flies which ubiquitously express guide RNA molecules and carry either *Ven*$^{+}$ or *Ven*^{Doc}. We then crossed each line carrying cas9 with the line carrying the gRNA and having the same *Veneno* allele. This results in the F1 generation having both cas9 and gRNA expressed in all of its cells, leading to somatic mutations caused by imperfect non-homologous end joining occurring in cells with intact *Veneno*. These flies will likely have mutations occurring in their early cells which then divide and likely produce a mosaic organism carrying different mutations. We infected the F1 generation flies with DAV, and tested for viral titre three days later.

To test whether the knockout was effective, we checked for indels in the knockout flies. We extracted DNA from pools of 10 F1 females, amplified PCR products that included the cas9 cutting site and used Ilumina MiSeq to sequence the sites targeted by the guide RNA. We then analysed the sequencing data using Cas-Analyzer to determine the percentage of sequences having

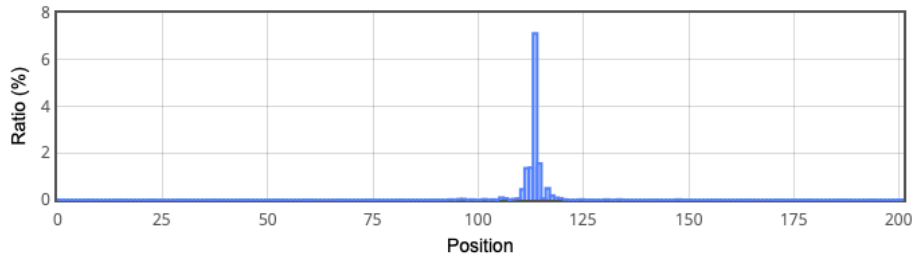
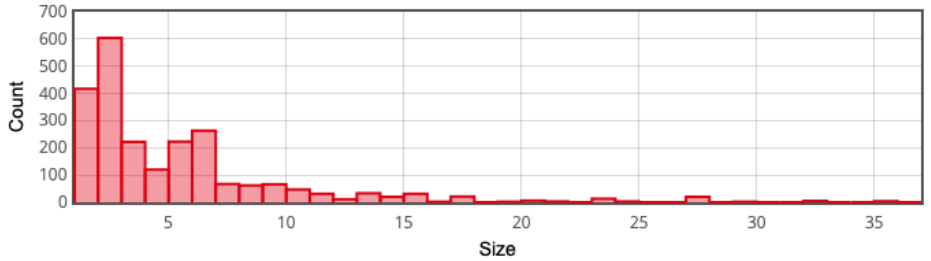
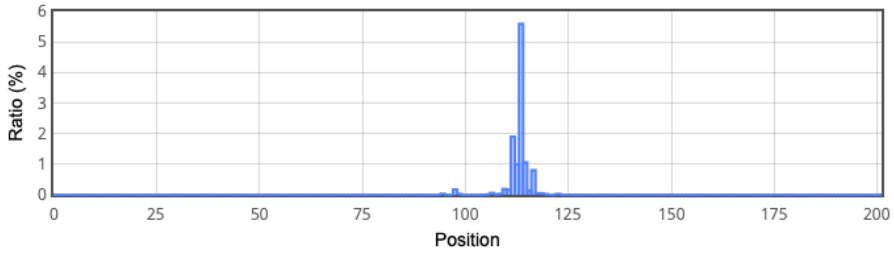
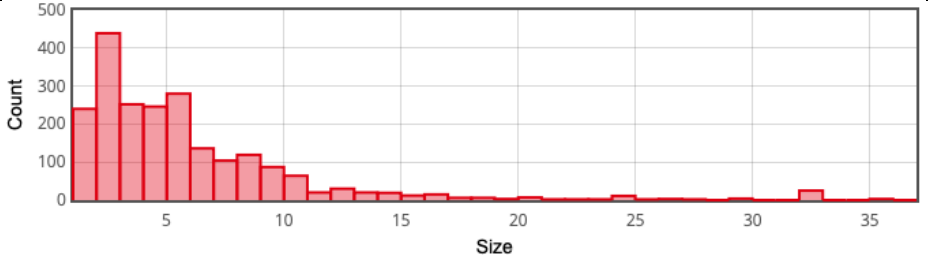
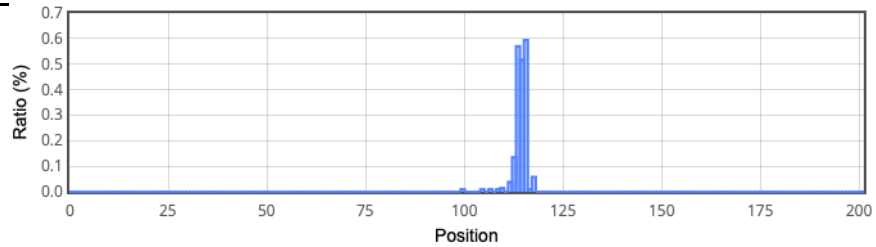
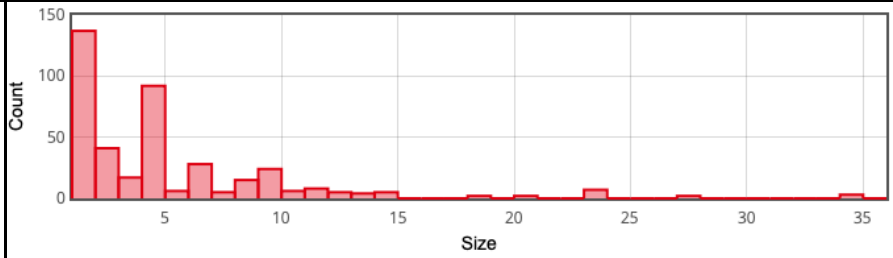
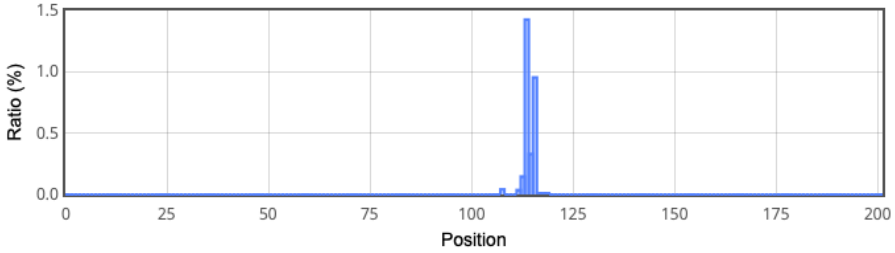
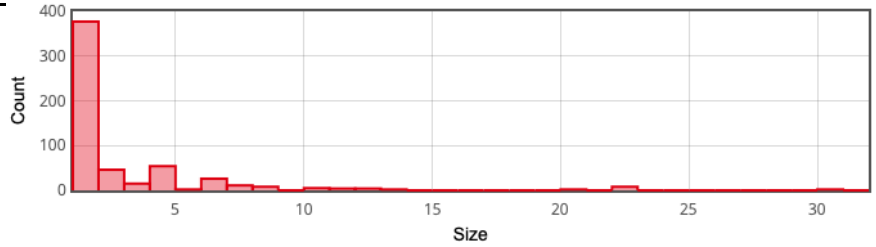
indels. We also calculated the percent of frameshifts by comparing the lengths of our sequences to the reference using Cas-Analyzer outputs. (Figure 1.10B). Interestingly, we found that the indels follow a repeatable pattern with, insertion positions, as well as deletion positions/sizes being highly similar for each CRISPR target.

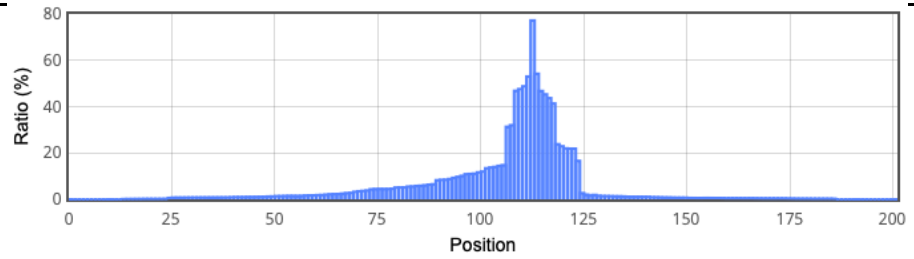
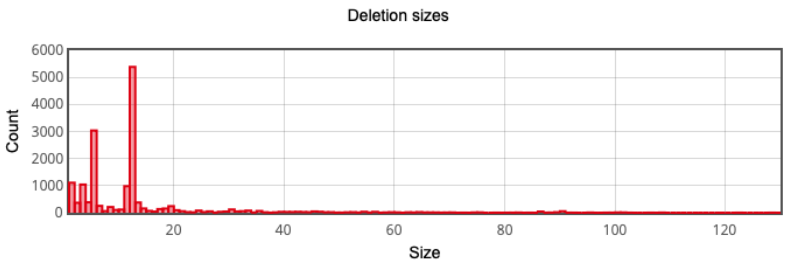
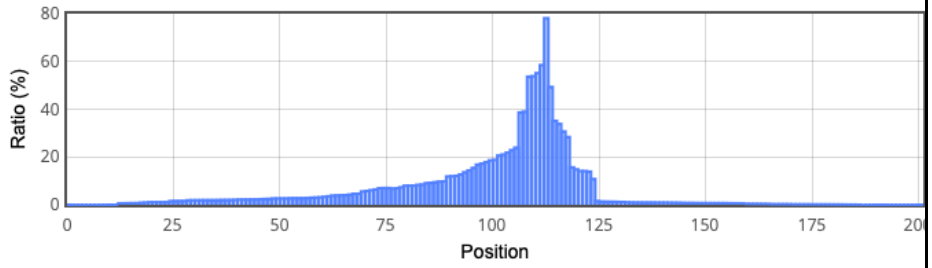
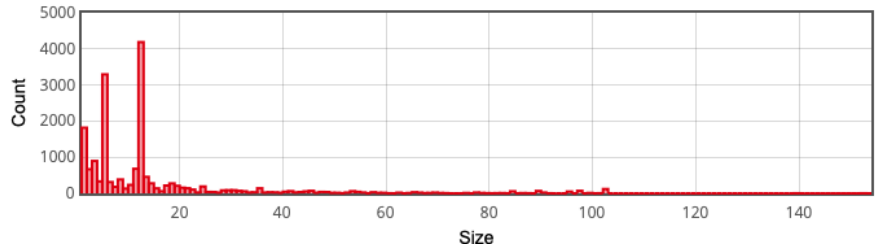
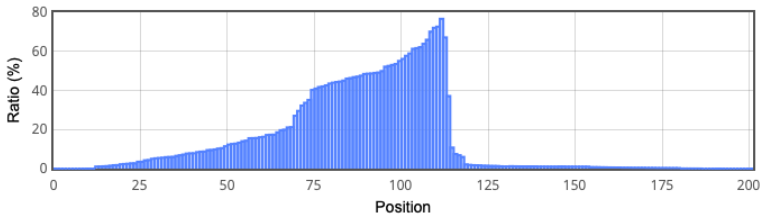
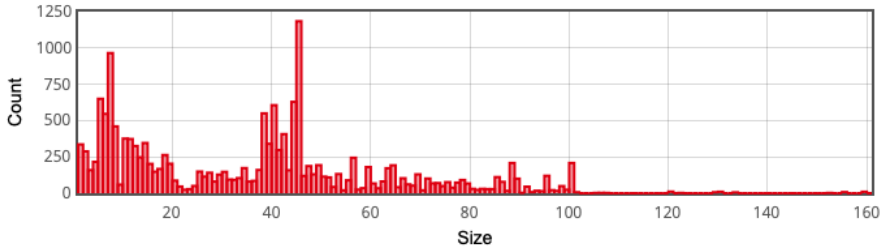
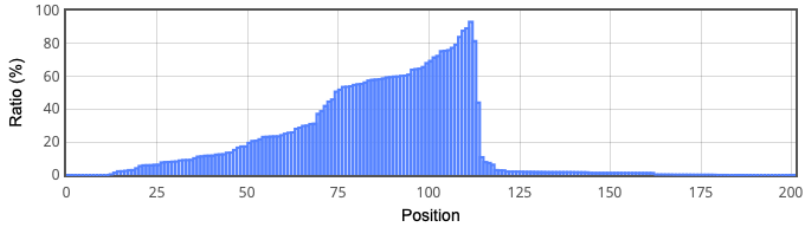
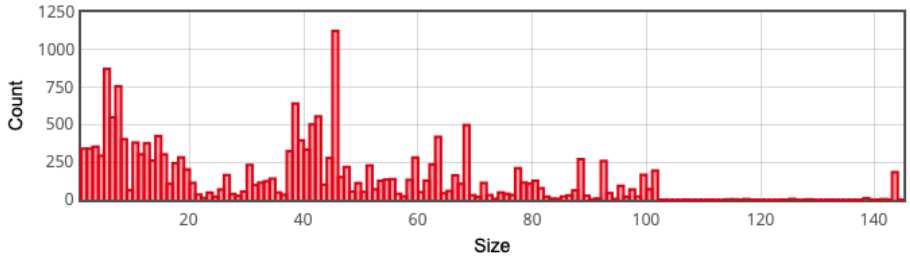
A

Line	Percent Indel	Percent Frameshift
Ven^{+} DownDoc	99%	53.38%
Ven^{Doc} DownDoc	98.2%	59.72%
Ven^{+} UpDoc	82.5%	56.38%
Ven^{Doc} UpDoc	99.7%	67.24%

B

Figure 1.10 A) Knocking out *Veneno* using CRISPR/Cas9. We selected two sites, updoc and downdoc in exons 2 and 5 respectively of *Veneno* to target with CRISPR/Cas9 for mutations.
B) *Veneno* Knockdown with CRISPR/Cas9 is highly efficient and most *Veneno* mutations result in frameshifts. With the exception of 59 UpDoc which had an indel frequency of 82.5%, the knockdown were highly efficient with indel frequencies close to 100%. Analysing the Cas-Analyzer output shows that between one half and two thirds of *Veneno* mutations result in frameshifts, which is consistent with the expectation that around 2/3 indels of random length will result in frameshift.
C) Indels in *Veneno* Knockdown with CRISPR/Cas9 follow a repeatable pattern. (next page) Insertion (i) and Deletion (ii) positions and frequencies of DownDoc and UpDoc knockout lines carrying *Ven*$^{+}$ and *Ven*^{Doc} produced by Cas-Analyzer. Y-axis of blue plot is the percentage of indels which occur at that position (where position 100 is the cleavage point of the CRISPR target). Y-axis of red plot is the number of indels for each size. The plots show similar indel patterns for each given CRISPR target. In particular, insertion positions, as well as deletion positions/sizes where highly similar for each CRISPR target.

C i	Insertion Position	Insertion Size
Ven^{+} DownDoc		
Ven^{Doc} DownDoc		
Ven^{+} UpDoc		
Ven^{Doc} UpDoc		

C ii	Deletion Position	Deletion Size
Ven^{+} DownDoc		
Ven^{Doc} DownDoc		
Ven^{+} UpDoc		
Ven^{Doc} UpDoc		

Ven$^{Doc,\ updoc}$ flies had a viral titre ~220,000-fold higher than *Ven*^{Doc} flies (Figure 1.11, Welch's t-test, t=-13.5, p=0.000283, df=4.13, p value adjusted using FDR with n=2), similar to the level seen in the *Ven*$^{+}$ flies, indicating that the mutation upstream of the *Doc* element resulted in a loss of the resistant phenotype. On the other hand, *Ven*$^{Doc,\ downdoc}$ flies were not significantly different from the *Ven*^{Doc} flies, which means that we found no evidence mutations downstream of the *Doc* element have an effect on the resistant phenotype (figure 1.11). This is consistent with the hypothesis that resistance is caused by *Ven_tra,* which includes exon 2 containing the updoc region, but not exon 5 containing the downdoc region.

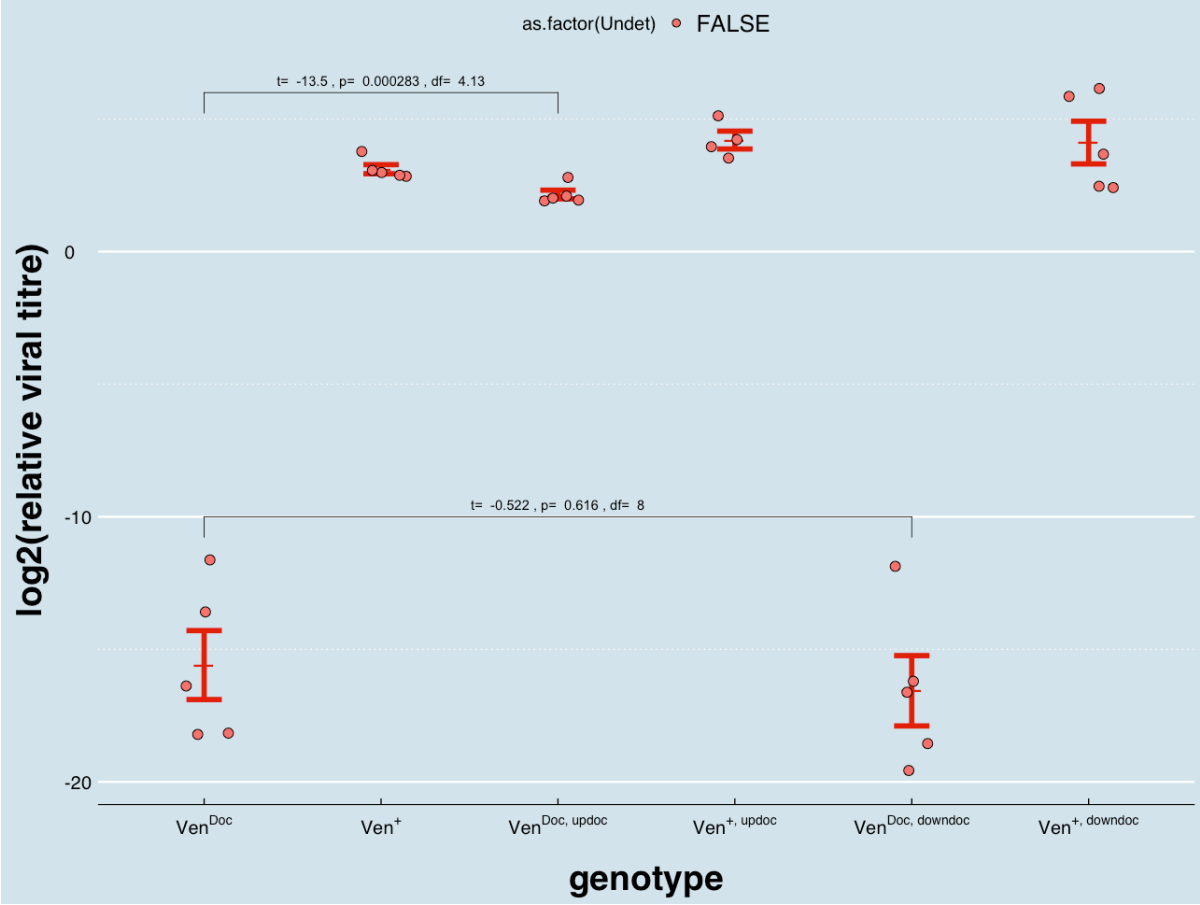


Figure 1.11. Mutating *Ven*^{Doc} upstream of the *Doc* element led to an elimination of the resistant phenotype. Flies having the resistant (*Ven*^{Doc}), and susceptible (*Ven*$^{+}$) alleles of *Veneno* were mutated using CRISPR/Cas9 at the two sites updoc and downdoc. Flies were infected with DAV, and viral titres were measured 3 days post infection. Y-axis shows the viral titre relative to *Actin5C* calculated using -ΔCT values. Each dot represents 6-15 infected flies. Red bars show the standard error of -ΔCT centred around the sample mean. P-values calculated using Welch's t-test, and adjusted for multiple tests using false discovery rate (FDR) method.

Ven_tra causes DAV resistance in cells

We hypothesized that *Ven_tra* might be responsible for DAV resistance. To test this, we expressed a truncated allele of the gene which only includes the translated sequences in *Ven_tra* (we call this allele Ven^{trunc}) in DL2 cells to check whether it is sufficient for increased DAV resistance.

To check whether DAV replicates in DL2 cells, we infected some DL2 cells with DAV. Indeed, we found that 3 days later, the infected cells had an increased viral titre from day 0 which had 2 samples with no virus detected, and a third sample with ~1/2800 of the average titre from day 3 (Figure 1.12). This indicated that cells were successfully infected with DAV meaning that DL2 cells are susceptible to DAV infection, and that this infection can be measured easily using qPCR, which makes these cells a viable system to test resistance. We then determined $TCID_{50}$ using the Reed and Muench method to be the dose in a 100µl of a $10^{-8.5}$ dilution of our viral extract by infecting DL2 cells with serially diluted DAV (See Chapter 1 Methods).

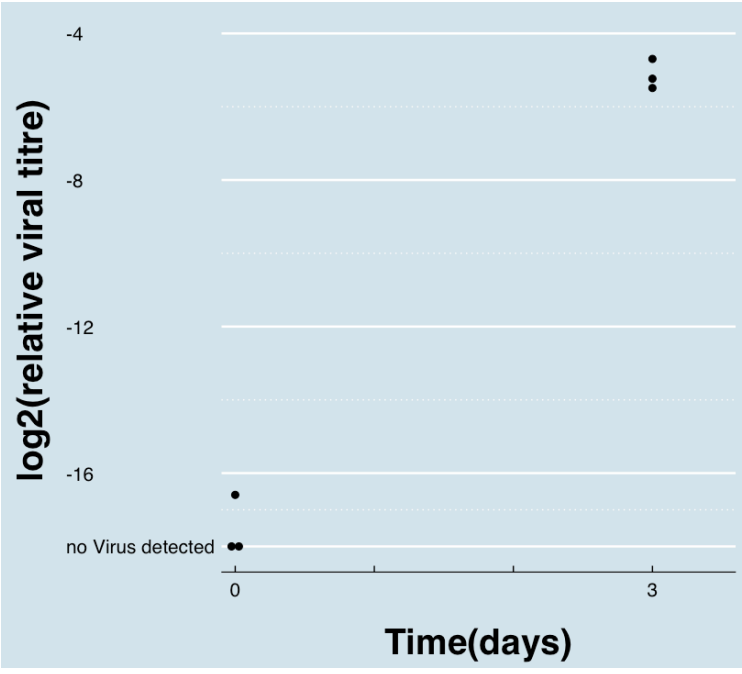


Figure 1.12. DAV titre increases with time in infected DL2 cells. Plot shows log_2(viral titre) of DL2 cells infected with DAV 0 and 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 6 well plate. There is an increase in viral titre with time indicating that DL2 cells are susceptible to DAV infection.

To check whether *Ven*^{trunc} is sufficient to cause DAV resistance in cells, we infected DL2 cells stably transfected with pMT-Puro-*Ven*^{trunc} plasmid and control DL2 cells stably transfected with the pMT-Puro plasmid (without *Ven*^{trunc}), and found that 3 days later, cells with pMT-Puro-*Ven*^{trunc} had a significantly lower viral titre (Figure 1.13, Welch's t-test, t=-6.78, p= 6.5x10^{-5}, df=9.42), which indicates that *Ven*^{trunc} (and therefore its encoded transcript *Ven_tra*) indeed does cause resistance.

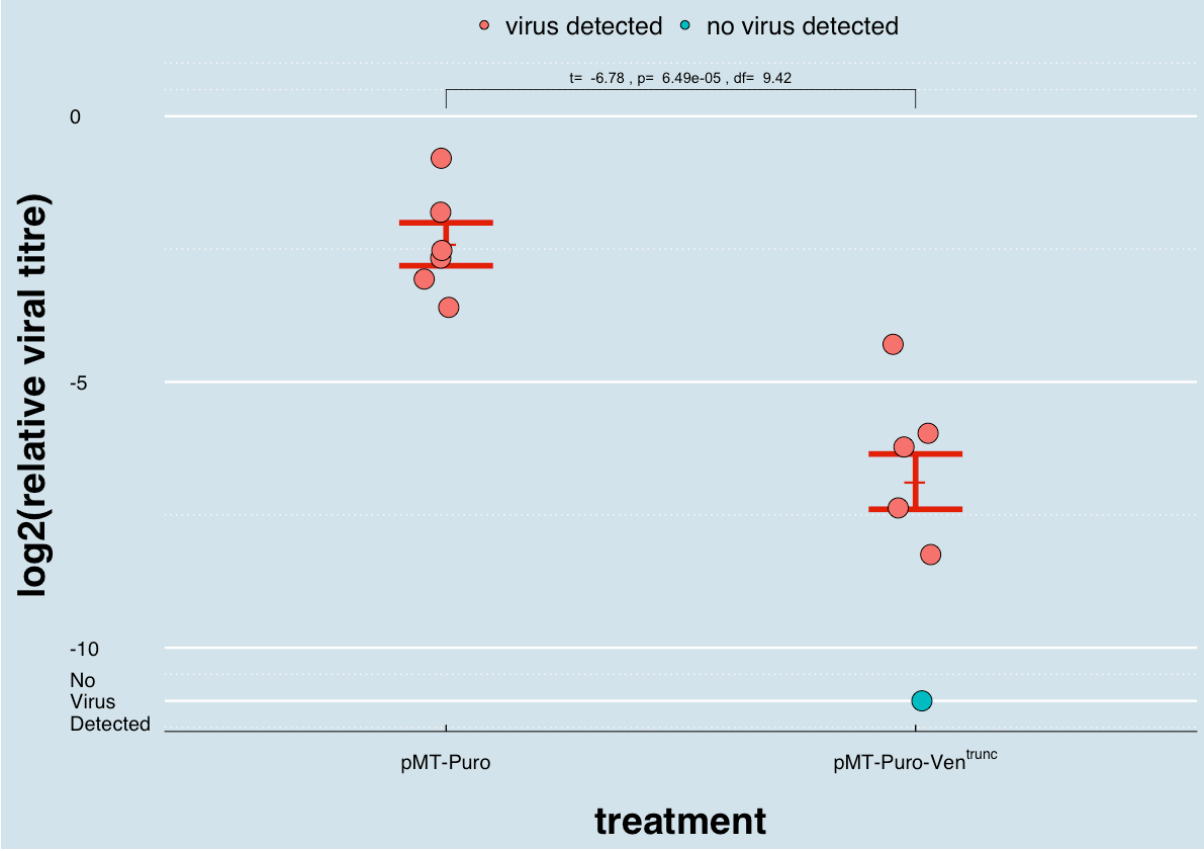


Figure 1.13. *Ven*^{trunc} results in lower DAV titre in DL2 cells. DL2 cells with and without the pMT-Puro-*Ven*^{trunc} were infected with DAV. Viral titre was measured 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. pMT-Puro-*Ven*^{trunc} causes a significant decrease in viral titre from the control, which indicates that the transcript is sufficient to increase resistance in cells. P-values calculated using Welch's t-test.

Discussion

We have shown that a *Doc* transposable element insertion in the coding sequence of the gene *Veneno* results in a dominant gain of anti-DAV antiviral function. This antiviral function appears to be specific to DAV since results from a previous experiment have shown no significant effect of the Ven^{Doc} allele on resistance to FHV, DCV, and DMelSV (Fabian and Cogni, unpublished result). This phenotype is lost if the *Veneno* DNA sequence upstream of the *Doc* insertion is mutated, but it seems unaffected by knocking down *Doc* RNA sequences. The insertion leads to the generation of a truncated transcript *Ven_tra* which contains mostly sequences from the region of *Veneno* upstream of the *Doc* insertion, with 18 bp of its coding sequence having *Doc* element origin. We've not only shown that the resistant phenotype relies on the transcription of the mutated gene, but also that expression of the truncated transcript is sufficient to produce the resistant phenotype. Given these findings, we can conclude that the resistant phenotype is due to the generation of a novel restriction factor coded for by *Ven_tra*, which is the result of a premature termination of transcription as a consequence of the *Doc* insertion in the coding sequence of Ven^{Doc}.

The fact that the coding sequence for *Ven_tra* does not contain the Downdoc site which we targeted for mutations with CRISPR/Cas9 or the DocBegin and DocEnd sites in the *Doc* sequence which we've targeted with RNAi explains why the resistant allele appears to be unaffected by those mutations and that knockdown. On the other hand, the coding sequence for *Ven_tra* does contain the site Updoc targeted for mutations with CRISPR/Cas9 and parts of the VenenoKD site which we targeted with RNAi, which explains why those experiments effectively led to the loss of the resistant phenotype.

We have shown that DL2 cells are an effective way to study DAV infection. We were able to successfully infect the cells with DAV, which replicated to a high titre within 3 days. We have also successfully transfected the cells with plasmids expressing our discovered restriction factor which had a clear effect in cells.

Our work has allowed us to optimize the use of DL2 cells in studying DAV infection. We have already successfully determined the TCID50, and estimated the number of viral particles in our sample, allowing us to use a more statistically robust dose in the experiments in Chapter 2 (~6.32 $TCID_{50}$ per 100µl well). We have also found that increasing cellular density to ~$1.5x10^6$ cells/ml led to a more pronounced, and therefore easier to measure, difference between the susceptible and resistant phenotypes which led us to use these higher densities in Chapter 2. More work can be done on this optimization. In particular, a time series of DAV infection on resistant and susceptible flies should help determine the optimal time point for infection.

Transposable elements are usually detrimental to the fitness of the host. In this case, we have a novel example of a transposable element insertion into a coding sequence resulting in the generation of a novel adaptive trait (whether the mutation leads to an increase in fitness is uncertain as there may be a cost to the increased resistance). This is not entirely unprecedented. There are numerous examples of transposable element insertions resulting in adaptive traits, to mention a few: there are multiple cases of transposable element insertions which have been domesticated by the host, and used to carry out important functions, such as *Het-A* and *TART* (Abad et al., 2004a). There are also well documented associations between transposable element insertions and the generation of insecticide resistance, such as the *Doc* element insertion upstream of *Cyp6g1* which leads to DDT resistance (Schlenke & Begun, 2004). Furthermore, the insertion of a *Doc* element into the coding sequence of *CHKov1* results in resistance to DMelSV (Magwire et al., 2011). That last example is particularly striking in its parallels to our findings here: a *Doc* element insertion into the coding sequence of a gene results in increased resistance to a virus. We will discuss this parallel further, and whether it is a coincidence or perhaps a consequence of the *Doc* element having a special antiviral property, in Chapter 2.

We do not have much data on the prevalence of Ven^{Doc} in nature, which can be obtained through genotyping of natural worldwide samples. Given that the insertion was in only 2 of the 162 DGRP lines which we've genotyped, and that we haven't found the insertion in any of the lab stocks which we've used in our experiments, our initial guess (our prior) is that Ven^{Doc} is a rare allele

despite the seemingly beneficial function it provides. However, this hypothesis is not inconsistent with our findings. It is possible that the allele is currently rare due to it having arisen recently, and that it is currently undergoing positive selection. Alternatively, it may be the case that despite its beneficial antiviral function, the Ven^{Doc} allele loses the function of Ven^{+} (which is currently unknown), or that Ven^{Doc} itself has a negative side-effect, and that this incurs a fitness cost equal or greater to the fitness benefit of having the insertion, leading to neutral or negative selection on the insertion preventing the spread of the allele. In the latter scenario, it may be possible that the fitness benefits of the insertion vary as a function of the prevalence of DAV infection in a certain population, and that, therefore, the allele is undergoing different types of selection in different populations. This could be tested by comparing Ven^{Doc} prevalence with DAV prevalence in different populations. Furthermore, the age of the allele could be estimated by checking for linkage disequilibrium with nearby sites, especially if more flies with Ven^{Doc} are found in nature.

In another scenario, the effect of the allele might be dependent on the strain of DAV which infects the flies. One limitation of our experiments is that we've only used one strain of DAV which originated in Australia, whereas our flies are either lab stocks or stocks from the DGRP line which originated in Raleigh North Carolina. It is therefore possible that the flies we've tested have not adapted to this particular strain of virus, and that our results might differ if we use a different viral strain to which our test flies have adapted. However, given the massive effect of the *Doc* insertion allele on resistance, which is very large compared to the variance of resistance in lines lacking the allele, this concern is unlikely to prove problematic.

Methods

Characterising genetic variation in resistance to DAV

To assay the viral load in the Drosophila Genetics Reference Panel (Mackay *et al.* 2012), we infected the lines with DAV. Prior to this, lines infected with *Wolbachia* were treated with tetracycline as in (Martinez *et al.* 2016). We infected flies with DAV (see main methods) and froze them three days post infection. Infected frozen flies were homogenized in 400µl of

Trizol (Ambion, USA), and RNA was extracted with Direct-zol 96-well plate kit (Zymo), without the DNase digestion step. RNA elution was performed in 100 µl of nuclease-free water. Viral titre was measured using one-step RT-qPCR (see main methods).

To determine dominance of the DAV resistance trait, we used the two resistant (239 and 362) and two susceptible DGRP lines (48 and 91). We generated flies from the following crosses: "239 x 239" and "362 x 362" (resistant); "48 x 48" and "91 x 91" (susceptible); "239 x 48", "239 x 91", "362 x 48" and "362 x 91" (heterozygotes). We performed up to four replicates for each cross each containing a pool of 20 females. We infected flies with DAV, froze them 3 days later, extracted their RNA using TRIzol (see main methods) then measured viral titre using one-step RT-qPCR (see main methods).

Mapping to chromosome

To map the resistant trait to chromosome, we used a double balancer line (w^*; If/CyO; MKRS/TM6b) to extract individual chromosomes from the two resistant DGRP lines (239 and 362). Because DAV resistance is dominant, we produced flies with an individual chromosome from a resistant line and two chromosomes from a susceptible line (DGRP 48 and 91). To extract the X chromosome, virgin females of the resistant lines were crossed with double balancer males. Males from this cross that have the resistant X chromosome were crossed with virgin females from a susceptible line and females carrying markers from the balancer stock for second and third chromosome present in the F1 male were selected and assayed for DAV resistance. To extract the second and third chromosome from the resistant lines, we crossed virgin females from the double balancer line with males from the resistant line. Males from this cross were crossed with virgin females from the susceptible lines. Females lacking markers on the second chromosome (carrying only second chromosome from the resistant line), females lacking markers on the third chromosome (carrying only third chromosome from the resistant line), and females with markers on both second and third chromosomes (controls with no chromosome from the resistant line) were selected and assayed for DAV resistance. DAV resistance was assayed in three replicates for each chromosome for each line, and because there was no difference between lines DGRP line 239 and 362, results for both lines are present together. To assay DAV resistance, we extracted

RNA using TRIzol extraction (see main methods) then measured viral titre using one-step RT-qPCR (see main methods).

QTL mapping

To produce recombinant third chromosomes, we first crossed virgin females of each resistant line with males of a susceptible line (239 x 373 and 362 x 306). Next, virgin females obtained from these crosses were crossed with males from a third chromosome balancers stock (TM2/TM6B). We assayed DAV resistance in this balancer stock and determined that it has the susceptible phenotype. 84 recombinant lines for DGRP-239 and 72 lines for DGRP-362 were produced and a single female from each line was infected with DAV (see main methods). 3 days later, infected frozen flies were homogenized in 400µl of Trizol (Ambion, USA), and RNA was extracted with Direct-zol 96-well plate kit (Zymo), without the DNase digestion step. RNA elution was performed in 100 µl of nuclease-free water, then viral titre was measured using one-step RT-qPCR (see main methods).

Using the VCF file from DGRP freeze 2 (http://dgrp2.gnets.ncsu.edu/data.html), we selected 10 SNPs along the third chromosome that were different between resistant and susceptible lines (resistant: 239 and 362; susceptible: 373 and 306), but shared within the two resistant and susceptible genotypes, for genotyping with high-resolution melt analysis (HRM). We avoided G/C and A/T SNPs since they are hard to separate using HRM, and designed the HRM primers to target SNPs that have no other SNPs 100bp on either side to avoid possible interference from other SNPs. We extracted DNA with ZR-96 Quick-gDNA kit (Zymo Research), then genotyped the 10 SNPs using HRM, performed with MeltDoctor HRM Master Mix (Applied Biosystems) according to manufacturer protocol. All HRM reactions were run with multiple homozygous and heterozygous controls.

Fine-scale mapping

To identify candidate genes involved in resistance to DAV we created lines recombinant between the resistant and the susceptible chromosomes in a 4.4 million bp large region on

chromosome *3R* that included the resistance QTL. Fly lines recombinant in this target region were generated through a series of crosses involving multiple lines. In brief, the recombinants lines are derived from a resistant DGRP-362 line (w^{1118}; 362/TM3 with segregating chromosome 2) and a susceptible line with two phenotypic markers at the ends of the target region (w^{1118}; Scr^{1},P(*mini-white*)/TM3), hereafter referred to as "parental lines". The first marker used was the Scr^{1} allele (Genome version 5 coordinates, 3R:2651265; BDSRC id = 2184), a recessive-lethal mutation that causes a reduced sex combs phenotype in heterozygous adult males. The second marker was a P-element insertion located in the gene *CG4820* (Genome version 5 coordinates, 3R:7042524; BDSRC id = 30182) that carries a *mini-white* gene that induces a reddish eye colour in white-eyed flies.

To confirm the presence of the Scr^{1} allele in the susceptible parent line we obtained allele information on FlyBase and identified that this allele is due to a G/A polymorphism at 3R:2651265 (genome release 5; Larkin et al., 2021). We then performed PCR in 15ul reaction volumes using Taq Polymerase (NEB) to amplify this SNP in the heterozygous susceptible parent to create a ~500 bp product and digested this using the enzyme BsrBI (NEB) for 2 hours at 37ºC followed by 20 min 80ºC inactivation. This process digested the reference allele (G), therefore resulting in three PCR product pieces in the susceptible parent. The P-element line without the Scr^{1} allele used to create the susceptible parent was taken as a control. We used a similar approach to genotype one variant (3R:4260142; A/G) in the recombinant lines, where we digested a ~600 bp PCR product with FokI (NEB) for 4 hours at 37ºC followed by 20 min inactivation at 65ºC. Here, only the amplicons from the resistant (G) but not susceptible lines (A) were digested.

We crossed the parental lines and collected ~800 virgin female offspring. These were then crossed to ~800 males of a line carrying the recessive lethal allele and a third chromosome balancer with heat-shock-inducible mortality (w*; Scr^{1}/TM3-*hs-hid*) in population cages. We added two large Petri dishes containing apple juice agar with yeast paste on top to the cages. Every 24 hours, Petri dishes were exchanged to supply fresh food. Using a brush and distilled H_2O, we collected eggs from the plates into a 50ml Falcon tube and transferred 500 µl of this egg-water solution to 1.5ml Eppendorf tubes. From these, we distributed 7 µl egg-water solution to vials with standard cornmeal-yeast food. This procedure allowed us to control the

number of eggs going into each vial (~100 eggs). To reduce the number of non-recombinant genotypes, we heat-shocked each vial in a water bath at 37ºC for 2 h at 3 and 4 days after egg collection to kill larvae carrying TM3-*hs-hid*. After eclosion, we were then able to distinguish the remaining non-recombinant genotype (white eyes, stubble) from the desired recombinant adult flies (reddish eyes). Single male and female recombinants (*w**; *Scr*^{1}/recombinant, P(*mini-white*)) were then crossed to a line carrying the recessive, lethal *Scr*^{1} allele (*w**; *Scr*^{1}/TM3) to obtain a final recombinant line. We are aware that using females might lead to further unwanted recombination, however, have not found this to be a major problem as recombinants close to the gene are extremely rare. Male and virgin female offspring of a single cross were then mated to create a stable recombinant fly line (with segregating TM3). In total we generated ~640 recombinant lines, each of which resulted from an independent recombination event of parental susceptible or resistant lines in the target region.

Using the VCF file from DGRP freeze 2 (http://dgrp2.gnets.ncsu.edu/data.html), we selected variants unique to 362 or both the resistant lines, and confirmed that these SNPs are rare and do not occur in the susceptible parental line (as we did not know the genomic sequence of the susceptible parent). We identified restriction site polymorphisms using RestrictionMapper (http://www.restrictionmapper.org). We extracted DNA from the recombinants using Chelex extraction (see main methods) and genotyped them by digesting PCR products with restriction enzymes. We also used Sanger sequencing on lines where the flanking markers came from different parents to more precisely define recombination breakpoints. Adult females were infected with DAV, including the parental lines as controls, and frozen 3dpi, their RNA extracted, then viral titre was measured using one-step RT-qPCR (see main methods). To identify potential differences in sex, we also determined viral load of infected males in a subset of recombinants and the parental lines. For most lines, we have only measured a single biological replicate, so that our level of replication was within groups that shared a similar recombination breakpoint. Recombinant lines that defined the location of the resistant allele on a smaller scale were measured for DAV viral load and genotyped in at least 4 biological replicates. Each replicate in the pricking experiment contained a pool of 5-20 flies (average = 16.6). Viral load was mainly determined in “heterozygous” recombinants that were carrying the balancer with the recombinant chromosome. We also measured several lines as “homozygous” with two copies of the recombinant chromosome, and for some lines both types were assessed. In line with the resistant allele being dominant, there was no detectable difference between heterozygous and homozygous genotypes.

RNAi knockdown of candidate genes

To test whether the genes within or near the region identified by genetic mapping had an antiviral function we performed RNAi-mediated knock-downs of candidate genes in a resistant genetic background. To do this we used lines that express RNAi constructs from the VDRC panel. Using a series of crosses, we combined each of the UAS-RNAi chromosomes of the VDRC lines with chromosome 3 from the resistant DGRP-362 (w*; UAS-RNAi; 362). As not all crosses yielded viable offspring, we could not create this line for all genes. We then crossed 2-3 males of the resistant UAS-RNAi lines to 2-3 female virgins of the *da-Gal4* or *tub-Gal80^{ts}*; *da-Gal4* lines at 18ºC. The *da* (*daughterless*) promotor is ubiquitously expressed. As some crosses resulted in complete lethality, too small number of offspring or morphologically impaired phenotypes with *da-Gal4*, we used the driver line in which Gal4 expression could be regulated by a temperature-sensitive *tub-Gal80* construct (*w*; tub-Gal80^{ts}; da-Gal4* – kindly provided by A. Leitao) for these lines. Female offspring at 0-3 days of age were then transferred to 29ºC and left for ~4 days before infections with DAV. Viral load was scored from 8 replicates with pools of 2-16 females (average = 13.1) as described above using one-step RT-qPCR (see main methods).

Locating *Doc* Element Insertion

We used TouchDown PCR with Taq (see main methods) to amplify the DNA of DGRP lines 306, 373, 362, and 239 using primer pairs 3JK_1a, 3JK_1b, 3JK_2a, and 3JK_2b (See Appendix 1) and ran the products through an agarose gel (see main methods).

We used ExoSap to clean up the products of the PCR reactions that had non-specific bands, and sanger sequenced using (3JK_1a, 3JK_1b, 3JK_2a, and 3JK_2b) (see appendix 1), which allowed us to obtain short sequences despite the products having multiple bands. Sequencing showed *Doc* element sequences which we used to design further primer pairs (CGthenDoc 1->7 and DocthenCG 1->7) (See appendix 2) which bridged the two breakpoints of the *Doc* insertion and

the flanking sequence. We used those primers to amplify and sequence the insertion site of the *Doc* element.

Sequencing *Doc* Element

To amplify the *Doc* element itself, we used Platinum Pfx DNA Polymerase (ThermoFisher) and primers spanning the entire *Doc* insertion (CGspanDoc 1->7) (See Appendix 2). For a 20µl reaction we used 2 µl of the 10X Pfx Amplification Buffer, 0.6µl of 10mM dNTP mix, 0.4µl of 50mM $MgSO_4$ 0.6µl of 10µM primer mix (forward and reverse), 1µl of template DNA (DGRP line 362), 0.2µl of Platinum Pfx DNA Polymerase, and 15.2 µl of nfH_2O and we followed the settings in table 1.1.

	Temperature	Time
1x	94°C	2 min
35x	94°C	15 s
	55°C per cycle	30 s
	68°C	6 min

Table 1.1 Cycling Conditions for Amplifying *Doc* element with Platinum Pfx

We ran the products in an agarose gel (see main methods) to find the primer pairs which worked best, and we cleaned the product of primer pairs CGspanDoc4 (FR) and CGspanDoc5 (FR) (Appendix 2) using ExoSap Cleanup (see main methods), and sequenced each product using CGspanDoc_R5, Doc 1 -> 7 (both F and R), DocthenCG_3 (F and R), and CGthenDoc_3R using sanger sequencing (see main methods; primers in appendix 2). We assembled the sequences to the reference sequence of the *Doc* element (O'Hare et al., 1991) as well as to the reference sequence of *Veneno* (dmel_r6.28_FB2019_3; Larkin et al., 2021), which we modified to include the sequence of the *Doc* element at the insertion site.

Genotyping *Veneno*

We genotyped DGRP lines' *Veneno* for presence/absence of the *Doc* element insertion using TouchDown Taq PCR amplification (see main methods) with 3 primers: DocthenCGR2, DocThenCGF2, and CGThenDocF5 (Figure 1.5B, Appendix 2). In the presence of the *Doc* insertion, DocthenCGR2 and DocThenCGF2 would amplify a 730bp region which includes sequences from the *Doc* element and the upstream *Veneno* sequence, and would outcompete the much larger product that would be formed by DocthenCGR2 and CGThenDocF5 (which was too long to be amplified anyway given the 1 minute extension time) (Figure 1.5C). In the absence of the *Doc* insertion, the only product amplified is the one formed by DocthenCGR2 and CGThenDocF5 which was 390bp (Figure 1.5C). We ran the amplified product through an agarose gel (see main methods) to determine the length (and therefore the genotype).

Veneno RNAi

We knocked down *Veneno* using VDRC fly line 24094 carrying the UAS driven RNA hairpin construct (UAS-VenenoKD) (Dietzl et al., 2007; provided kindly by the Vienna Drosophila Resource Center). The construct, which is on the second chromosome, was combined with Ven^{Doc} (DGRP 362) on the third chromosome and crossed with a second chromosome daughterless::Gal4 (daGal4) line, resulting in a UAS- VenenoKD /dagal4 heterozygote ubiquitously expressing the hairpin construct. The construct is then processed via endogenous machinery into 21 bp RNAi targets which are then used by the siRNA pathway to break down *Veneno* transcripts.

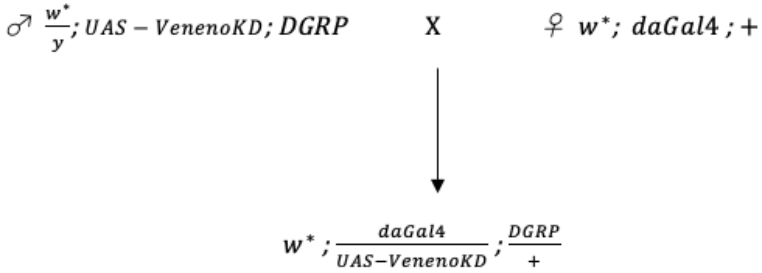


Figure 1.14 Crossing Scheme For Veneno RNAi Experiment. We used these crosses to produce w*; daGal4/UAS-VenenoKD; Ven^{Doc}/Ven^{+}

We infected lines w*; daGal4/UAS-VenenoKD; Ven^{Doc}/Ven^{+}, w*; UAS-VenenoKD; Ven^{Doc}, and w*; daGal4; Ven^{+} with DAV then extracted their RNA with TRIzol, reverse transcribed it, and tested it for DAV infection using qPCR (see main methods).

Targeting *Doc* element with RNAi

Two sites in the *Doc* element insertion (each in one of the two coding sequences) were selected as targets for RNAi knockdown. To design the oligos, we used methods on the DRSC/TRiP Functional Genomics Resources website (https://fgr.hms.harvard.edu/cloning-and-sequencing): We chose a 21 nucleotide sequence using the DSIR website (http://biodev.cea.fr/DSIR/DSIR.html), swapped uracil for thymine to get the guide strand DNA, and reverse complemented the sequence to get the passenger strand DNA. For the top strand oligo, we added ctagcagt to the 5' end of the passenger strand DNA, tagttatattcaagcata between the passenger strand DNA and the guide strand DNA, and gcg to the 3' end of guide strand DNA. For the bottom strand oligo, we added aattcgc to the 5' end of the passenger strand DNA, tatgcttgaatataacta between the passenger strand DNA and the guide strand DNA, and actg to the 3' end of the guide strand DNA. (See Appendix 3)

We carried out the rest of the cloning into a WALIUM 20 vector using the methods on the DRSC/TRiP Functional Genomics Resources website (https://fgr.hms.harvard.edu/cloning-and-sequencing). We maxiprepped the plasmid, and sent it to The University of Cambridge Department of Genetics Fly Facility to be injected into *phiC31; attP40* embryos, which express phiC31 driven by *vasa*, and has a chromosome 2 landing site (*attp40*) to obtain w*;UAS-gene/+;+.

We then followed the following crossing scheme in Figure 1.15 using either this w*;UAS-gene/+;+ line or an untransformed control line w*;attP/+;+ as well as a double balancer line, DGRP lines 362 and 850, and a line carrying Actin5C::Gal4 (referred to as gal4 in figure 1.15) with TM6B/MKRS on Chr 3, which was provided by Chuan Cao to produce lines w*;gal4/UAS-DocKD;DGRP and w*;gal4/attP;DGRP which we then infected with DAV then extracted their RNA with TRIzol, reverse transcribed it, and tested it for DAV infection using qPCR (see main methods). We also used qPCR to measure expression of the *Doc* element in the two regions targeted (using BeginExpFR and EndExpFR primers in Appendix 19).

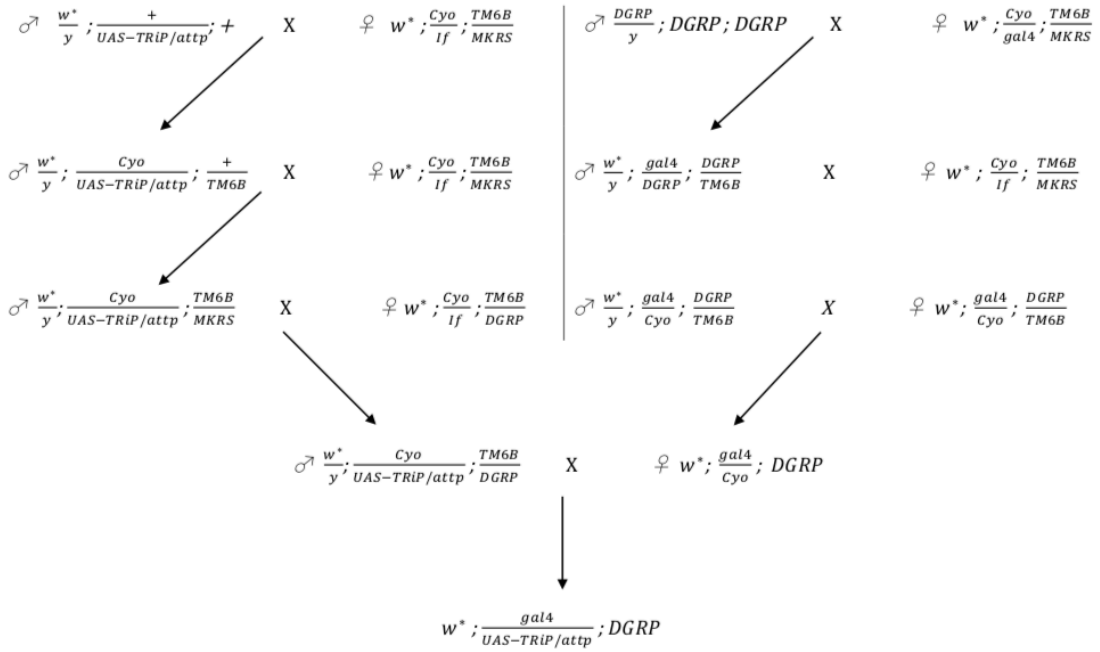


Figure 1.15. Crossing Scheme for *Doc* Element RNAi Experiment. We used these crosses to produce w*;UAS-gene/Gal4;DGRP_362 and w*;UAS-gene/Gal4;DGRP_850.

Mutagenesis with CRISPR/Cas9

To perform mutations using CRISPR/Cas9, we generated transgenic flies that constitutively express both the guide RNA and the Cas9 protein. To create the gRNA (guide RNA) transgenic line, we used the plasmid pCFD3:U6:3-gRNA (Addgene 49410; (Port et al., 2014)), which expresses a single gRNA using U6:3 which is the strongest U6 promoter in Drosophila (this was kindly provided by Simon Bullock). gRNA expression oligos (Appendix 4) were designed and cloned with pCFD3 following the instructions on the crisprflydesign website: http://www.crisprflydesign.org/wp-content/uploads/2014/05/Cloning-with-pCFD3.pdf. We prepared maxipreps of the plasmids, and sent them for transformation onto the second chromosome *attp40* landing site using phiC31 expressing stocks (driven by a *nanos* promoter) with a *vermillion* mutation at The University of Cambridge Department of Genetics Fly Facility, producing flies with gRNA/Cyo on the second chromosome. We used this line as well as double balancers and DGRP lines 59 and 362 to produce w*;Cyo/gRNA;DGRP (Figure 1.16).

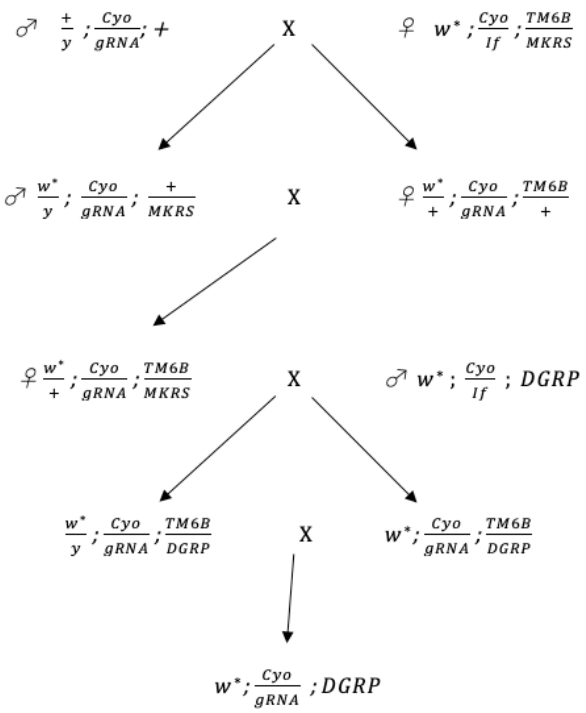


Figure 1.16. Crossing Scheme for CRISPR gRNA flies. We used these crosses to produce w*;Cyo/gRNA;$DGRP_{59}$ and w*;Cyo/gRNA;$DGRP_{362}$.

We used CFD1 (Bloomington: 54590), which ubiquitously expresses cas9, (kindly provided by The University of Cambridge Department of Genetics Fly Facility), and DGRP lines as well as double balanced flies to produce the lines act5c-cas9;+;DGRP_59 and act5c-cas9;+;DGRP_362 as shown in the following crossing scheme (Figure 1.17).

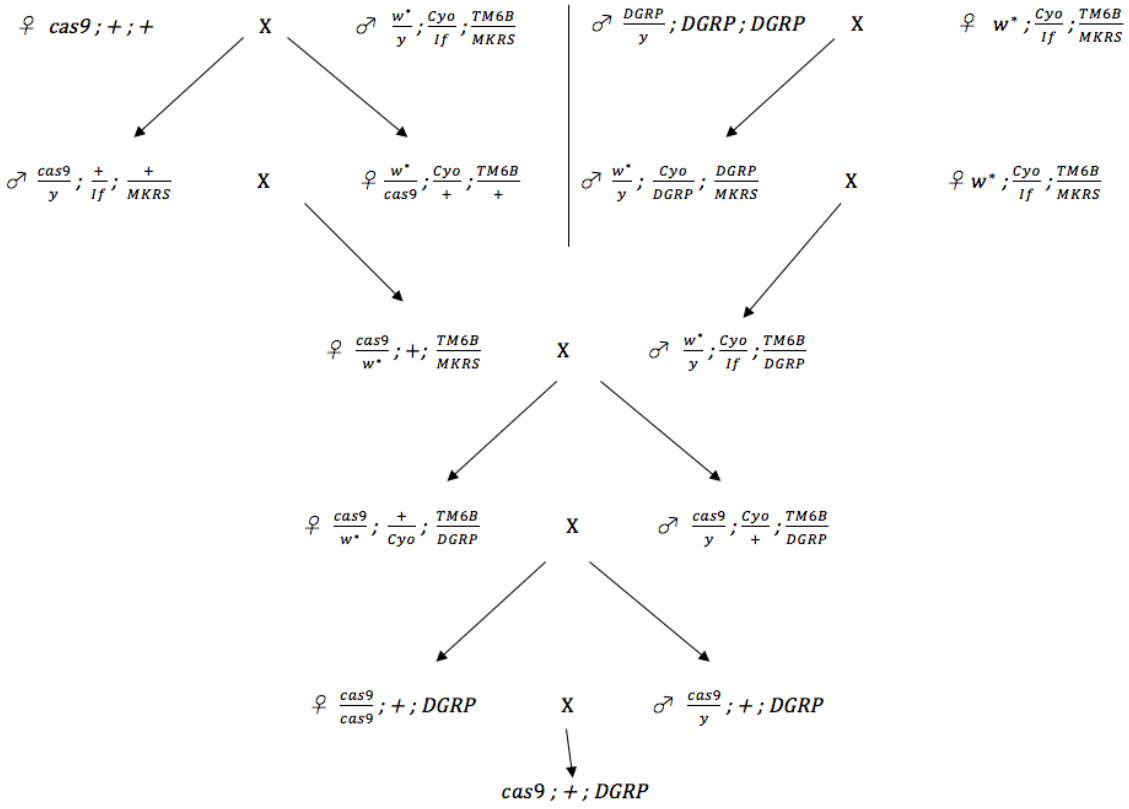


Figure 1.17. Crossing scheme for Cas9 flies. We used these crosses to produce act5c-*cas9*;+;DGRP_59 and act5c-*cas9*;+;DGRP_362.

Finally we crossed w*;Cyo/gRNA;DGRP with cas9;+;DGRP to obtain fly line cas9/w*;gRNA/+;DGRP, which expresses both cas9 and gRNA throughout somatic and germline cells, and has a DGRP background on chr 3. We infected lines cas9/w*;gRNA/+;DGRP, and cas9/w*;attP/+;DGRP with DAV then extracted their RNA with TRIzol, reverse transcribed it, and tested it for DAV infection using qPCR (see main methods).

Sequencing CRISPR target sites with Illumina

We used Illumina sequencing to check whether the sites targeted for mutation using CRISPR/Cas9 were indeed successfully mutated. We designed primers to amplify the target sites for amplification: we selected regions around DownDoc or UpDoc to amplify, then created primer pairs amplifying these regions then, to create overhangs for the Nextera XT index primers, added TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG to the 5' end of the forward primer and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG to the 5' end of the reverse primer. (See appendix 5)

We amplified target sites by PCR using Phusion® High-Fidelity DNA Polymerase with settings in table 1.2. Reaction set up: 6µl Phusion HF buffer, 1.5µl dNTP (0.5mM), 20.4µl nfH_2O, 0.8µl primers (10mM each), 0.3µl Phusion HF polymerase, 1µl DNA template.

	Temperature	Time
1x	98°C	30 s
10x	98°C	10 s
	62°C + 1°C per cycle	15 s
	72°C	15 s
15x	98°C	
	72°C	30 s
1x	72°C	5 mins

Table 1.2 Cycling Conditions for TouchUp 62-72 used to amplify Ilumina targets and add overhangs

We used KAPA Pure Beads (Roche) to clean up the PCR product, following the manufacturer's protocol. We prepared dual indexed libraries from our PCR product by adding adapters using Q5® Hot Start High-Fidelity DNA Polymerase, and Nextera XT Index Kit v2 primers (N716 -> N720 and S521-S522) (Ilumina), using the cycle settings in table 1.3.

	Temperature	Time
1x	95°C	3 mins
8x	95°C	30 s
	55°C	30 s
	72°C	30 s
1x	72°C	5 mins

Table 1.3 Cycling Conditions for library preparation with Nextera XT Index Kit

We then used KAPA Pure Beads (Roche) to clean up the indexed samples following manufacturer's protocol. We checked for the correct product size by running our samples on an agarose gel, and quantified the DNA using Qubit (see main methods).

Libraries were normalized along with others in the pool, and sent to the Biochemistry sequencing service for MiSeq using the 2x250bp paired end reads, nano v2 kit.

Nanopore (direct RNA)

We used Nanopore direct RNA sequencing to sequence the whole transcriptome of flies carrying Ven^{Doc}: We extracted RNA from ~12 female flies from the 362 DGRP line using TRIzol extraction. We checked for RNA quality using TapeStation analysis (kindly done by Vicki Murray and Maike Paramor at the Systems Biology Department at the University of Cambridge). We selected for polyadenylated RNA using Dynabeads mRNA purification kit (Invitrogen) substituting KAPA wash buffer for Invitrogen wash buffer, and otherwise following the manufacturer protocol. We sequenced the RNA using Direct RNA Sequencing Kit SQK-RNA002 (Oxford Nanopore) and the MinIon (Oxford Nanopore) connected to a MinIT device (Oxford Nanopore).

Nanopore (PCR cDNA)

We used Nanopore PCR cDNA sequencing to sequence the whole transcriptome of flies carrying Ven^{Doc}: We extracted RNA from ~12 female flies from the 362 DGRP line using TRIzol extraction, then purified using Qiagen Rneasy Mini Kit column following manufacturer protocol. While the RNA was bound to the column, we digested any possible contaminant DNA using the Qiagen Rnase-Free Dnase Set following manufacturer protocol. We ran our sample through an agarose gel to check for quality. We sequenced the RNA using the cDNA-PCR Sequencing Kit (SQK-PCS109) following the manufacturer's protocol, using KAPA pure beads instead of AMPure XP beads and RNAsin Ribonuclease inhibitor (Promega) instead of RNAseOUT (ThermoFisher).

Bioinformatics

To analyse nanopore data we first used pychopper (ver 2.4.0, https://github.com/nanoporetech/pychopper) to identify, trim, and orient full length cDNA reads, we then used Minimap2 (ver 2.1, https://github.com/lh3/minimap2) to align our sequences to the *D. melanogaster* genome (dmel_r6.28_FB2019_3; Larkin et al., 2021) modified to include the *Doc* element insertion. We followed the recommended minimap2 settings for Nanopore transcriptome data. We used samtools (ver 1.9, https://github.com/samtools/samtools) to sort and index our alignment, then used the pinfish pipeline (https://github.com/nanoporetech/pinfish) (which clusters reads with similar exon/intron structures then generates consensus exon boundaries for each cluster) to find likely transcripts based on the aligned transcriptome. We used the default recommended settings with pinfish except for the minimum cluster size (smallest number of reads that can comprise a cluster) which we set to 4 to account for less abundant transcripts. The script which we used for all of this Bioinformatic analysis can be found on github: https://github.com/osamabrosh/NanoporeDataAnalysis .

Testing DAV infection in cells

To test whether DAV replicates in DL2 cells, we infected DL2 cells with DAV: We diluted the

cells to ~1.5x10^5 cells/ml, counting the cells under the microscope using Fast-Read 102 cell counting chambers (Biosigma). We placed 3ml of our diluted cells in a Costar 6-well TC plate (Corning). We infected the cells with DAV diluted with medium at a 10^{-8} dilution (using 3µl of 10^{-5} dilution DAV in 3ml of cells. 72 ± 3 hours later, cells were spun down, medium was removed, and cells were dissolved with 30µl nfH_2O and 200µl of TRIzol Reagent (ThermoFisher), then frozen at -80°C. Then RNA was extracted, reverse transcribed, and viral titre was measured using qPCR (see main methods).

Determining TCID50 (Median Tissue Culture Infectious Dose)

To determine TCID50, we diluted DL2 cells to ~150,000 cells/ml of medium . 24 hours later, we moved the cells to 96 well plates with 90µl of the cell solution per well. We then added DAV dissolved in medium diluting it serially at each column of the plate (by adding 30µl of the initial solution of DAV diluted in medium to the first column, then moving 30µl from the first column to the next column and so on) thereby infecting the plate with DAV at varying concentrations, with each column of 8 replicates having ¼ of the DAV concentration of the previous column. 72 ± 3 hours later, we spun down the cells, removed the medium, and dissolved the cells in 50µl nfH_2O then froze them in 150µl TRIzol LS Reagent (ThermoFisher) at -80°C until RNA extraction. After RNA extraction, reverse transcription, and measurement of viral titre, we counted the proportion of wells infected at each dilution. We then used the Reed and Muench method (Reed & Muench, 1938) to calculate the dilution of our viral extract at which 50% of our 100µl wells become infected ($10^{-8.5}$). Assuming that the number of viral particles in a well follows a Poisson distribution based on the concentration of virus and the volume in the well, we calculated that our original viral sample had 2.18x10^9 infective viral particle/ml.

Expressing *Ven*^{trunc} in cells

We created plasmids expressing *Ven*^{trunc} using NEBuilder (see main methods): We inserted the sequence of *Ven*^{trunc} with both UTRs removed into pMT-puro (addgene 17923, Figure 1.18) which has been digested with NotI-HF. The *Ven*^{trunc} sequence was amplified in three parts from DGRP_362 DNA and was inserted either without an added stop codon to produce pMT-*Ven*^{trunc} –V5 -puro (Figure 1.19, primers in appendix 7: Ven_exon 1FR, Ven_exon2FR and Ven_exon3FR_V5) or with a stop codon to produce pMT- *Ven*^{trunc} -puro (Figure 1.20, Primers in appendix 6: Ven_exon 1FR, Ven_exon2FR and Ven_exon3FR_stop).

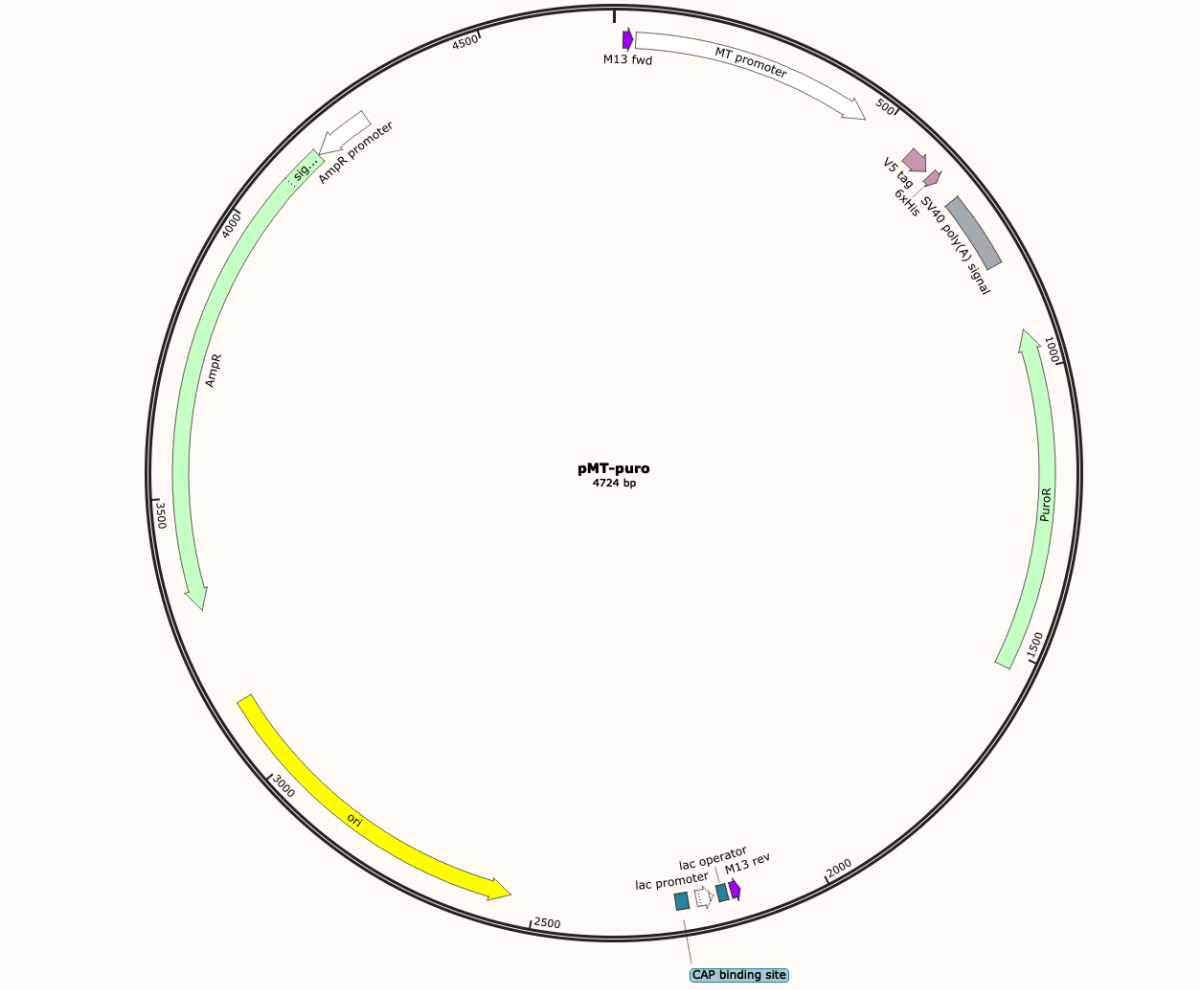


Figure 1.18 pMT-puro Plasmid (addgene). Expresses PuroR gene causing puromycin resistance, and the MT promoter which expresses the downstream sequence in the presence of $CuSO_4$.
https://www.addgene.org/17923/

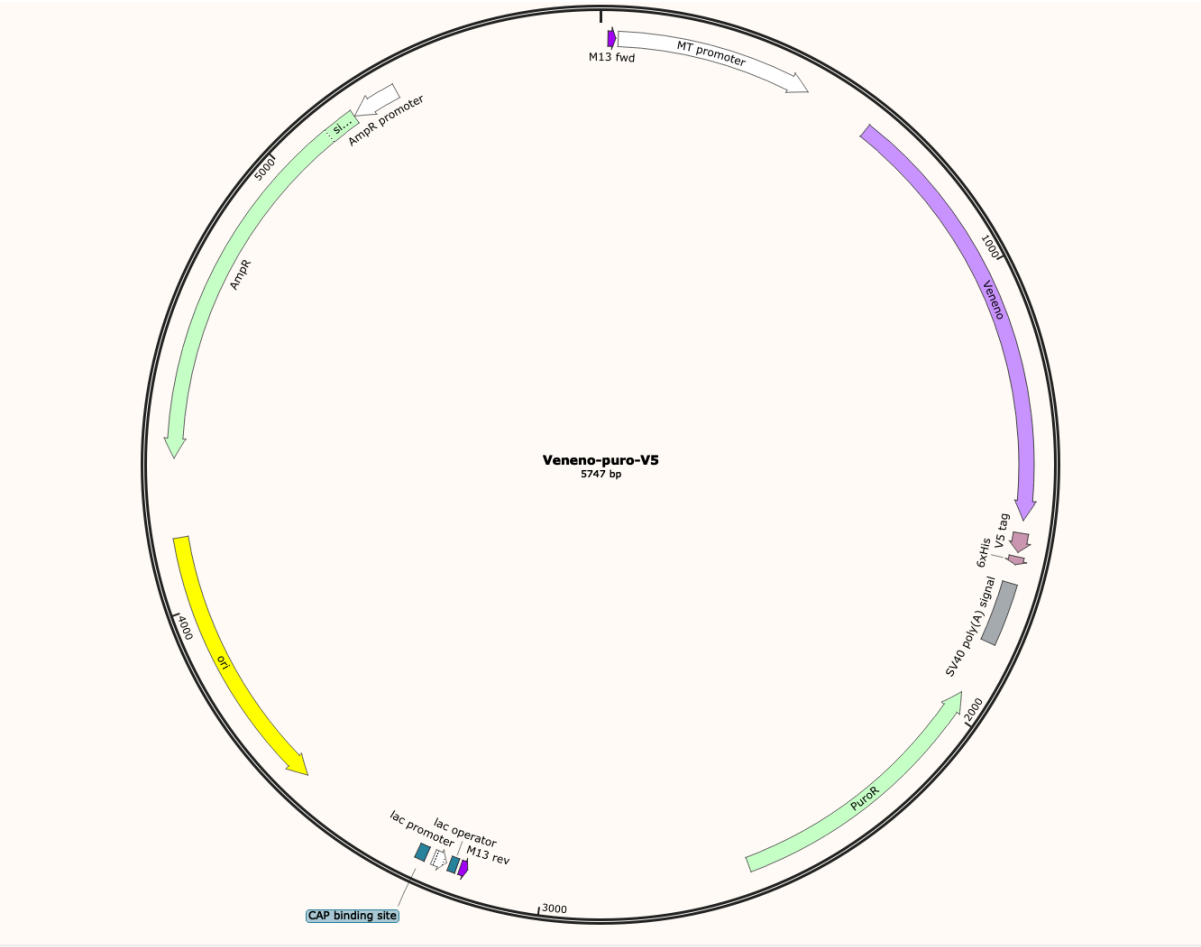


Figure 1.19 pMT-Ven^{trunc} -V5-puro Expresses PuroR gene causing puromycin resistance, and *Ven^{trunc}* attached to the V5 tag under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder (see main methods) to insert *Ven^{trunc}* in three fragments which we amplified from DGRP_362 DNA using primers in appendix 7.
(Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

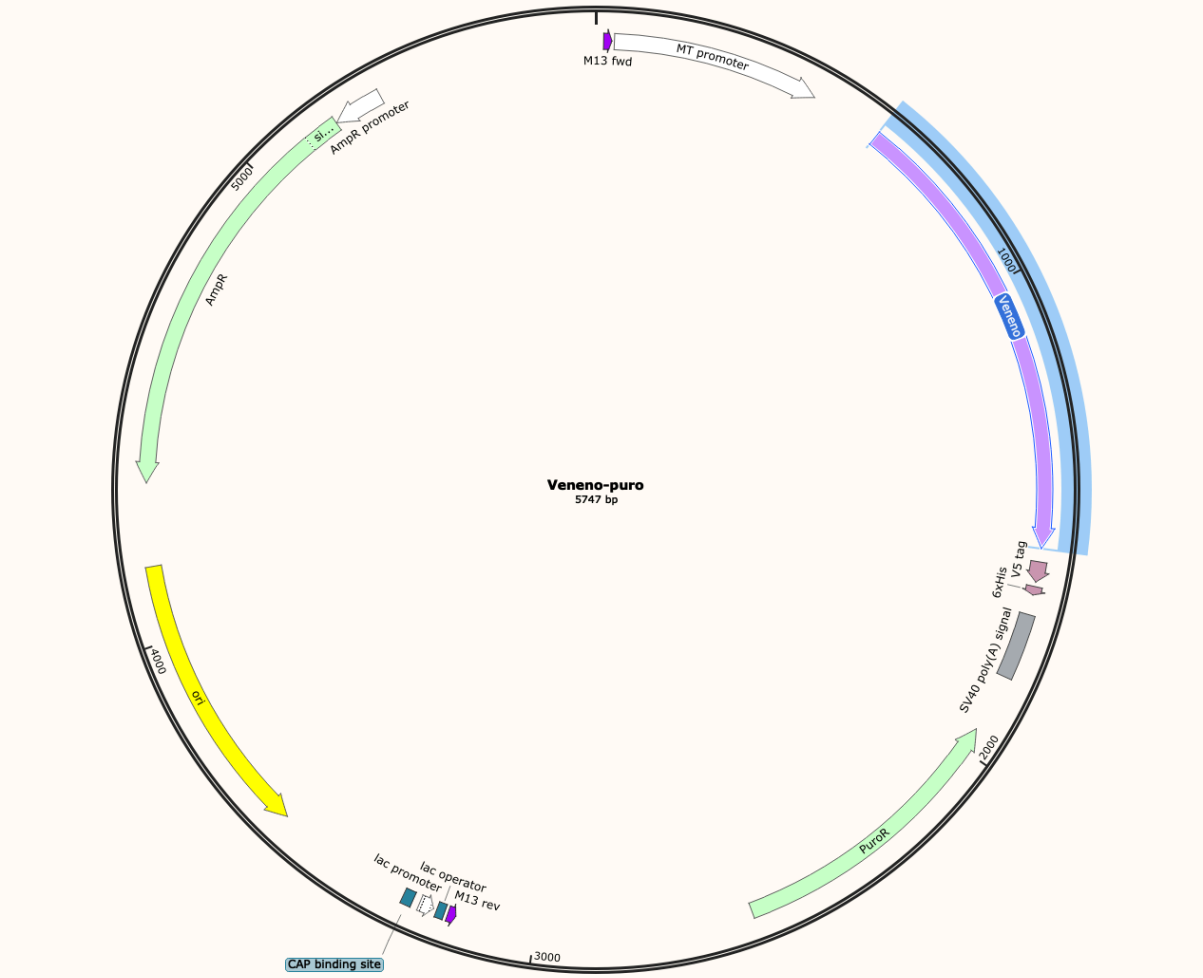


Figure 1.20 pMT-Ven^{trunc} -puro Expresses PuroR gene causing puromycin resistance, and *Ven^{trunc}* under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder (see main methods) to insert *Ven^{trunc}* with an alteration resulting in a stop codon at the end of the *Ven^{trunc}* sequence before the V5 tag, in three fragments which we amplified from DGRP_362 DNA using primers in appendix 6.
(Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

Infecting Cells Expressing *Ven^{trunc}*

We infected with DAV (see main methods) DL2 cells transfected with pMT-*Ven^{trunc}* -puro and DL2 cells transfected with pMT-puro (as control) with cells at a 1.5x10^5 cells/ml density and ~3.16*$TCID_{50}$ per 100µl well. 3 days later we extracted the RNA using TRIzol extraction, reverse transcribed it, and used qPCR to measure viral titre (see main methods).

Chapter 2 – Mechanism of DAV Resistance Caused by *Ven^{Doc}* Allele

Introduction

In Chapter 1 we showed that a *Doc* transposable element insertion into the coding sequence of *Veneno* results in increased resistance to Drosophila A Virus. In this chapter, we will work to characterize how this insertion causes that phenotype. In general, transposable element insertions are detrimental to host fitness (Houle & Nuzhdin, 2004; Pasyukova et al., 2004); however, in this case it results in a beneficial trait. Whether it does so by altering the protein product of the gene or by altering its expression patterns or by some other mechanism can indicate different things about how transposable elements affect host evolution. It is therefore important to understand how this occurs to further our understanding of the role of transposable elements in host evolution.

The wild type *Veneno* gene is 2597 base pairs long, with the 4685 bp *Doc* insertion occurring 1290 bp from the 5' end in the *Ven^{Doc}* allele. A stop codon occurs 18 base pairs from the 5' end of the *Doc* insertion, effectively splitting the gene into two halves. The *Doc* insertion leads to the expression of a truncated transcript *Ven_tra,* which ends 137 bp downstream of the gene-TE junction, and encodes a novel restriction factor. We've shown in Chapter 1 that expressing the translated region of this transcript is sufficient to induce DAV resistance in *Drosophila* cells.

Veneno contains two instances of the Tudor domain, which has a protein-protein interaction function, and is involved in the piRNA pathway (Lasko, 2010). The truncated transcript *Ven_tra* only contains sequences for the first of the two Tudor domains. Although there is no antiviral piRNA pathway in *Drosophila* (Petit et al., 2016), Joosten et al. have found that in *Aedes aegypti* the orthologue of *Veneno* is involved in antiviral piRNA biogenesis (Joosten et al., 2018). The *Aedes Veneno* codes for two Tudor domains which facilitate its interaction with Ago3, Vasa and Yb; together they form a protein complex, which also includes Piwi5, thereby allowing for ping-pong amplification of viral piRNAs (Joosten et al., 2018). The binding of Veneno to Ago3 is

dependent on symmetric dimethyl arginine (sDMA) recognition by the Tudor domain (Joosten et al., 2018).

In addition to the Tudor domains, *Veneno* also contains an MYND-type Zinc finger. Zinc fingers have been implicated in protein, nucleic acid, and lipid binding (Matthews & Sunde, 2002). In *Aedes aegypti*, the Zn-finger in Veneno plays an important role in localizing the protein to Ven-bodies, which are cytoplasmic foci where the ping-pong amplification of vpiRNAs takes place, and which are similar to nuages in *Drosophila* (Joosten et al., 2018). This is particularly interesting since the ORF1 of the *Doc* element codes for a protein which also localizes to clusters in the cytoplasm (Rashkova et al., 2002).

In addition, *Veneno* in *Drosophila* has been shown to interact with *Hen1* (which stabilizes miRNAs, siRNAs, and piRNAs by methylating them) in a high throughput study which used co-affinity purification and mass spectrometry to find a protein complex map for *Drosophila* (Guruharsha et al., 2011). It has also been shown to interact with *R2D2* in another co-affinity purification and mass spectrometry study on the siRNA pathway (Majzoub, 2013) which is involved in the siRNA pathway. The siRNA pathway is the main antiviral pathway in *Drosophila* which makes interactions between it and *Veneno* particularly interesting.

The high throughput tissue expression data from the modENCODE project show that *Veneno* is most highly expressed in the ovaries of 4-day old adult virgin female flies (Brown et al., 2014; Graveley et al., 2011). This is confirmed by high throughput anatomical expression data from the FlyAtlas project, which shows that the greatest expression of *Veneno* is in the adult ovary (Chintapalli et al., 2007). This indicates that wild type *Veneno* might have a function in the ovary. Additionally, temporal expression data from the modENCODE project show a peak of *Veneno* expression in the first two hours of the embryonic stage, and on day 5 in the adult female (Brown et al., 2014; Graveley et al., 2011), suggesting a possible role in early development and supporting the idea of a role in the ovary. Proteomic analysis has also detected the *Veneno* protein in the spliceosomal B and C complexes as well as affinity selected mRNPs, indicating a possible role it has in mRNA splicing (Herold et al., 2009).

We found that *Ven*^{Doc} has a lower expression level than *Ven*$^{+}$ in the abdomen of females, but similar levels elsewhere. We also showed that neither the 18 base pairs of *Doc* element origin in the translated sequence of *Ven_tra* nor the Tudor and MYND Zn finger domains are necessary for the antiviral function. We find no evidence for the involvement of the siRNA pathway nor any of several immune pathways including piRNA in the antiviral function of *Ven*^{Doc}.

Results

Ven^{Doc} has lower expression level than *Ven*$^{+}$ in female abdomens

To test if *Ven*^{Doc} is expressed differently from *Ven*$^{+}$, we checked for the expression of *Veneno* in the thorax and abdomen of DAV infected flies in both resistant and susceptible lines. We found that in the male thorax and abdomen and in the female thorax there is no significant difference in *Veneno* expression between the two alleles (Figure 2.1). However, in the abdomen of females, *Veneno* expression is significantly lower in the resistant line than in the susceptible line (Figure 2.1, Welch's t-test, t=-6.41, p=0.00162, df=6.82, p value adjusted using FDR with n=4). In our experiments, we have used females to test for viral resistance; however, previous work has shown that the resistant allele has the same effect in males (Fabian, unpublished data).

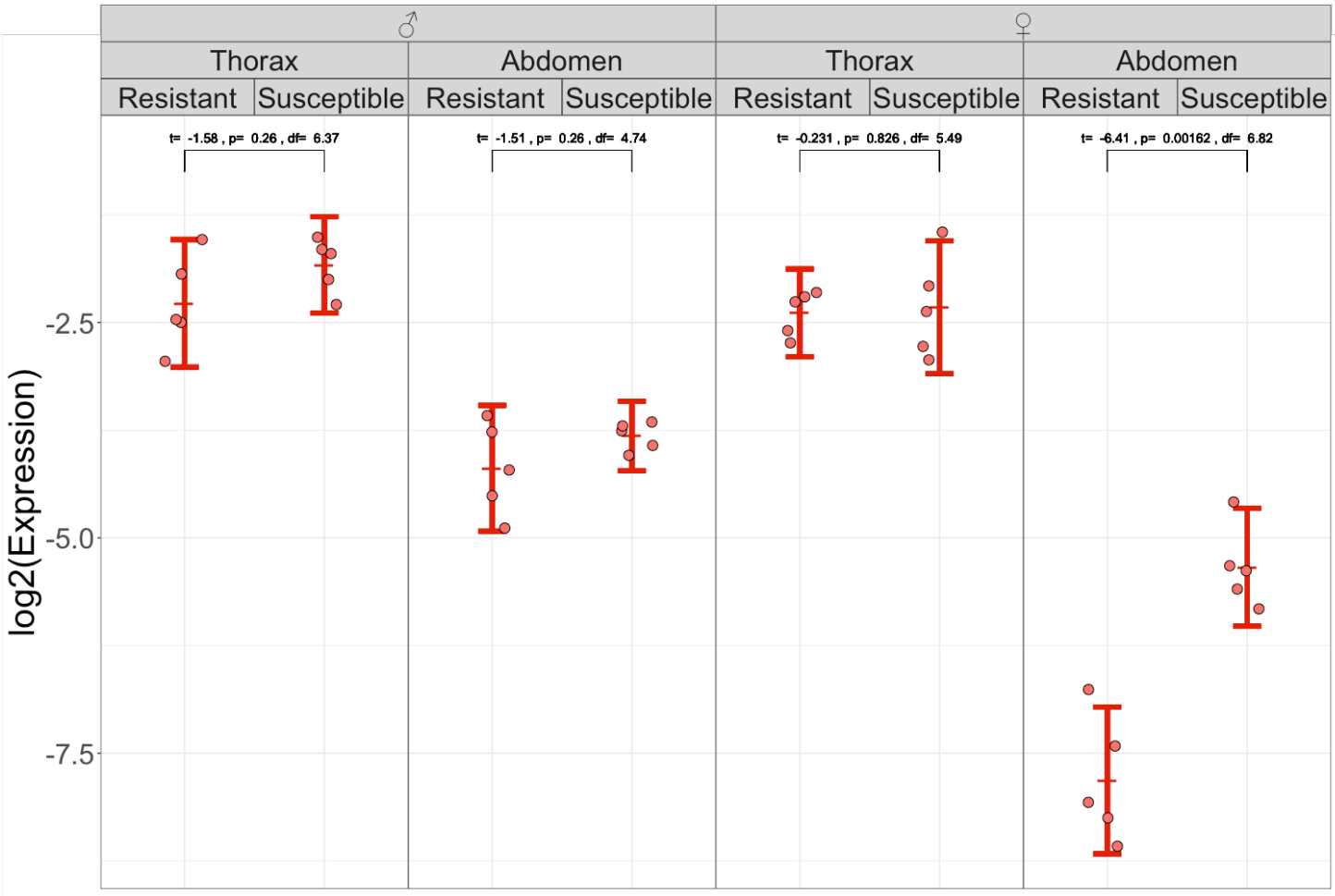
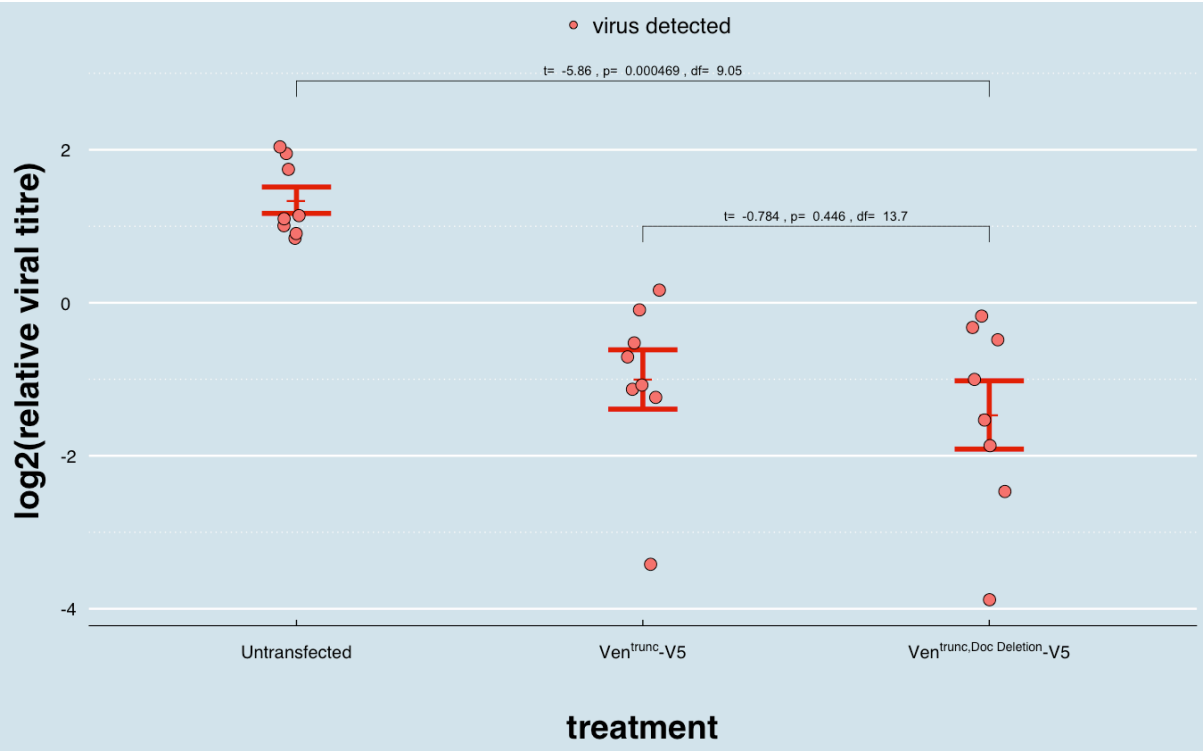


Figure 2.1 Expression of *Veneno* in the abdomen of DAV infected female flies is reduced by the *Doc* insertion. Flies having the resistant (Ven^{Doc}), and susceptible (Ven^{+}) alleles of *Veneno* were infected with DAV, and dissected 3 days post infection to test *Ven* expression in the different tissues. Y-axis shows the expression level relative to Actin5c calculated using -ΔCT values. Each dot represents tissue from 1 infected fly. Red bars show the standard error of -ΔCT centred around the sample mean. P-values calculated using Welch's t-test, and adjusted for multiple tests using false discovery rate (FDR) method.

Veneno Sequences of *Doc* element origin are not necessary for antiviral function

Ven_tra contains 18 base pairs which are of *Doc* element origin before the stop codon. To test whether the sequences of *Doc* element origin in the *Ven_tra* were necessary for the DAV resistance phenotype, we created a stably transfected cell line expressing the Ven^{trunc} without the *Doc* element sequence. To do so, we used Q5 site directed mutagenesis to excise the *Doc* sequence from a plasmid expressing Ven^{trunc}-V5, then stably transfected DL2 cells with both the mutated and unmutated plasmids. To ensure all of our cells are transfected, we cloned single cells from the transfected population. We used fluorescent staining of the V5 tag and flow cytometry (Appendix 21) as well as qPCR (in DAV infected cells, Appendix 22) to test that Ven^{trunc} is being expressed by the cloned population. We infected our cell lines with DAV and measured viral titre 3 days post infection using qPCR to test the effect of the deletion on the resistant phenotype (Figure 2.2). To confirm the Q5 mutagenesis and cloning process did not alter resistance, we repeated this process with a synonymous mutation PremCont (See Appendix 20).

We found that cells expressing $Ven^{trunc,\ Doc\ Deletion}$-V5 had a viral titre that was significantly lower than that of the untransfected line (Figure 2.2, Welch's t-test, t=-5.86, p=0.000469, df=9.05, p value adjusted using FDR with n=2), and quite similar to that of the cells expressing Ven^{trunc} (Welch's t-test, t=-0.7846, p=0.446, df=13.7, p value adjusted using FDR with n=2). This indicates that the sequence in *Ven_tra* of *Doc* element origin is not required for DAV resistance, and that the sequence of host origin is sufficient to cause that phenotype.



B

Figure 2.2 *Ven*^{trunc} results in lower DAV titre in DL2 cells despite deleting *Doc* Insertion sequence
Cells with and without a deletion of the *Doc* element sequence from *Ven*^{trunc} were infected with DAV. Viral titre was measured 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the mean. Expressing *Ven*$^{trunc,\ Doc\ Deletion}$ causes a significant decrease in viral titre from the control to a level similar to that of *Ven*^{trunc}, which indicates that sequences from the *Doc* element are not required to increase resistance in cells. P-values calculated using Welch's t-test with and FDR adjustment for multiple tests with n=2. Transfected cells express a transgene fused to a V5 tag and are clones of a single cell.

Neither the MYND Zn Finger Domain nor the Tudor Domain are Necessary for Antiviral Function

Veneno codes for 2 Tudor domains (which are likely involved in protein-protein interactions (Lasko, 2010)) and an MYND Zinc finger domain (which may be involved in protein-protein interaction and nucleic acid binding (Matthews & Sunde, 2002)). The MYND Zinc finger domain and one of the Tudor domains are encoded by *Ven_tra*, which suggests they may be involved in the antiviral function. One hypothesis is that *Veneno* has a protein scaffolding function which facilitates the formation of a protein complex (similar to the role it plays in *Aedes aegypti*), and that this role is affected (or induced) by the truncation caused by the *Doc* element.

To test whether this is the case, and whether the protein domains coded by *Ven_tra* are involved in resistance, we generated mutations in the plasmid expressing Ven^{trunc} –V5 using Q5 site directed mutagenesis. We targeted the Zinc Finger MYND Domain with the H51A which is similar to a mutation found to inactivate the domain by Liu et al. (Y. Liu et al., 2007). The Histidine which we targeted is required to chelate zinc, and is therefore required for correct function and folding of this domain (Y. Liu et al., 2007). We also targeted the Tudor domain with the Y198A mutations which is similar to mutations shown to inactivate the Tudor domain (Joosten et al., 2018; H. Liu et al., 2010). The tyrosine is involved in forming the binding pocket for the symmetric demethylation of arginine, which allows protein-protein interactions (Joosten et al., 2018; H. Liu et al., 2010). In addition we generated mutants with synonymous mutations in the same codons as our mutations in both domains (H51H and Y198Y) to act as controls. We then stably transfected DL2 cells with these plasmids. . To ensure all of our cells are transfected, we cloned single cells from the transfected population. We used fluorescent staining of the V5 tag and flow cytometry (Appendix 23) as well as qPCR (after DAV infection, Appendix 24) to test that Ven^{trunc} is being expressed by the cloned population. We infected our cell lines with DAV and measured viral titre 3 days post infection using qPCR to test the effect of the mutation on the resistant phenotype (Figure 2.3 B-C).

We found that both $Ven^{trunc,H51A}$ (Figure 2.3B Welch's t-test, t=-3.77, p=0.0491, df=2.35, p value adjusted using FDR with n=2) and $Ven^{trunc,\ Y198A}$ (Figure 2.3B Welch's t-test, t=-11.3, $p=6.56x10^{-7}$, df=10.5, p value adjusted using FDR with n=2) had viral titres which were significantly lower than that of the untransfected control, and similar to those of their respective controls. We then repeated the experiment without the Tudor Synonymous control and found similar results with both $Ven^{trunc,H51H}$ (Figure 2.3C Welch's t-test, t=-5.09, p=0.000961, df=7.95, p value adjusted using FDR with n=2) and $Ven^{trunc,\ Y198A}$ (Figure 2.3C Welch's t-test, t=-6.55, p=0.000242, df=8.74, p value adjusted using FDR with n=2) having viral titres which are significantly lower than that of the untransfected control. This indicates that these sites likely do not necessarily have to be functional in order for the resistance effect to take place.

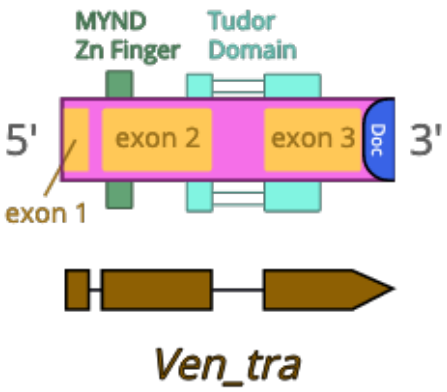
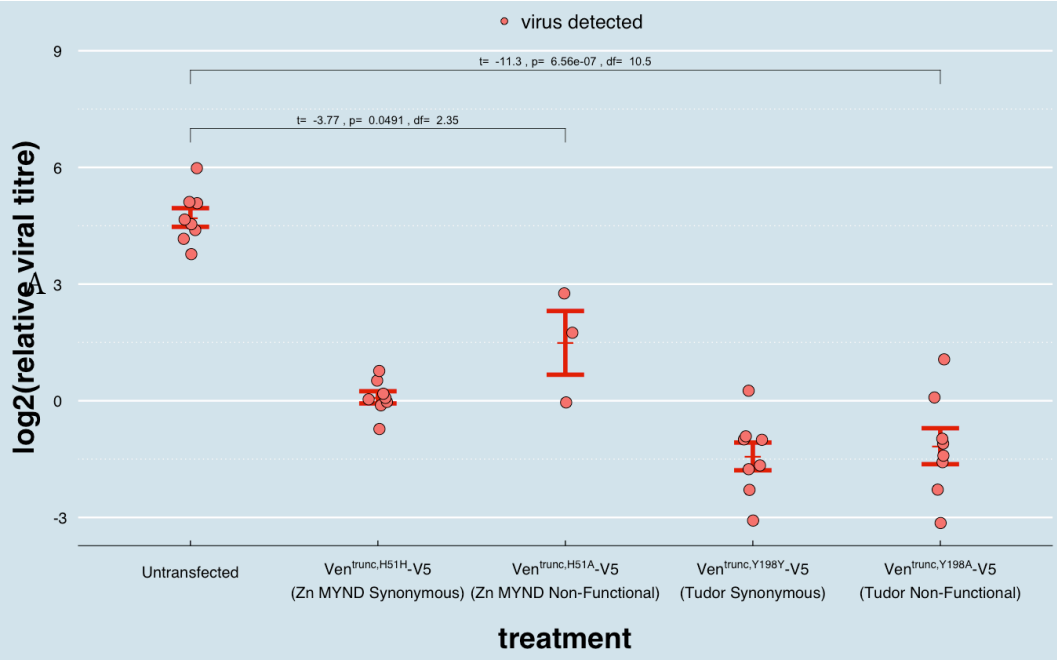


Figure 2.3. *Ven*^{trunc} results in lower DAV titre in DL2 cells despite mutations in Tudor and Zn MYND domains.

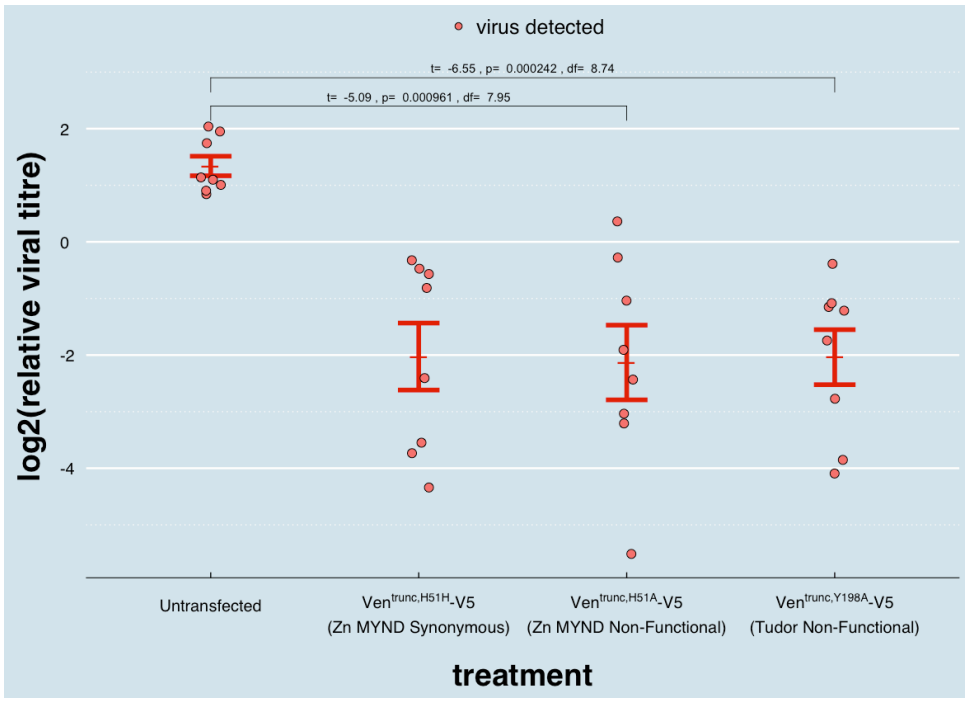
A) Diagram showing the protein domains in the region of *Veneno* expressed in *Ven*^{trunc}.

B,C) Cells with synonymous and non-synonymous mutations in the Tudor and Zn MYND domains of *Ven*^{trunc} were infected with DAV. Viral titre was measured 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the mean. pMT-Puro-*Ven*^{trunc} causes a significant decrease in viral titre from the control despite the mutations, which indicates that the mutations are not necessary increase resistance in cells. P-values calculated using Welch's t-test with and FDR adjustment for multiple tests with n=2. Transfected cells express a transgene fused to a V5 tag and are clones of a single cell.

A



B

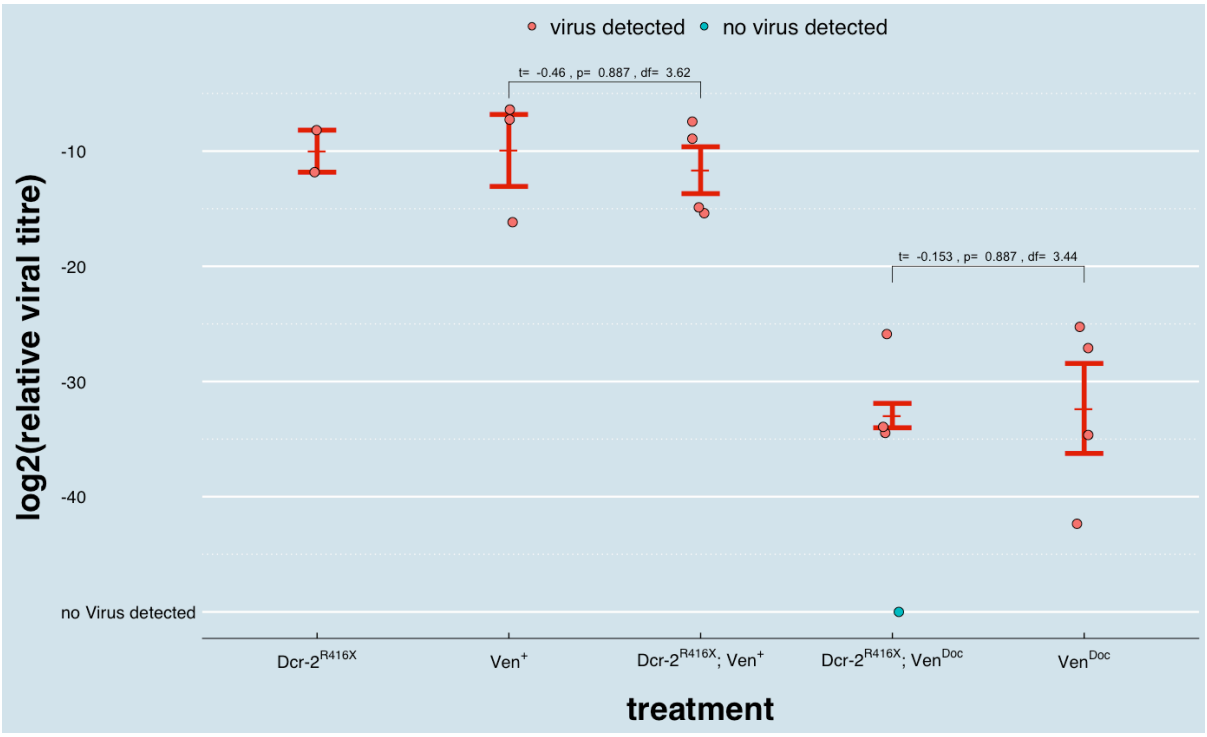


C

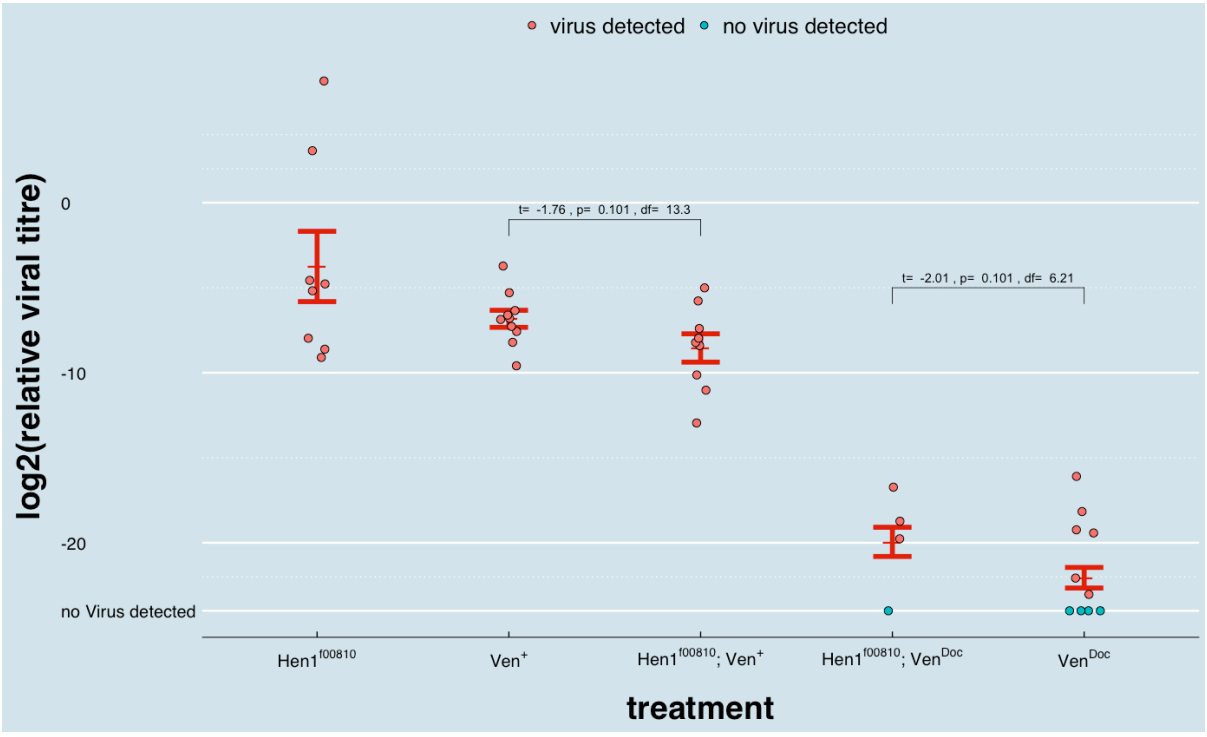
RNAi Pathway is Not Required for *Veneno's* Antiviral Function

Given *Veneno*'s interactions with Hen1 (Guruharsha et al., 2011) and R2D2 (Majzoub, 2013), as well as its Tudor domain which links it to small RNA pathways, we hypothesized that its resistance phenotype may require the RNAi pathway, the main antiviral pathway in *Drosophila.* To check whether that is the case, we tested whether the resistant phenotype caused by *Veneno* interacts with mutations in RNAi pathway genes. Using fly lines carrying mutations in RNAi pathway genes *Dicer-2* (Dcr-2^{R416X}) which is required for siRNA production and *Hen1* (Hen1^{f00810}) which stabilizes siRNAs by methylating them (provided by Maria Carla Saleh from the Institut Pasteur), we generated lines carrying each mutation and either *Ven^{Doc}* (DGRP line 362) or *Ven^{+}* (DGRP line 373 or 306).

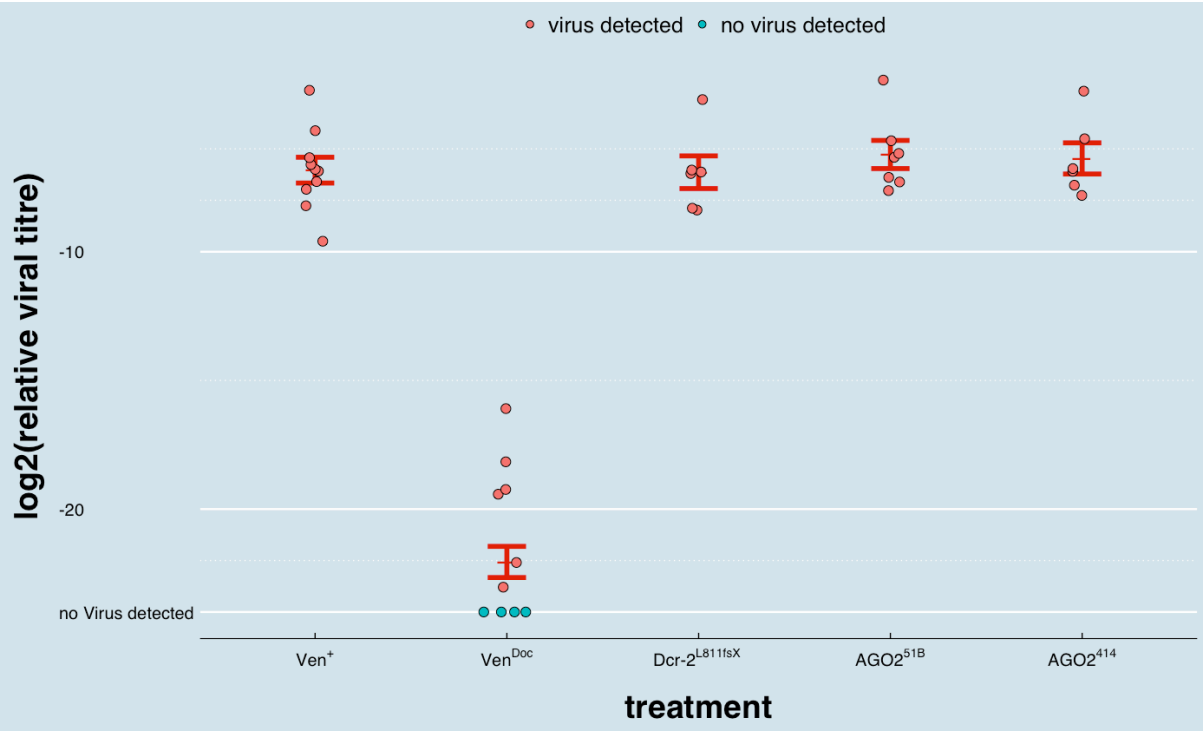
We froze the flies 24 ± 1 hours after infecting them with DAV, then we extracted their RNA, reverse transcribed it, and measured viral titre using qPCR (see main methods). We found that the differences in viral titre between *Ven^{+}* and *Ven^{Doc}* were not affected by the RNAi mutations, indicating that resistance caused by *Veneno* does not require the siRNA pathway (Figure 2.4A-B). Not only that, but we also found no significant influence of the RNAi mutations on DAV titre at all (Figure 2.4A-B). To confirm this, we infected flies from three further RNAi mutant lines: a second *Dicer-2* mutant Dcr-2^{L811fsx} and 2 mutants for gene *Ago*-2 (*Ago2^{51B}* and *Ago2^{414}*) which codes for the siRNA pathway protein which makes the cuts in the target RNA, and found no increased resistance in those either (Figure 2.4C). This indicated that not only is the resistance caused by *Ven^{Doc}* independent of the RNAi pathway, but that there is also no evidence the RNAi pathway is involved in DAV resistance at all.


virus detected
no virus detected
t= -0.46 , p= 0.887 , df= 3.62
t= -0.153 , p= 0.887 , df= 3.44
log2(relative viral titre)
-10
-20
-30
-40
no Virus detected
Dcr-2R416X
Ven+
Dcr-2R416X; Ven+
Dcr-2R416X; VenDoc
VenDoc
treatment

A


virus detected
no virus detected
t= -1.76 , p= 0.101 , df= 13.3
t= -2.01 , p= 0.101 , df= 6.21
log2(relative viral titre)
0
-10
-20
no Virus detected
Hen1f00810
Ven+
Hen1f00810; Ven+
Hen1f00810; VenDoc
VenDoc
treatment

B



C

Figure 2.4. RNAi pathway likely not involved in resistant phenotype. A,B) Flies of resistant (DGRP 362) and susceptible (DGRP 306/373) backgrounds with mutation in Dcr-2 (A) or Hen1 (B) genes were infected with DAV. There is no significant difference between lines with or without the mutation.
C) Flies with Dcr-2 and Ago-2 mutations were infected with DAV. There appears to be no increase in viral tire in mutant flies relative to unmutated flies with *Ven*$^{+}$ (wild type).
In all panels: Viral titre was measured 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents 7-15 flies. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution.

Immune Pathway Knockdown Test (ImPaKT)

To test if other immune pathways are required for the DAV resistance phenotype caused by *Ven*^{Doc}, we designed a panel of RNAi flies carrying targets for various genes of interest either involved in immune pathways or related to *Veneno* through various ways: see table 2.1.

We then generated flies which express a ubiquitously expressed Gal4 driver (actin promoter), and a UAS-driven siRNAs for each target. These flies had either *Ven*^{Doc}(DGRP 362) or *Ven*$^{+}$(DGRP 850) on their third chromosome. We infected these lines with DAV, then extracted their RNA three days later, reverse transcribed it, and measured viral titre with qPCR (see main methods).

With the exception of *Ge-1* and *Atg5*, we were unable to detect any significant effects of the RNAi on the resistance caused by the *Doc* insertion (Figure 2.5). With *Atg5* the low sample size (caused by low survivorship in flies with the knockdown) prevents us from making a definitive conclusion. The *Ge-1* data has a similar problem despite the higher sample size since the two data points for the resistant background suggest contradicting conclusions (with one having no viral replication at all and the other having similar levels of viral replication, Figure 2.5). Therefore, further investigation into these two genes is required before drawing a conclusion. We have therefore failed to find evidence that any of the tested genes and corresponding pathways are involved in the resistant phenotype caused by the *Doc* insertion.

Gene	BDSC#	Pathway	why it's there
Ago-2	34799	siRNA	The siRNA pathway is the main antiviral pathway in Drosophila. Additionally, R2D2 and Hen-1 associate with *Veneno*. (Guruharsha et al., 2011; Majzoub, 2013)
R2D2	34784	siRNA	
Hen-1	33385	siRNA	
Ago3	34815	piRNA	Joosten et al. found that the orthologue of *Veneno* in *Aedes aegypti* is involved in virus resistance in the mosquito through its interactions with the piRNA pathway including proteins Vasa and Ago3 (Joosten et al., 2018)
vasa	34950	piRNA	
vret	38212	piRNA	
Ge-1	32349		Cao et al. found a polymorphism in Ge-1 which affects DMelSV resistance (Cao et al., 2016)
Atg7	34369	autophagy	The autophagy pathway has an antiviral role: Lamiable et al. showed that mutating Atg7 led to increased VSV susceptibility (Lamiable et al., 2016). Nakamoto et al. showed that Toll-7 is involved in VSV recognition (Nakamoto et al., 2012). Shelly et al. showed that knocking down Atg5, Atg7, and Atg18 (and other autophagy genes) causes increased susceptibility to VSV (Shelly et al., 2009).
Toll-7	30488	autophagy	
Atg5	34899	autophagy	
Atg18	34714	autophagy	
Sting	31565	autophagy	
ref2p	36111	autophagy	*ref(2)p* has a polymorphism which confers resistance to DMelSV (Bangham, Kim, et al., 2008)
relish	33661	IMD	Costa 2009 showed that knocking down *relish* causes increased susceptibility to CrPV
DptB	28975	IMD	Huang 2013 showed that knocking down *DptB* causes increased susceptibility to SINV
Dif	29514	Toll	Zambon et al. presented evidence that *Dif* is involved in DXV resistance (Zambon et al., 2005)
toll	31477	Toll	Ferreira et al. showed that *toll* mutants are more susceptible to DCV (Ferreira et al., 2014)
p53	41720	apopotosis	*p53* mutation led to increased FHV susceptibility (B. Liu et al., 2013)
dfoxo	32427		Regulates siRNA; *dFOXO* mutation led to increased susceptibility to CrPV (Spellberg & Marr, 2015)
CG6283	51498		Interaction with *Ven*
CG9925	35811		Paralog of *Ven*

Table 2.1. Genes targeted by ImPaKT

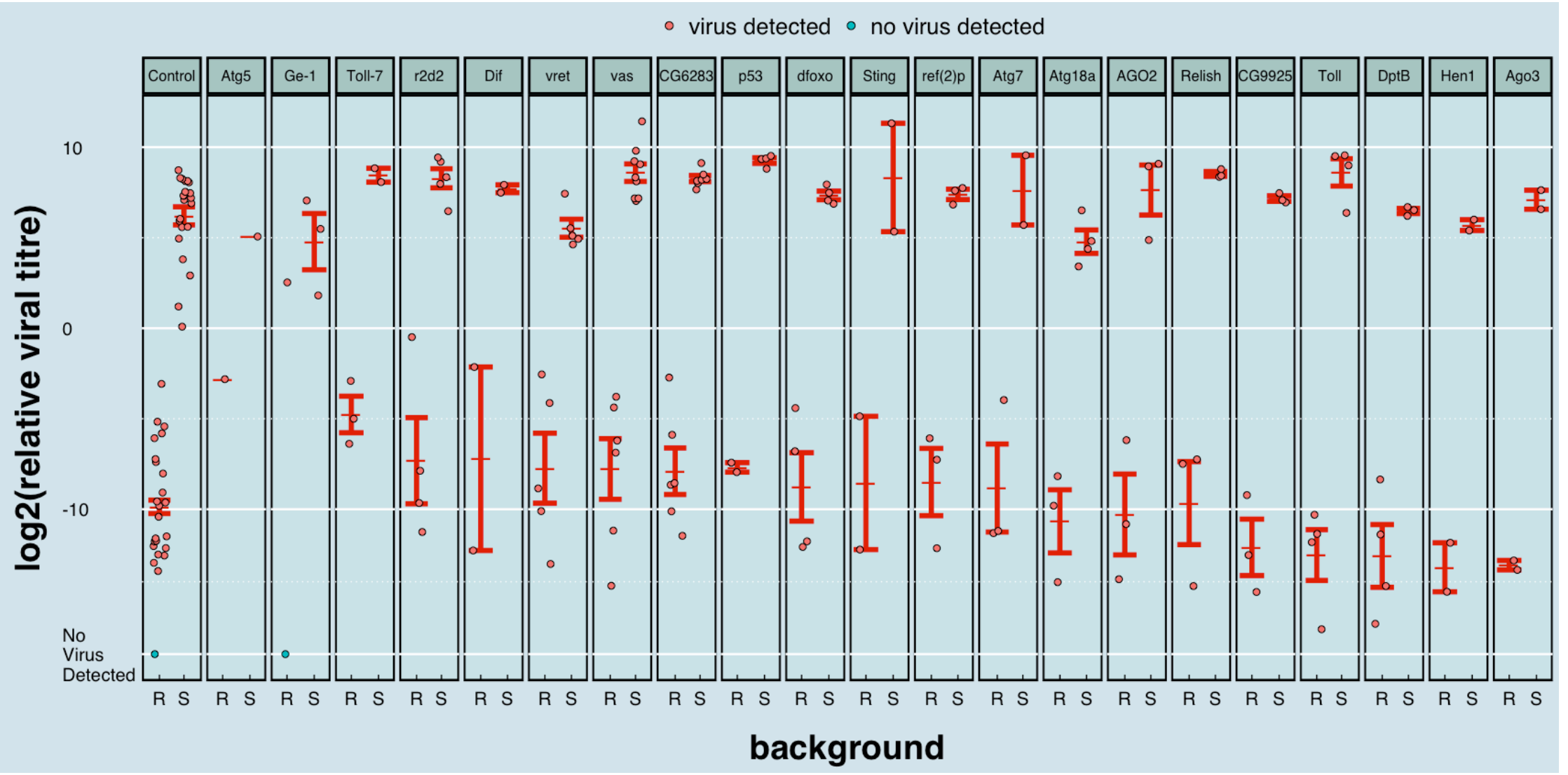


Figure 2.5. No evidence ImPaKT panel genes are involved in resistant phenotype. Flies of resistant (DGRP 362) and susceptible (DGRP850) backgrounds with RNAi targeting a panel of genes were infected with DAV. Viral titre was measured 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents 7-15 flies. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. It is not clear that any of the knockdowns resulted in a loss of the resistant phenotype in flies with resistant background.

Proteomics

To find the pathways *Ven*^{Doc} is involved in, we generated cell lines to look for the protein interactions involving *Ven*^{trunc} using Co-IP protein sequencing. We generated a cell line expressing line *Ven*^{sus}*-V5*, the RNA sequence of the wild type *Veneno* allele attached to a V5 tag to use as a control. As a baseline for interactions, we generated a line expressing *GFP-V5*, a GFP tag attached to the V5 tag. We infected these lines with DAV to check that they are indeed susceptible to DAV infection (Figure 2.6). The idea is to pull down *Ven*^{sus}*-V5* and *Ven*^{trunc}*-V5* using the V5 tag and sequencing the resulting protein samples using mass spectrometry to find which proteins *Ven*^{trunc} preferentially interacts with compared to *Ven*^{sus}. At the time of the thesis submission the Co-IP sequencing has not been completed.

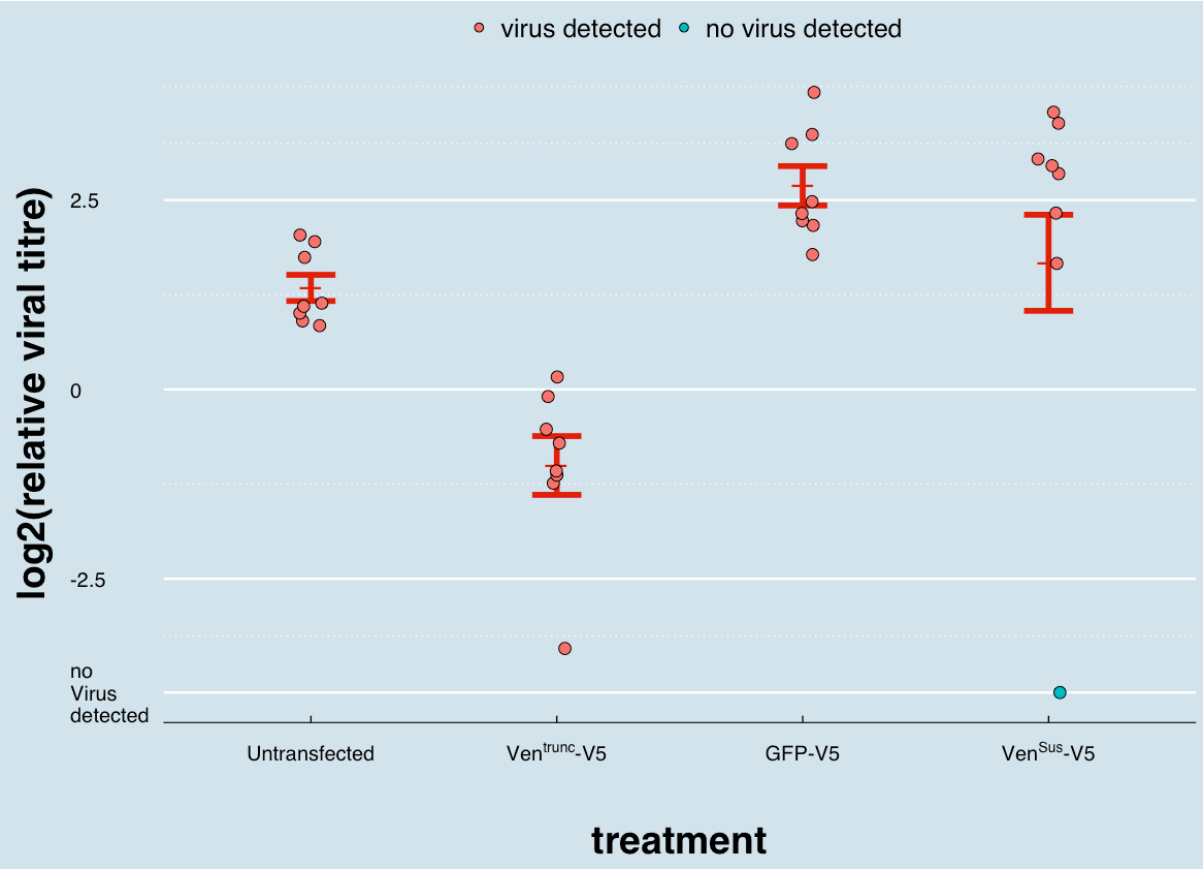


Figure 2.6. Cells transfected with pMT-Ven^{Sus}-V5-Puro and pMT-GFP-V5-Puro are not DAV resistant Cells transfected with pMT-*Ven*^{Sus}-V5-Puro and pMT-GFP-V5-Puro as well as untransfected and *Ven*^{trunc}-V5 expressing cells were infected with DAV. Viral titre was measured 3 days post infection. *Ven*^{Sus} and GFP do not cause increased resistance to DAV. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. Transfected cells express a transgene fused to a V5 tag. *Ven*^{trunc}-V5 line is a clone of a single cell.

Discussion

We have investigated how the *Doc* insertion into *Veneno* produces virus resistance. There are several ways a transposable element can cause novel phenotypes. Rostant et al. (2012) and ffrench-Constant (2006) mention a few, which we shall go through. A TE can insert into the 3' region of a gene, contributing to its mRNA stability, which is not applicable in the case of *Veneno* since *Doc* inserts in the middle of the sequence, not at its 3' end. A TE can excise a gene into a different genomic location affecting its expression patterns, but that is not the case for Ven^{Doc} since that *Doc* insertion does not contain sequences from other genes. A TE can result in a gene duplication due to ectopic recombination, which leads to the amplification of the gene, but again this is not the case since there are no duplications in the surrounding regions of Ven^{Doc}. Another possibility is that the TE can insert upstream or in the 5' end of a gene and alter its expression pattern such as the *Doc* inserting upstream of *Cyp6g1* which leads to DDT resistance (Schlenke & Begun, 2004). However, all these possibilities, which involve manipulating the expression of *Veneno* cannot be the mechanism for resistance since, as we will highlight in the next paragraph, it is unlikely that the *Doc* element causes resistance by altering the expression patterns of Ven^{Doc}.

We have found that expression of Ven^{Doc} is lower than that of Ven^{+} in the abdomen of females, but not significantly different in the female thorax, the male abdomen, or the male thorax. This may be due to the anti-TE piRNA pathway, which is active in the germ line as well as OSS cells, targeting *Ven_tra* or preventing the transcription of Ven^{Doc} which includes some transposable element sequence. More experiments need to be done in order to determine whether that is the case; in particular, measurement of Ven^{Doc} expression in flies with deactivated piRNA pathways. Our findings show it is unlikely that the *Doc* element results in DAV resistance by altering expression patterns. The only effect on expression the *Doc* insertion has is that it causes lowered expression in the abdomen of females. However, that is unlikely to cause resistance. To start with, DAV resistance induced in both males and females by the *Doc* insertion, so simply reducing expression in females is not enough to cause that resistance. Furthermore, we have shown in Chapter 1 that knocking down *Ven_tra* results in the loss of resistance, suggesting that lowered expression does

not induce resistance. Finally, we have also shown in Chapter 1 that expressing Ven^{trunc} in cells already expressing Ven^{+} is sufficient to induce resistance, which is strong evidence that lowered expression is not the mechanism causing resistance. We can therefore discount all the possibilities laid out by Rostant et al. and ffrench-Constant et al. which say resistance is caused by altering the expression patterns of Ven^{Doc}.

A final mechanism mentioned by (Rostant et al., 2012) and (ffrench-Constant et al., 2006) is that a TE inserting into the coding sequence of a gene can alter its transcript sequence. This could either result in a gene-TE chimera or a truncated transcript sequence. The gene-TE chimera is an interesting possibility as there are examples in nature of TEs being recruited to host function by being incorporated into the coding sequence of a gene, which results in such chimeras such as *SETMAR* in primates which is a chimera of the gene *SET* and the TE *Hsmar1* (Cordaux et al., 2006) which retains the function of the TE domain (D. Liu et al., 2007), and combines it with that of the gene to result in a novel function (Tellier & Chalmers, 2019). Another possible example of this, which we highlight in Chapter 3, is $CHKov1^{Doc1420}$ which expresses multiple transcripts encoding a varying number of *Doc* element base pairs and causes DMelSV resistance. Although we do not yet have evidence that the *Doc* element sequence is required in for the resistant phenotype produced by $CHKov1^{Doc1420}$ (a future experiment will investigate this), this brings up an interesting possibility: it may be possible for *Doc* element sequences to be recruited into taking part of a new antiviral function by being expressed as a part of a larger coding sequence in a gene-TE chimera. However, we have shown that this is not the case for Ven^{Doc}. To start with, only a small part of the *Doc* element in *Ven* appears to be transcribed, and a stop codon occurs 18 base pairs after the gene-TE junction, resulting in a transcript containing very little of the *Doc* element. But we also have strong evidence that sequences of *Doc* element origin in *Ven_tra* are not required for DAV resistance, since resistance is maintained even when those sequences are deleted. We can therefore discount the gene-TE chimera hypothesis.

This leaves us with the hypothesis resistance is caused simply by the truncation in *Veneno* caused by *Doc* prematurely terminating the protein. This appears to be the case here since we have shown

that expressing a truncated transcript which does not include any TE sequences is sufficient to cause DAV resistance in cells. There have been reports about TEs causing truncations in transcripts by introducing stop codons into the coding sequence. One such is example is the *Hel-1* LTR retrotransposon insertion into the HevCalP coding sequence in *Heliothis virescens,* which contains a stop codon 30 bases into its first LTR, resulting in a truncated transcript and predicted protein (Gahan et al., 2001). The allele produced by this insertion is resistant to *Bacilus thuringiensis* (Bt) toxins which are used as pest control; however, this resistance is recessive since the truncation results in a loss of function of the host protein (Gahan et al., 2001). What is particularly interesting about the *Doc* insertion in Ven^{Doc} is that it results in a *gain* of function by truncating the gene.

Since Ven^{Doc}'s phenotype does not rely on the *Doc* element sequence, the similarities between it and $CHKov^{Doc1420}$, and between the phenotypes that they generate, appear to be superficial; but are they coincidental? We can argue that the same evolutionary forces which have led to the *CHKov1* insertion have led to the *Veneno* insertion. The prevalence of viral infection in a population leads to there being a fitness advantage in evolving a mechanism of resistance, especially since viruses can evolve more rapidly than hosts, leading to the need for new host adaptations to counteract those of the virus. As a source of mutations, transposable elements can sometimes result in changes to resistance which can, by chance, be beneficial. Future studies should investigate whether this is due to random insertion in or near genes with a potential antiviral sequence, or transposable elements preferentially inserting in or near such genes. Some have argued that transposition may be induced in hosts undergoing stress from novel environments (Schrader & Schmitz, 2019; Stapley et al., 2015), and future work should investigate whether this occurs when viruses become prevalent in host population, or when a new insecticide is encountered by a population.

The *Doc* element leading to two different antiviral functions might be pure coincidence; however, future work should investigate whether *Doc* elements are more likely than other elements to cause antiviral mutations. This may be the case perhaps due to *Doc* elements 1) preferentially inserting into the coding sequences of genes (or simply avoiding coding sequences less than other TEs), 2) preferentially inserting into genes with potential antiviral functions, 3) containing sequences

which are more likely to result in truncations, such as a polyadenylation signal or other signals required to terminate transcription, or 4) simply because they currently have a high transposition rate compared to other elements. We have not found any evidence in the case of the *Doc* insertion in *Veneno* that the *Doc* element has a special antiviral property due to its sequence since the restriction factor coded by Ven^{Doc} does not rely on the *Doc* sequence to cause that resistance.

We have also investigated which of the protein interaction domains in *Ven_tra* are involved in the resistance phenotype. In *Aedes aegypti*, the orthologue of *Veneno* is involved in the antiviral piRNA pathway, in which it relies on its MYND Zn finger to localize to the nuage-like Ven-bodies, and on its Tudor domain to form a complex with Ago3 and Piwi5 (Joosten et al., 2018). We had hypothesised that *Ven_tra* might encode a protein with a similar scaffold or adaptor protein role in *Drosophila* which enable its antiviral function. However, our results show that in fact, the phenotype caused by *Ven_tra* does not rely on its MYND Zn or Tudor domains to be functional. Although a negative result, this in itself is quite an interesting result, since it shows that this restriction factor has a somehow different role in *Drosophila* than it does in *Aedes*. *Veneno* is a gene with an antiviral effect in *Aedes*, and which interacts with antiviral genes in *Drosophila*, but has no antiviral role unless around half of it is removed through truncation; when that happens, it gains an anti-viral role which is maintained despite inactivating its key domains. *Ven_tra* may interact with cellular pathways exploited by the virus, perhaps as a result of the gene's possible original role in RNA biology. However, it may also be the case that our mutations did not completely eliminate the functions that they are supposed to eliminate in theory, or that they eliminate some functionality of these domains, but not their entire functionality. For example, the tyrosine we targeted to deactivate the Tudor domain is involved in forming the binding pocket for the symmetric demethylation of arginine, which allows protein-protein interactions (Joosten et al., 2018; H. Liu et al., 2010). However, it is possible that the Tudor domain can interact with other proteins independently of the aromatic cage. Indeed, Webster et al. have shown that the Tudor domain in Krimper can bind to the protein Ago3 even when the latter has mutations in its arginine residues which are the target of symmetrical dimethylation and is therefore methylation deficient (A. Webster et al., 2015). In other words, the Tudor domain can interact with other proteins independently of sDMA binding, and therefore possibly independently of its aromatic cages. In

that case, it is possible that we did not eliminate this binding function by targeting the tyrosine. Nonetheless, our results provide insightful evidence into the functioning of this restriction factor.

We have also attempted to investigate which pathway *Ven^{Doc}* is involved in. Given its Tudor domains and predicted role, we hypothesised that *Veneno* is involved in small RNA pathways, an idea bolstered by its interactions with *Hen1* and *R2D2*, both of which are involved in the RNAi pathway (*Hen1* is also involved in the miRNA and piRNA pathways). Given that there is no antiviral piRNA pathway in *Drosophila* (Petit et al., 2016) , the prime candidate was the siRNA pathway (we also show below a knockdown of the piRNA pathway which also failed to show an interaction). However, we have found that the antiviral function of *Veneno* does not require the siRNA pathway, and no evidence the siRNA pathway is involved in DAV resistance at all.

To solve the mystery of which pathway *Veneno* is involved in, we cast a wide net by targeting multiple immune pathway genes (including piRNA) and potentially connected genes with the Immune Pathway Knockdown Test (ImPaKT). We failed to find any evidence that *Veneno* is connected to any of those. However, this shouldn't be taken as definitive evidence that *Ven^{Doc}* doesn't rely on any of those pathways, since RNA interference knockdowns may not completely eliminate gene function, and further studies with mutants should be done to provide stronger evidence for that claim. ImPaKT itself is a good tool for future study to quickly look for associations between a resistance gene and immune pathways, or to search for pathways providing resistance to certain pathogens. That latter application has already been successfully used in our lab to narrow down which pathways are involved in resistance to La Jolla Virus and New Field Virus (Bruner, unpublished data). Therefore, despite the negative result, the panel of flies which we've assembled does have effects with other viruses, and can be useful in future studies.

In order to narrow down which pathways may be involved in the phenotype of *Ven^{Doc}* . Its interaction partners could be investigated by sequencing with mass spectrometry of co-immunoprecipitated product. We have already begun that investigation, which, unfortunately, has not been completed in time for this thesis.

In the absence of evidence implicating *Ven*^{Doc} in any known immune pathways, it joins the ranks of restriction factors which provide virus resistance whose mechanisms are poorly understood. These include *Vago* (Deddouche et al., 2008), *ref(2)P* (Bangham et al., 2008), *CHKov1* (Magwire et al., 2011), *pastrel* (Magwire et al., 2012), and *Ge-1* (Cao et al., 2016) 2016). In the arms race between hosts and viruses, hosts have developed many novel alleles which seem to provide resistance to viruses by encoding restriction factors. These restriction factors seem to be specific to certain viruses, and their mechanisms are poorly understood. It is important to study how these restriction factors function, and whether they do so in a variety of ways, independent of any pathway, as that can have wide implications in understanding virus immunity in insects and even other species such as humans.

Methods

Measuring Expression of *Veneno*

To measure *Veneno* expression in the thorax, and abdomen, we dissected DAV infected resistant and susceptible (DGRP lines 362 and 306 respectively), male and female flies using forceps into the head, thorax, and abdomen, then extracted RNA from the thorax and abdomen sections using TRIzol extraction (see main methods) then reverse transcribed the RNA and measured expression using qPCR (see main methods, VenExpFR Primers in Appendix 19).

Q5 site directed mutagenesis

Mutating Plasmids

We created plasmids expressing versions of *Ven*^{trunc} with the mutations in table 2.2 by mutating the pMT-*Ven*^{Doc}-V5 -puro plasmid using the Q5 site directed mutagenesis kit (NEB) and the primers in appendix 8. We followed the manufacturer guidelines with a scaled down PCR reaction (to

10µl) for the exponential amplification: 5µl Q5 Hot Start High-Fidelity 2X Master Mix, 0.5µl of 10µM primer mix (forward and reverse), 0.4µl template DNA, and 4.1µl nfH_2O. (See cycle settings in table 2.3)

Mutation	**Mutation Description**
Doc Deletion	deletion of the *Doc* element sequence
ZnMut	nonsynonymous (H51A) mutation in the MYND Zn Finger
ZnCont	synonymous (H51H) mutation in MYND Zn Finger (control for ZnMut)
TudMut	nonsynonymous mutation (Y198A) in the Tudor domain
TudCont	synonymous mutation (Y198Y) in the Tudor domain (control for TudMut)
PremControl	synonymous mutation in codon 4 of Ven^{Doc}.

Table 2.2 Mutations made to Ven^{trunc} using Q5 site directed mutagenesis. A mutation similar to H51A was found to inactivate the MYND Zn finger domain (Y. Liu et al., 2007). Mutations similar to Y198A were found to inactivate the Tudor domain (Joosten et al., 2018; H. Liu et al., 2010). (Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

	Temperature	Time
1x	98°C	30 s
25x	98°C	10 s
	T_a*	20 s
	72°C	3 min
1x	72°C	2 mins

Table 2.3 Q5 site directed mutagenesis protocol *:Calculated for each primer pair using NEBaseChanger (http://nebasechanger.neb.com/)

We then added 1µl of the PCR product to 5µl of KLD Reaction Buffer, 1µl of KLD enzyme mix, and 3 µl of nfH_2O and incubated in room temperature for 5 minutes. We added 5µl of the mix to 50µl NEB® 5-alpha chemically competent cells which had been thawed on ice. We incubated the cells on ice for 30 minutes, heat shocked them at 42°C for 30 seconds, incubated them on ice for 5 more minutes, then added 950µl of room temperature SOC media (NEB) and incubated the mixture at 37°C in a shaking incubator at 150 RPM.

Since all the plasmids had Amp resistance, we plated the transformed bacteria onto Lysogeny broth agar (LB-agar) plates with 100µg/ml Ampicillin for selection of transformants carrying the plasmid, and incubated overnight at 37°C. We picked off colonies and verified the presence of an insert of correct size by running a PCR reaction targeted around the insertion site (pMT_For and BGH_Rev primers in Appendix 18) using the bacteria directly as a template, and running the product through a gel. We then further verified colonies which had the correct band size by Sanger sequencing the insert region of the plasmids (using pMT_For and BGH_Rev primers in Appendix 18). Samples with the correct insert were then grown further for freezing and for plasmid extraction. We then prepared the plasmids and transfected the cells following the same methods in the main methods section.

We cloned the mutant cell lines (and the unmutated pMT-*Ven*^{trunc} –V5-puro line) by following methods similar to (Zitzmann et al., 2018): In 96 well plates (Cellstar), we added 100µl of medium containing on average 1-1.5 transfected cells as well as $5x10^4$ untransfected cells (present to generate growth factors to allow the transfected cells to grow despite their low density). 2 days later, we added 30µl of medium with enough puromycin to bring the total concentration of puromycin to 10µg/ml. This selected for our transfected cells which have puromycin resistance. We allowed for growth of our colonies until they were visible by eye, and we confirmed them by microscopy before picking them by pipette and moving them to fresh medium in another 96 well plate. We subsequently the expression of the V5 tag in the selected cells via flow cytometry.

Infecting Cell Lines expressing mutated *Ven^{trunc}*

We infected with DAV (see main methods) untransfected DL2 cells and selected DL2 cells expressing the PremControl mutant. In a separate experiment we infected with DAV (see main method) untransfected DL2 cells and selected DL2 cells expressing the ZnControl, ZnMut, TudControl, and TudControl mutants. In a third experiment we infected with DAV (see main methods) untransfected DL2 cells and selected DL2 cells expressing the DocDeletion, ZnControl, ZnMut, and TudControl mutants as well as the unmutated pMT-*Ven^{trunc}* –V5-puro line. For all three experiments we infected the cells at a $1.5x10^{6}$ cells/ml density and ~6.3 x $TCID_{50}$ per 100µl well. 3 days later we extracted the RNA using TRIzol extraction, reverse transcribed it, and used qPCR to measure viral titre (see main methods).

siRNA pathway mutants

We combined third chromosomes carrying different alleles of *Ven* with mutants affecting the siRNA pathway on chr 2. Lines with mutations in siRNA pathway genes were kindly provided by Maria Carla Saleh from the Institut Pasteur. We used line ;*Dcr-2^{R416X}*; which has a Dcr-2 mutations on the second chromosome, and line ;*Hen1^{f00810}*; which has a Hen1 mutation on its second chromosome. Using the mutant lines, DGRP lines 362 (*Ven^{Doc}*), 306 (*Ven^{+}*), and 373 (*Ven^{+}*), and double balancer lines, we generated the lines ;*Dcr-2^{R416X}*; DGRP_373, ;*Dcr-2^{R416X}*; DGRP_362, ;*Hen1^{f00810}*; DGRP_306, and ;*Hen1^{f00810}*; DGRP_362 using the following crossing scheme, and then we infected those lines with DAV, froze them after 24±1 hours, then extracted their RNA with TRIzol, reverse transcribed it, and tested it for DAV infection using qPCR (see main methods).

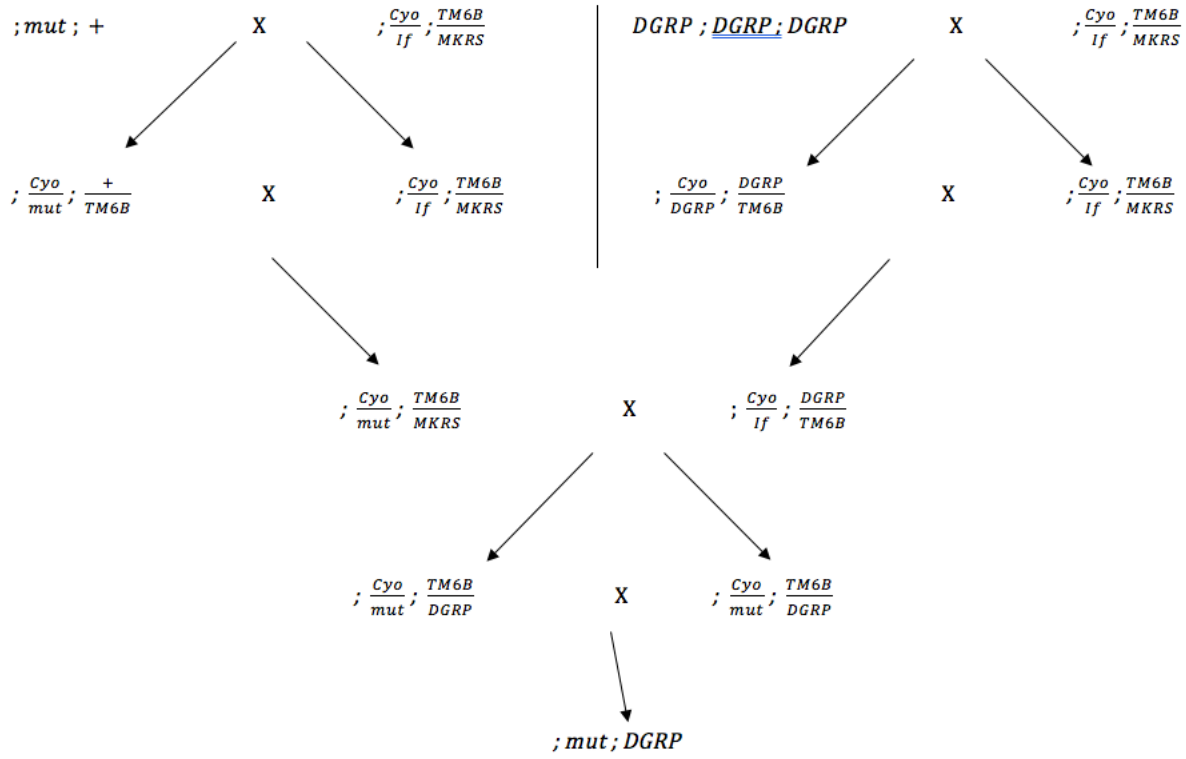


Figure 2.7. Crossing scheme for siRNA Pathway Mutants. We used these crosses to produce ;Dcr-2^{R416X}; DGRP_373, ;Dcr-2^{R416X}; DGRP_362, ;Hen1^{f00810}; DGRP_306, and ;Hen1^{f00810}; DGRP_362

Immune Pathway Knockdown Test (ImPaKT)

To carry out the Immune Pathway Knockdown Test (ImPaKT), we selected a panel of Bloomington fly stocks which carried TriP RNAi (Zirin et al., 2020) constructs targeting genes of interest. We selected genes which are a part of an immune pathway or genes which interact with *Veneno*. Using a fly line which expresses *Actin5c* driven Gal4 (provided by Alex Whitworth), w*;Cyo/If;DGRP lines (either 362 or 850) which had been created for another experiment and double balancers, we created lines w*;Act-gal4/UAS-TriP;DGRP/+ or w*;Act-gal4/+;DGRP/UAS-TriP (depending on whether the TriP-RNAi construct was on the second or third chromosome) according to the crosses in Figure 2.8. We reared these flies at 25°C, and infected them with DAV then froze them after 72 ± 3 hours extracted their RNA with TRIzol, reverse transcribed it, and tested it for DAV infection using qPCR (see main methods).

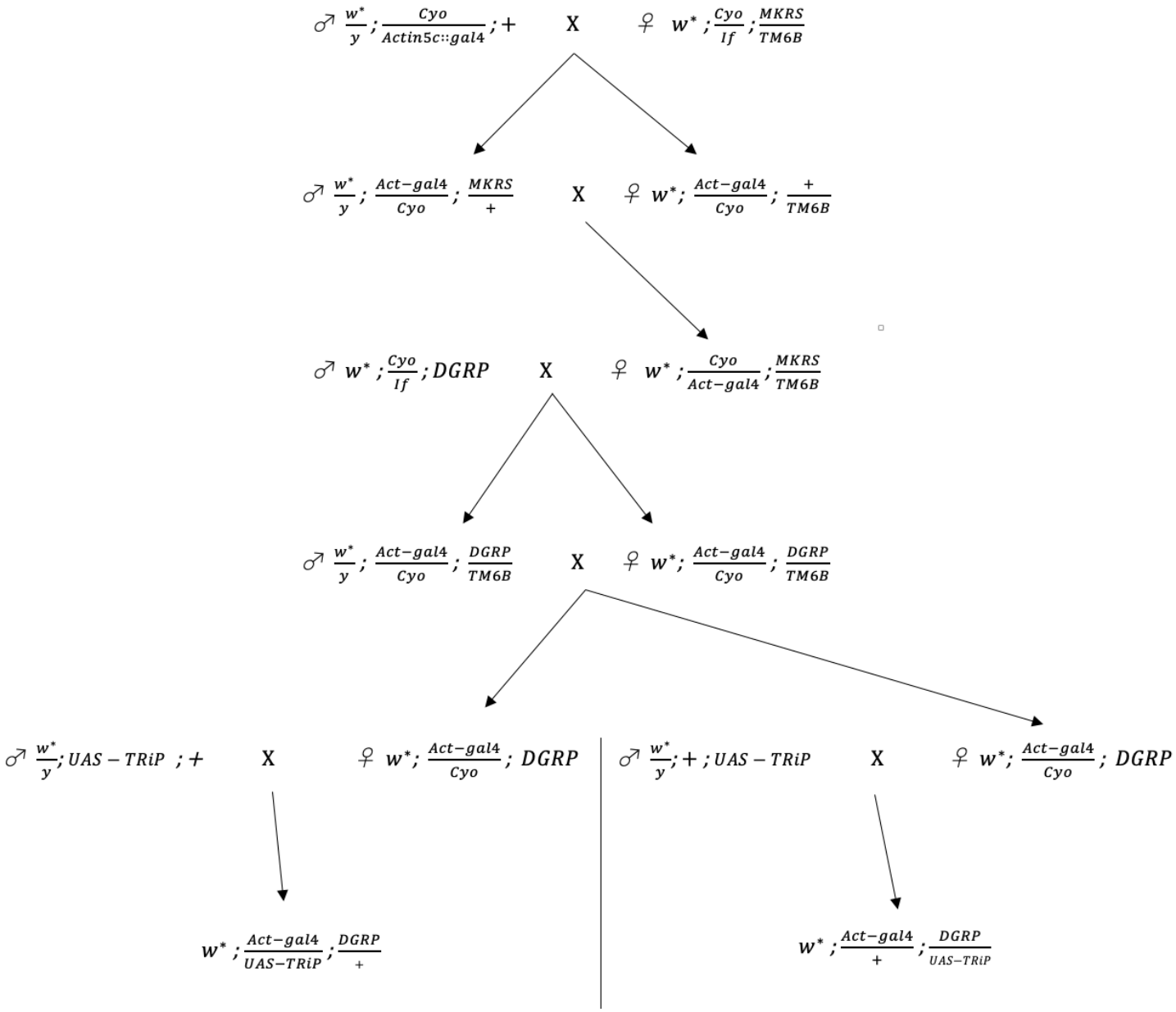


Figure 2.8. Crossing scheme for ImPaKT. We used these crosses to produce flies expressing ImPaKT knockdown constructs carrying *Ven*$^{+}$/*Ven*$^{+}$ (+/DGRP$_{850}$) or *Ven*$^{+}$/*Ven*^{Doc} (+/DGRP$_{362}$ or UAS-TriP/DGRP$_{362}$) on chromosome 3.

Proteomics

We generated the plasmids to generate cell lines to be used as controls for the proteomics experiment using the NEBuilder method described in the main methods section. pMT-*Ven*^{sus}-V5-

puro (Figure 2.9) was made using inserts *Ven*^{trunc} and VenCompleting, which we amplified from pMT-*Ven*^{trunc} -V5 -puro and DGRP_373 cDNA respectively using primers Ventrunc_FR and VenCompleting_FR in Appendix 9. pMT-GFP-V5 -puro (Figure 2.10) was made by inserting GFP, which we amplified from pUAST-attp-GFP (plasmid provided by Jinghua Gui (Karam-Teixeira lab)) using primers in appendix. We inserted the sequences into pMT-puro (addgene 17923, Figure 1.18) which has been digested with NotI-HF. We transfected cells with these plasmids.

To test that these cells are indeed susceptible, we infected with DAV (see main methods) untransfected DL2 cells, a resistant control (a clones expressing the pMT-*Ven*^{trunc} –V5-puro line) as well as these cells transfected with pMT-*Ven*^{sus}-V5-puro and pMT-GFP-V5 -puro at a 1.5x10^{6} cells/ml density and ~6.3 x $TCID_{50}$ per 100µl well. 3 days later we extracted the RNA using TRIzol extraction, reverse transcribed it, and used qPCR to measure viral titre (see main methods).

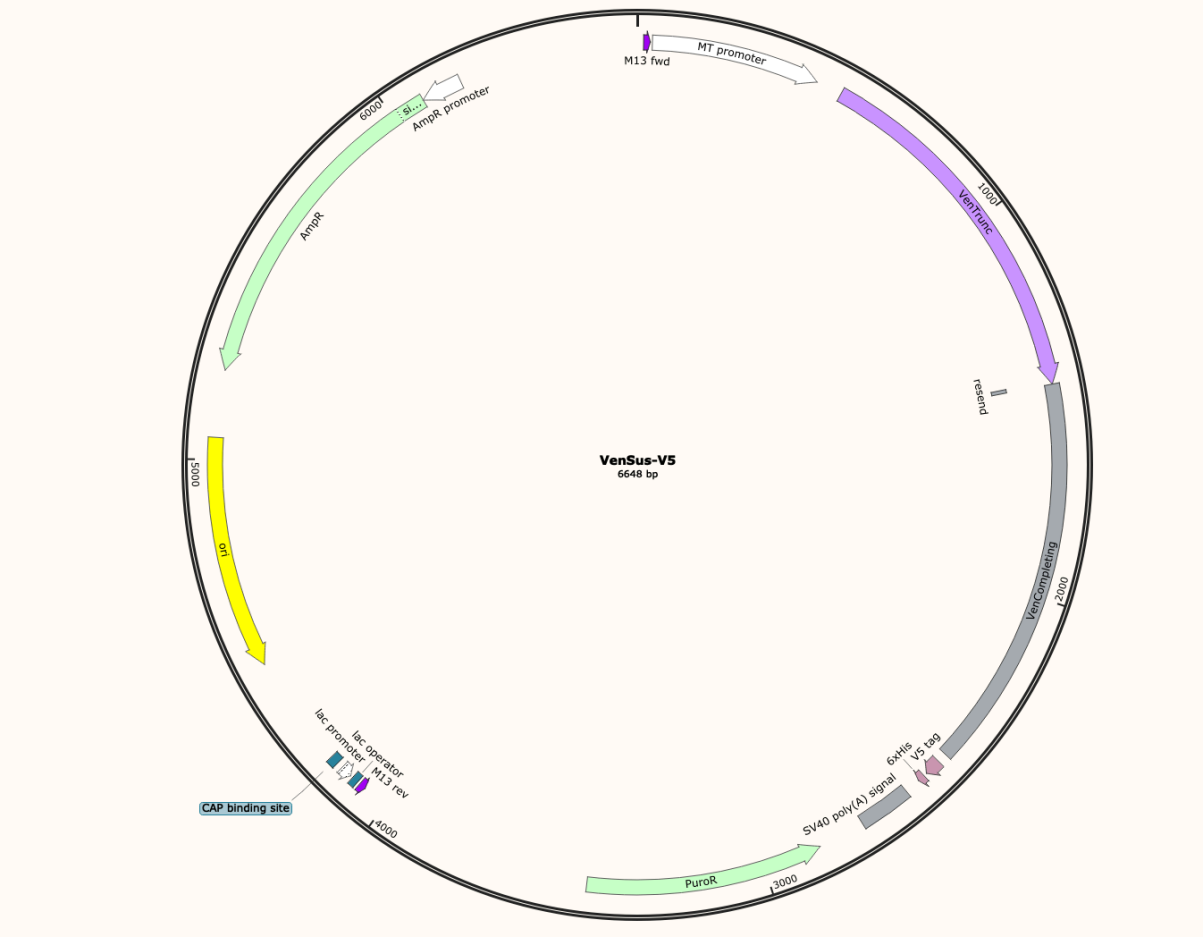


Figure 2.9 pMT-*Ven^{sus}*-V5 -puro Expresses *Ven^{sus}-V5*, the RNA sequence of the wild type *Veneno* allele without UTRs attached to a V5 tag under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder (see main methods) to insert *Ven^{sus}* in two fragments, *Ven^{trunc}* and VenCompleting, which we amplified from pMT-*Ven^{trunc}* -V5 -puro and DGRP_373 cDNA respectively using primers in appendix 9.
(Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

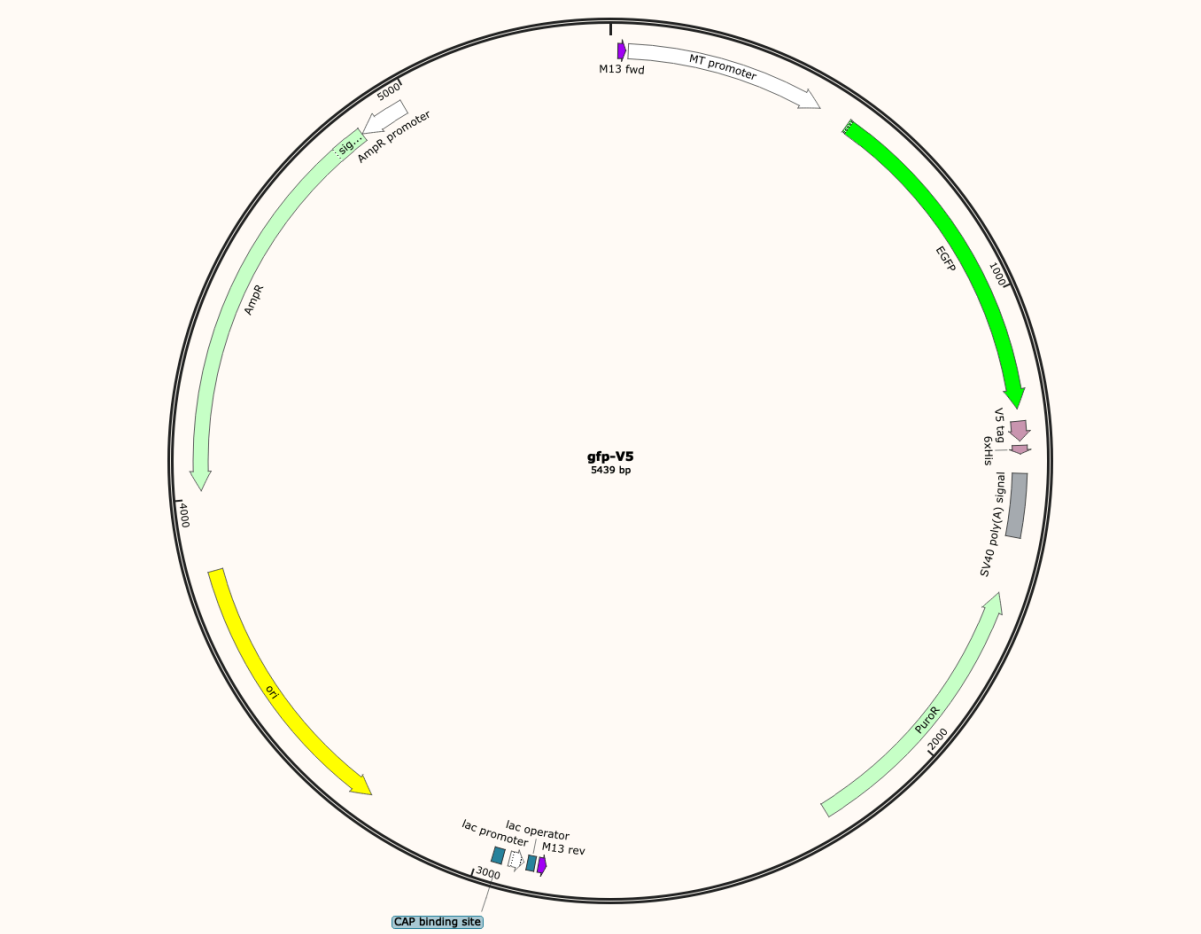


Figure 2.10 pMT-GFP-V5-puro As a control for proteomics, expresses *GFP-V5*, a GFP tag attached to the V5 tag under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder (see main methods) to insert GFP, which we amplified from pUAST-attp-GFP (plasmid provided by Jinghua Gui (Karam-Teixeira lab)) using primers in appendix 9.
(Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

Chapter 3- *Doc* element insertion in *CHKov1* Causes Resistance to DMelSV by Encoding Novel Chimeric Transcript

Introduction

In this chapter, we will discuss the case of a *Doc* element insertion in the coding sequence of *CHKov1*. Little is known about the function of the wild type *CHKov1.* The high throughput anatomical expression data from the FlyAtlas project shows that the greatest expression of *CHKov1* is in the adult salivary gland, which exhibits a high level of expression whereas all other tested categories had moderate, low, or no expression (Chintapalli et al., 2007). On the other hand, the high throughput tissue expression data from the modENCODE project show *CHKov1* to be most highly expressed in the head of a mated 20-day male with moderate expression; other tissues also had moderate expression such as the head of 1 and 4 day virgin females, the head of 1 day mated females, the head of 1 and 4 day mated males, the digestive system and the carcass of a 1 day adult, and the accessory gland of the 4 day mated male (though the adult salivary gland was not one of the tested categories) (Brown et al., 2014; Graveley et al., 2011). Though inconclusive, these expression data suggest the possibility that *CHKov1* might carry out a role in the salivary glands. Additionally, temporal expression data from the modENCODE project show a peak of *CHKov1* expression in the larva L3 puffstage 7-9 (late larval) and the white prepupa stage (early pupal) (Brown et al., 2014; Graveley et al., 2011), suggesting a role of *CHKov1* in pupation.

The *Doc* insertion in *CHKov1* has been shown to cause resistance to DMelSV (Magwire et al., 2011). The insertion is also thought to result in novel derived transcripts derived transcripts 1 and 2 (dtr1-2) (Aminetzach et al., 2005), although the exact nature of these transcripts is uncertain as these authors did not make their sequencing data publicly available. There are also contradicting

data on an anti-pesticide role played by this insertion with some evidence showing that it causes increased pesticide resistance (Aminetzach et al., 2005), and other experiments failing to replicate that finding (Battlay et al., 2016). The insertion is under positive selection which has led to it becoming common worldwide (Aminetzach et al., 2005), which could be explained by the fitness advantage to having DMelSV resistance or, if true, having pesticide resistance.

In a previous experiment, our group has shown that the viral resistance caused by the *Doc* element insertion is a novel function, which could be eliminated by knocking out a region of the host genome coding for dtr1, which is on the 5' side of the *Doc* element insertion (and not the region coding for dtr2, which is on the 3' side of the insertion). (Unpublished data). Here we extend this work, showing that expressing a transcript starting upstream of the *Doc* element and ending at the first stop codon within the *Doc* element (at the end of the first open reading frame in the *Doc* element) is sufficient in causing resistance to DMelSV.

Results

CHKov1^{Doc1420} Produces 4 New truncated Transcripts

In order to find the *CHKov1* transcript which might be responsible for the increased resistance caused by the *CHKov1*^{Doc1420} allele, we sequenced the transcriptome of $DGRP_{362}$ which carries the insertion. We extracted RNA from $DGRP_{362}$ flies and performed Oxford Nanopore sequencing on the polyadenylated RNA which we had reverse transcribed and amplified by PCR. The PCR-cDNA sequencing had a throughput of ~8.8 million mapped reads (SRA: SRR15541957). We mapped the sequences to the Drosophila melanogaster genome, then found likely *CHKov1* transcripts by clustering reads with similar exon/intron structures then generating consensus exon boundaries for each cluster using the pinfish pipeline. (Figure 3.1, tr1-4)

We identified four transcripts tr1 (18 reads), tr2 (5 reads), tr3 (5 reads), and tr4 (24 reads) (Figure 3.1, sequences on benchling: https://benchling.com/obrosh/f_/Zoz9jidS-thesis-sequences/). All transcripts

start at the beginning of the gene and express varying amounts of *Doc* element sequences with tr4 expressing the entire *Doc* element (up until the start of the run of As at the end of the insertion), and tr1 resembling derived transcript 1 from Aminetzach et al. (Aminetzach et al., 2005). The sequences from ORF1 of the *Doc* insertion are expressed in-frame which means the transcripts encode a gene-TE chimera. tr3 which encodes the entirety of ORF1 up until the stop codon, has a predicted protein 648 amino acids long, 85 of which are of gene origin and 563 of which are of *Doc* origin.

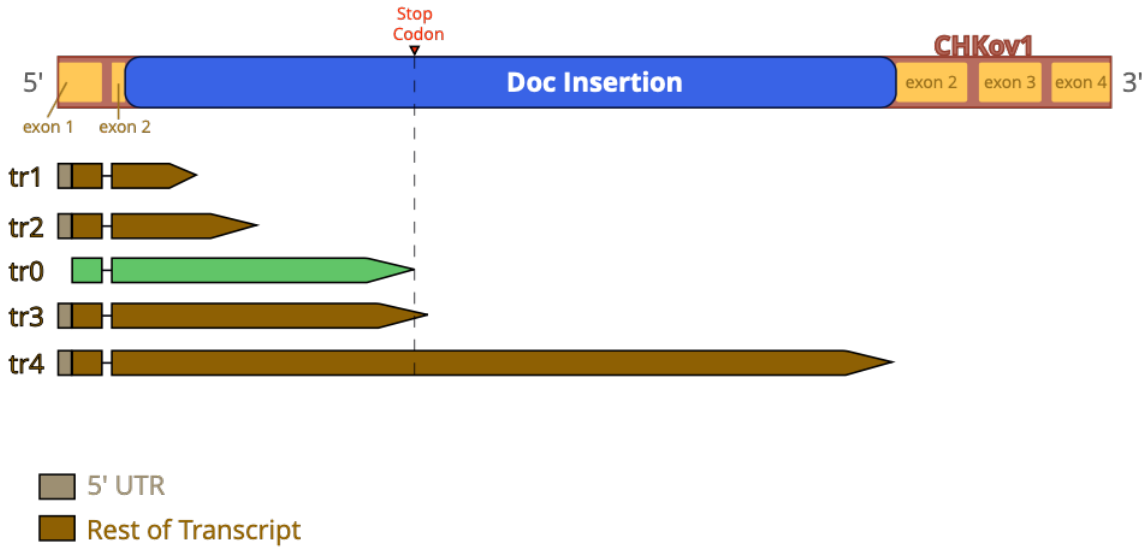


Figure 3.1. Nanopore sequencing shows four transcripts mapping to *CHKov1*. Using pinfish we were able to narrow down the *CHKov1*^{Doc1420} transcript to four possibilities (tr1-4). We designed transcript tr0 (green) which is the sequence of tr3 with the 5'UTR and the 3' UTR removed to test whether a truncated transcript is sufficient for DMelSV resistance. Sequences on benchling https://benchling.com/obrosh/f_/Zoz9jidS-thesis-sequences/ .

Novel *CHKov1* Transcript Causes Increased Resistance to DMelSV

To test whether a truncated *CHKov1* transcript is sufficient for generating resistance, we created a fly line which expresses truncated transcripts in a susceptible *CHKov1* background lacking the Doc1420 insertion (*CHKov1*$^{+}$). For that, we designed transcript tr0 (See figure 3.1), which is the sequence of tr3 with the 5'UTR and the 3' UTR removed. We cloned tr0 with an attached Venus tag on the 3' end into vector pUb-attb. This was used to create a fly line expressing *CHKov1*^{tr0}-Venus driven by a ubiquitin promoter (second chromosome) which has chromosome 3 of

$DGRP_{850}$ (the susceptible allele of *CHKov1*). Additionally, we created a control line which has chromosome 3 of $DGRP_{850}$, and the empty attP40 insertion site on the second chromosome. We infected both lines with Hap23 DMelSV to test for resistance.

We found that 12 days later, the viral titre in the control line was ~2500 times higher than the line with the insertion, a difference which was highly statistically significant (Figure 3.2, Welch's t-test: t=-9.19, $p=1.31x10^{-5}$, df=8.23). This indicates that the truncated transcript tr0 is sufficient in causing increased DMelSV resistance.

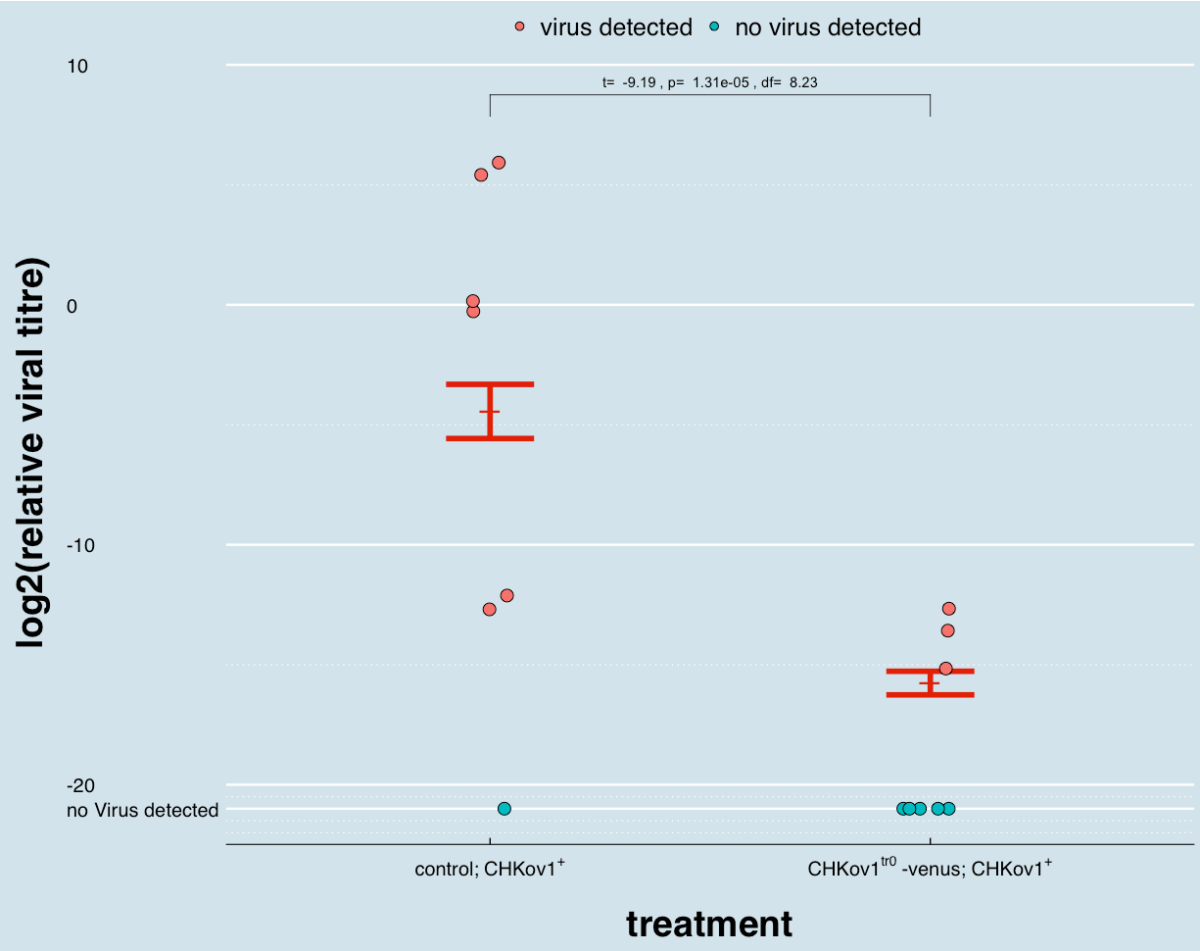


Figure 3.2. *CHKov1^{tr0}*-Venus results in lower DMelSV titre in flies. Fly lines with and without the *CHKov1^{tr0}*-Venus were infected with DMelSV. Viral titre was measured 12 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents 7-15 flies. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. *CHKov1^{tr0}*-Venus causes a significant decrease in viral titre from the control, which indicates that the transcript is sufficient to increase resistance in cells. P-values calculated using Welch's t-test.

Additionally, we created fly lines expressing *CHKov1*^{tr0}-GFP, *CHKov1*^{tr1}-GFP (no 5'UTR), and *CHKov1*^{tr2}-GFP (no 5' UTR) driven by UAS (second chromosome) (Primers in Appendix 12). These will be used in future experiments to test how much of the *Doc* element sequence is involved in resistance.

Knockdown of *CHKov1* with RNAi

To test whether the transcript of *CHKov1* is involved in its observed effect on resistance, we used RNAi to target the *CHKov1* transcript to check whether this results in a loss of the resistance function. We created fly lines carrying both the UAS driven TRiP (Zirin et al., 2020) construct targeting *CHKov1* as well as a Gal4 driver expressed ubiquitously by an actin promoter. We then injected these flies with Hap23 sigma virus, and incubated for 12 days at 25°C before extracting their RNA and using RT-qPCR to test for DMelSV viral titre relative to *Rpl32* (see main methods).

The results were inconclusive, with the knockdown causing only a (non-significant) increase in viral titre in the *CHKov1*$^{+}$ line, an unexpected result given our hypothesis. We attempted to repeat the experiment at 29°C to increase the activity of the RNAi knockdown. Unfortunately, the virus did not sufficiently replicate at 29°C for accurate measurement.

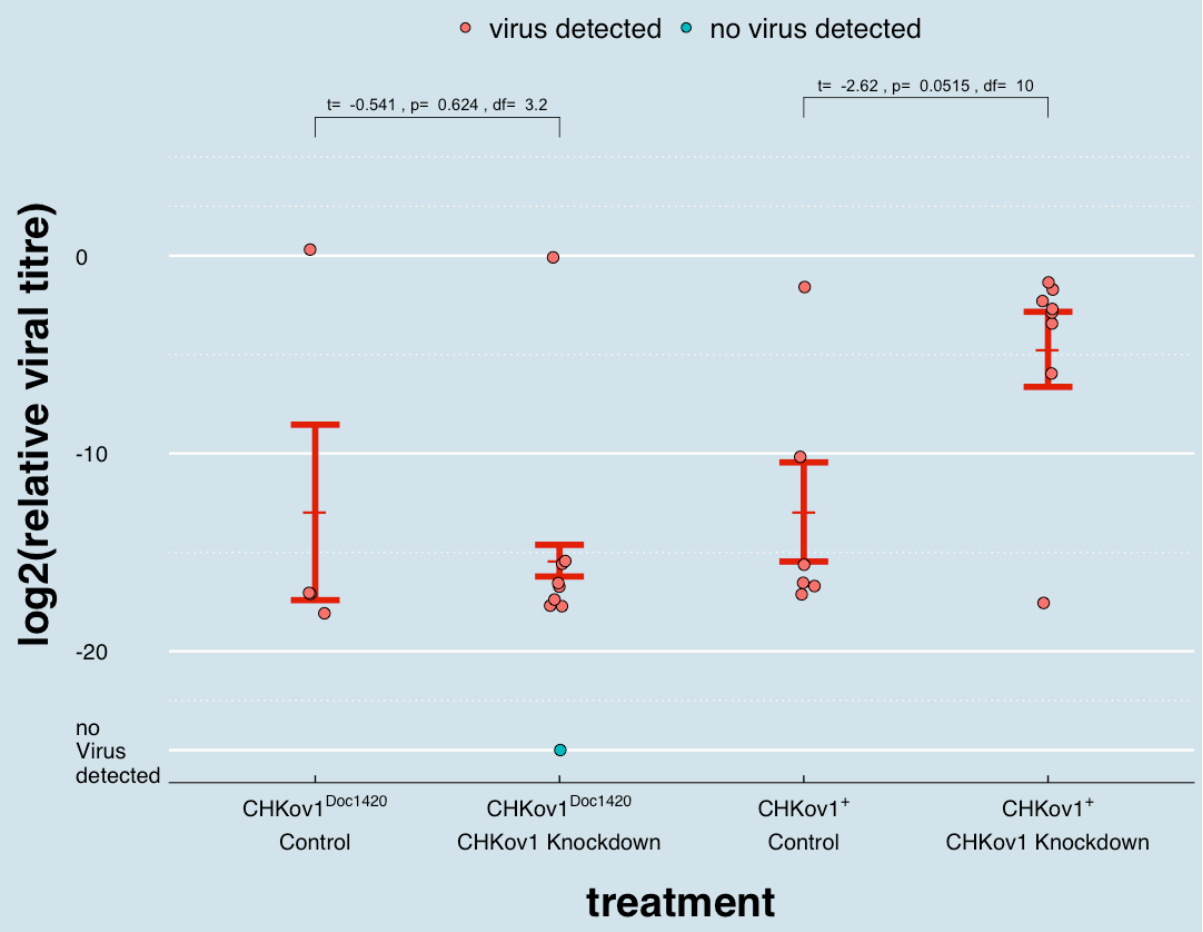


Figure 3.3. RNAi knockdown of *CHKov1* produced inconclusive results. *CHKov1*$^{+}$ and *CHKov1*^{Doc1420} lines with and without RNAi constructs targeting *CHKov1* were infected with DMelSV. Viral titre was measured 12 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents 7-15 flies. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. No significant effect of the knockdown is detected. P-values calculated using Welch's t-test, adjusted for multiple tests using the false discovery rate method with n=2.

Infecting Cells with DMelSV

We attempted to use cells as a system of studying DmelSV infection and the effect of *CHKov1*^{Doc1420} on resistance. To investigate this possibility, we infected with DMelSV DL2 cells and DL2 cells expressing the B2 protein from FHV (provided by Gaspar Bruner), which causes increased susceptibility to viral infection. However, we were unable to detect any observable viral replication (Figure 3.4).

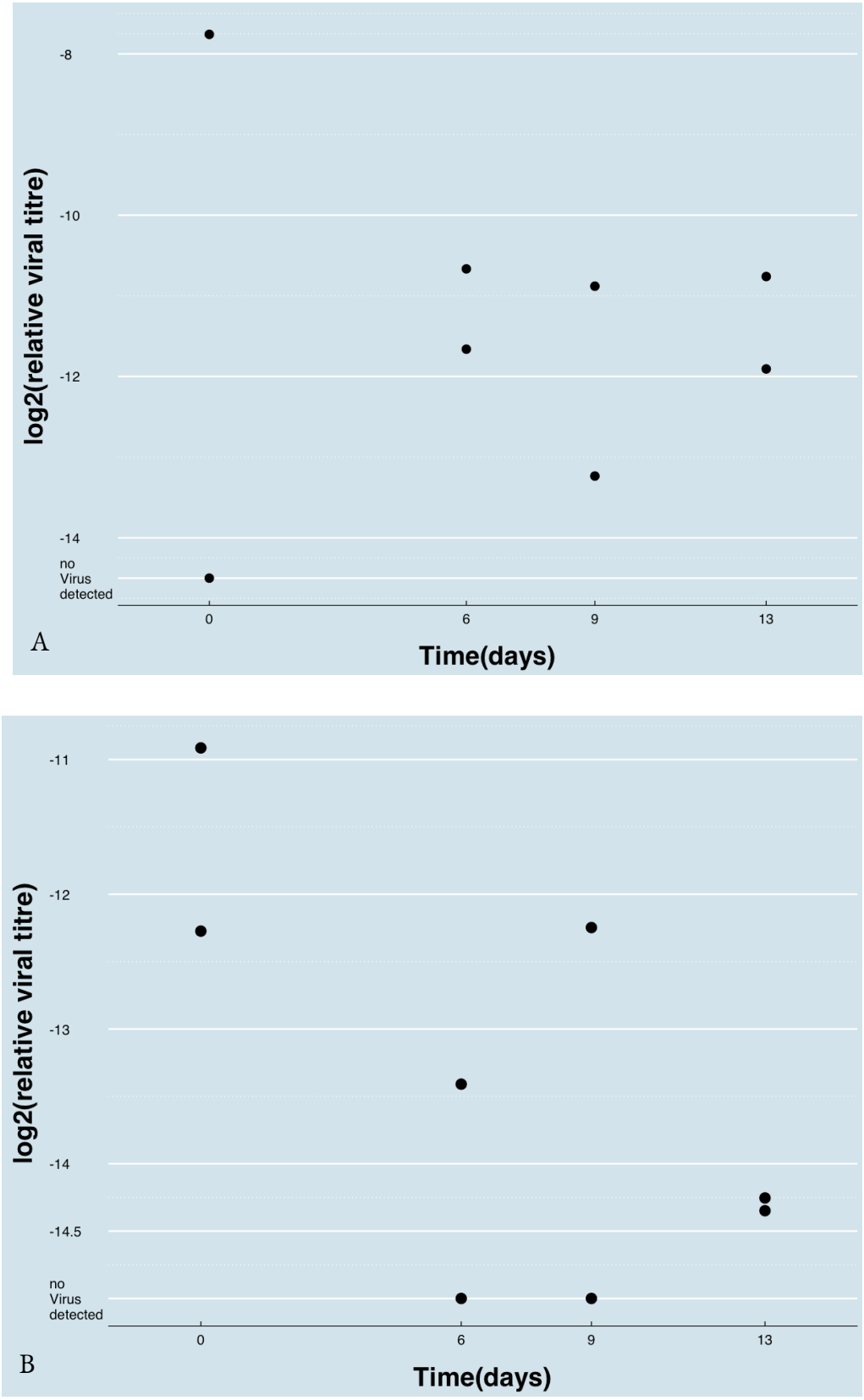


Figure 3.4. No evidence DMelSV titre increases with time in infected cells. Plot shows log_2(relative viral titre) of **(A)** DL2 cells and **(B)** DL2 cells expressing B2 infected with DMelSV 0 ,6, 9 and 13 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. In both cases, there does not seem to be an upward trend in viral titre with time.

Discussion

Because of their large fitness costs and rapid adaptations to host defences, viruses are a source of continual selection pressure on their hosts. This generates a lot of evolutionary novelty in host antiviral genes, as mutations which result in antiviral functions are selected for, and become more abundant. In this chapter we discuss a novel restriction factor which appears as a result of an insertion of the *Doc* retrotransposon into *CHKov1* that causes resistance to DMelSV. This is strikingly similar to the result discussed in Chapters 1 and 2, wherein a TE of the same type inserts into the gene *Veneno*, generating a novel restriction factor which causes DAV resistance. In both cases, the *Doc* element causes a truncation of the host gene. However, whereas the insertion in *Veneno* results in a transcript which encodes very few amino acids of *Doc* element origin, which are not required for the resistant phenotype, the results we discuss here suggest a different possibility for the *Doc* insertion in *CHKov1*.

We have shown that the Doc1420 transposable element insertion results in 4 truncated transcripts of *CHKov1*: tr1-tr4 (Figure 3.1). Tr1 is similar to the derived transcript 1 detected by Aminetzach et al. (Aminetzach et al., 2005), only differing by a few base pairs. We have also added to that finding transcripts tr2-tr4, which include larger regions of the *Doc* element than Aminetzach et al. detected. One of these transcripts, tr3 encodes the entirety of ORF1 (with the exception of the first five amino acid which are not encoded in this instance of the *Doc* insertion), in frame, up until the stop codon, forming a predicted gene-TE chimera protein 648 amino acids long, 563 of which are of *Doc* origin.

Furthermore, we have shown that tr0, which is the translated region of tr3, is sufficient to cause resistance to DMelSV (Figure 3.2). When overexpressed, the transcript results in a massive decrease of ~50,000-fold in DMelSV titres in infected flies. This confirms the hypothesis that the *Doc1420* insertion results in increased resistance to DMelSV by creating a truncated transcript of *CHKov1.* This truncated transcript may or may not require the regions of *Doc* element origin in causing the resistant phenotype.

In the former case, the chimeric protein this transcript is predicted to encode could rely on an added functionality requiring the amino acids coded for by ORF1 of the *Doc* element. ORF1 in *Doc* is predicted to have a nucleic acid (possibly RNA) binding function (O'Hare et al., 1991) and has been found to localize in irregular clusters in the cytoplasm (Rashkova et al., 2002). In *I* factors, another non-LTR transposon, the protein expressed by ORF1 binds both RNA and DNA, and assists in the annealing of complementary strands of nucleic acid, appearing to play a similar role to Gag proteins in retroviruses (Dawson et al., 1997). In *LINE-1*, ORF1 codes for a protein p40 which binds to RNA in a sequence specific manner (Hohjoh & Singer, 1997). ORF1 in *Doc* might potentially be able to bind to DMelSV sequences, or assist antiviral RNA, such as one wielded by an siRNA RISC complex, in annealing to DMelSV strands. If so, *CHKov1*^{Doc1420} would encode a chimeric gene with a novel function.

However, the entirety of the tr0 sequence might not be necessary for this antiviral function. In that case, simply creating the shorter transcript might enough for the gain of function to occur. This would be similar to the example we've highlighted in our Chapter 1, where we show that a *Doc* element insertion into the sequence of *Veneno* creates a truncated transcript with a novel antiviral function which does not require the sequences of ORF1 or ORF2 to perform its function. Further experimentation is required to determine which of the two possibilities is true. In particular, an experiment comparing the effects of expressing tr1 and tr2 (for which we've also prepared transgenic flies) with that of expressing tr0 should give us a better idea about how much of the *Doc* element sequence is required for DMelSV resistance.

Magwire et al. have previously shown that a second polymorphism involving *CHKov1*, namely two duplications of parts of *CHKov1* (including parts of the Doc1420 insertion) and *CHKov2* result in added resistance to DMelSV (Magwire et al., 2011). Although they find that the expression of *CHKov1* is not increased by these duplications (Magwire et al., 2011), the primers they use do not measure the expression of tr1-4. Therefore, it is possible that the duplication acts by increasing the expression of one of these transcripts. Later measurements have shown that the expression of one

(or more) of these transcripts is indeed increased as a result of the duplication (Chuan Cao, per. comm.).

Although our attempts at using RNAi to show the involvement of the transcript in resistance have been unsuccessful (figure 3.3), the experiments could be repeated in the future with slightly different conditions to improve the likelihood of success. One way this could be done is to grow the RNAi flies at 29°C to improve the efficiency of the RNAi, but then to move them to 25°C upon infection to allow for better viral replication. A future experiment could also measure knock-down efficiency (which we did not do) to get a clearer picture of what is happening. However, the results shown in (figure 3.2) could be seen as enough evidence for the involvement of the transcript in DMelSV resistance, making a repeat of the RNAi experiment unnecessary.

The same could be said about our attempts to infect cells with DMelSV. We had prepared a DL2 cell line which expresses tr0 in order to measure the effect this allele has on virus resistance in cells, but we were unable to infect any cells with DMelSV (Figure 3.4). Although unsuccessful our results do not preclude the possibility this could be done successfully in the future with different parameters. One way to improve the likelihood of success would be to increase the amount of virus added initially to each well. Indeed, Liao et al. have recently successfully infected S2 cells by adding 100µl of virus extract to 50µl of medium containing cells (Liao et al., 2019). In comparison we used 10µl of 40 times diluted cell extract for 90µl of medium containing cells. With that said, using cells to study infection provides several advantages, including the ease of preparing transgenic cell lines compared to preparing transgenic fly lines; however, the advantages of using the cells as a system as a mechanism for studying DMelSV resistance might be outweighed by the drawbacks caused by the difficulty in infecting the cells with the virus. On the other hand, we have had success in expressing transgenic constructs in flies and measuring their effects using fly injection. Until we overcome the challenges in infecting cells with sigma virus, flies remain the more powerful system.

Methods

Nanopore (PCR cDNA)

We used Nanopore PCR cDNA sequencing to sequence the whole transcriptome of $DGRP_{362}$ flies carrying *CHKov*^{Doc1420}: We extracted RNA from ~12 female flies from the 362 DGRP line using TRIzol extraction, then purified using Qiagen Rneasy Mini Kit column following manufacturer protocol. While the RNA was bound to the column, we digested any possible contaminant DNA using the Qiagen Rnase-Free Dnase Set following manufacturer protocol. We ran our sample through an agarose gel to check for quality. We sequenced the RNA using the cDNA-PCR Sequencing Kit (SQK-PCS109) following the manufacturer's protocol, using KAPA pure beads instead of AMPure XP beads and RNAsin Ribonuclease inhibitor (Promega) instead of RNAseOUT (ThermoFisher).

Bioinformatics

To analyse nanopore data we first used pychopper (ver 2.4.0, https://github.com/nanoporetech/pychopper) to identify, trim, and orient full length cDNA reads, we then used Minimap2 (ver 2.1, https://github.com/lh3/minimap2) to align our sequences to the *D. melanogaster* genome (dmel_r6.28_FB2019_3; Larkin et al., 2021) (which includes the *CHKov1*^{Doc4120} allele. We followed the recommended minimap2 settings for Nanopore transcriptome data. We used samtools (ver 1.9, https://github.com/samtools/samtools) to sort and index our alignment, then used the pinfish pipeline (https://github.com/nanoporetech/pinfish) (which clusters reads with similar exon/intron structures then generates consensus exon boundaries for each cluster) to find likely transcripts based on the aligned transcriptome. We used the default recommended settings with pinfish except for the minimum cluster size (smallest number of reads that can comprise a cluster) which we set to 4 to account for less abundant transcripts. The script which we used for all of this Bioinformatic analysis can be found on github: https://github.com/osamabrosh/NanoporeDataAnalysis .

CHKov1 in DL2 cells

We prepared pMT-puro-*CHKov1*^{tr0}-V5 (Figure 3.5) using NEBuilder (See main methods). We inserted *CHKov1*^{tr0} (Figure 3.1) which we amplified in two parts from DGRP_362 (Appendix 10: Primers pMTchkov_1FR and pMTchkov_2FR) into pMT-puro (addgene 17923, Figure 1.18) which has been digested with NotI-HF.

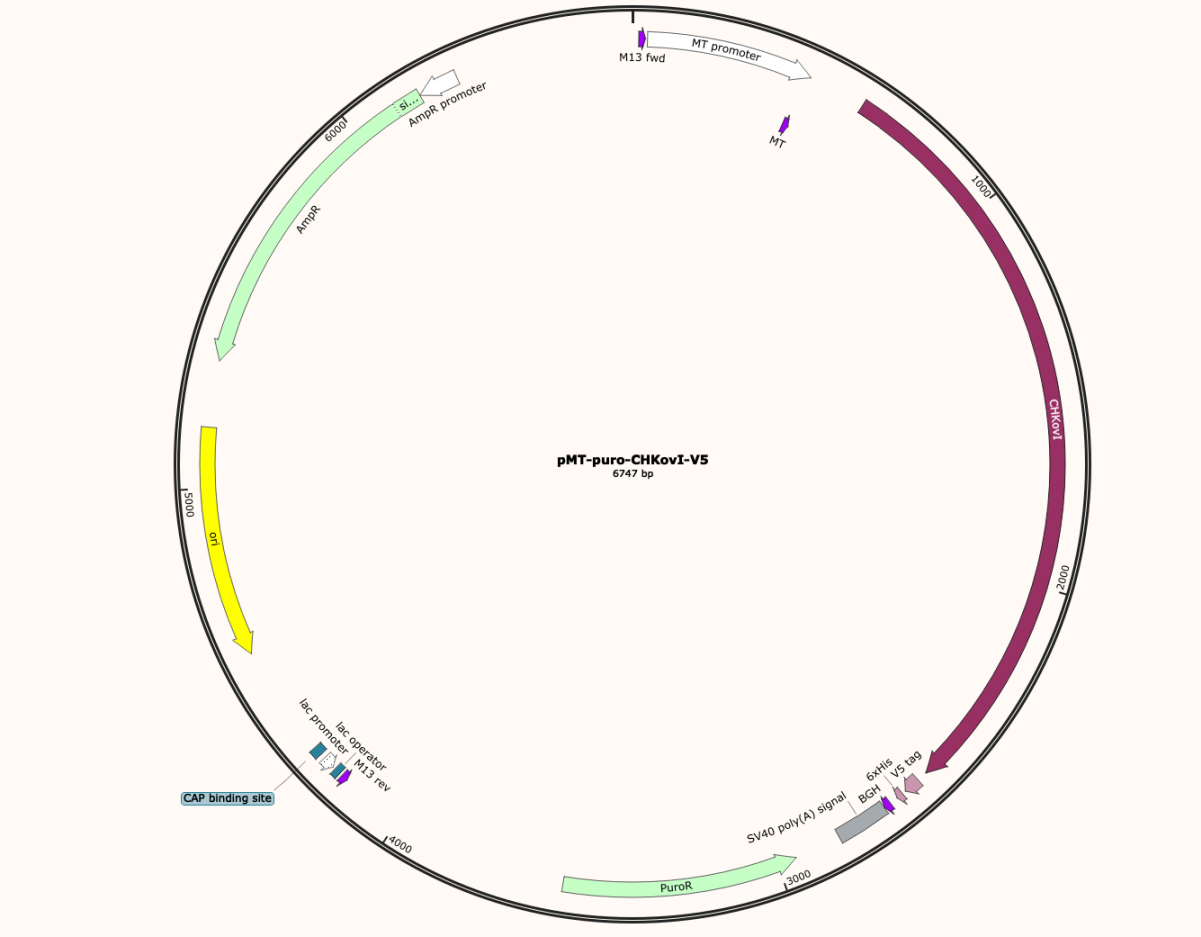


Figure 3.5 pMT-puro-*CHKov1*^{tr0}-V5 Expresses *CHKov1*^{tr0}-V5, the sequence of *CHKov1* transcript without its UTRs attached to a V5 tag, under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder (see main methods) to insert *CHKov1*^{tr0} in two fragments, which we amplified from DGRP_362 DNA using primers in appendix 10.

We infected the cells following the same methods for DAV infection (see main methods) using 90µl of 1.5x10^6 cells/ml with 10µl of Hap23 DMelSV which has been diluted 40 times with medium containing 5mM $CuSO_4$.

CHKov1 RNAi

We used lines w*;UAS-TRiP;TM6B/MKRS (from Chuan Cao, UAS-TRiP construct made using oligos in Appendix 13), w*;attp40;DGRP, w*;If/Cyo;DGRP, and w*;Act5C-Gal4/Cyo;TM6B/MKRS (all previously produced in the lab with DGRP lines 850 and 362), to make w*;UAS-TRiP/ Act5C-Gal4;DGRP_850, w*;UAS-TRiP/ Act5C-Gal4;DGRP_362, w*;attp40/ Act5C-gal4;DGRP_850, and w*;attp40/ Act5C-gal4;DGRP_362 using the crossing scheme in Figure 3.6. We infected the resulting flies with DMelSV then extracted their RNA with TRIzol, reverse transcribed it, and tested it for sigma virus infection using qPCR (see main methods).

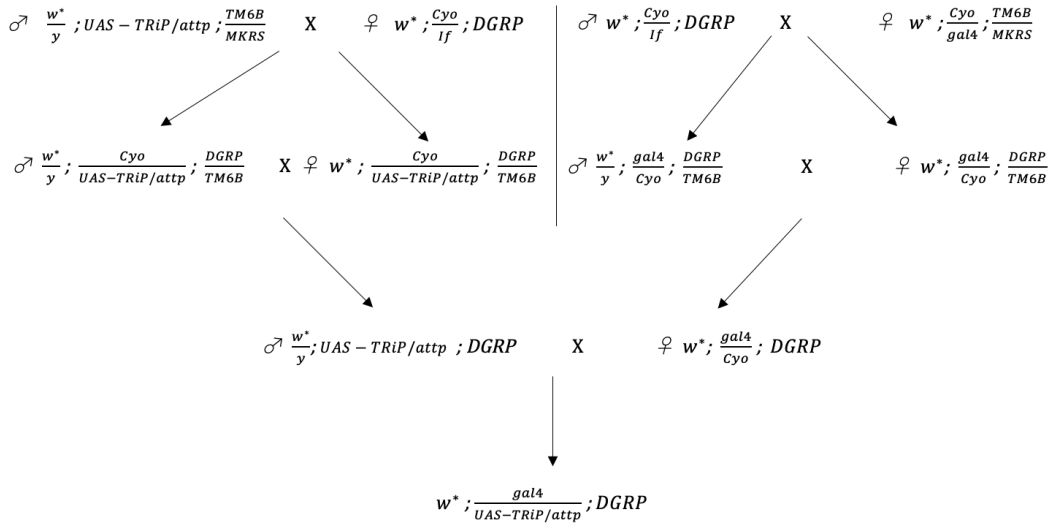


Figure 3.6 Crossing scheme to produce w*;TRiP/ Act5C-Gal4;DGRP_850 , w*;TRiP/ Act5C-Gal4;DGRP_362, w*;attp40/ Act5C-gal4;DGRP_850, and w*;attp40/ Act5C-gal4;DGRP_362

CHKov1 truncated transcript in flies

We prepared plasmids pUB-*CHKov1*^{tr0} (Figure 3.7) and pUB- *CHKov1*^{tr0}-Venus (Figure 3.8) using NEBuilder (see main methods) for insertion into fly lines. We used as insert the sequence *CHKov1*^{tr0} (amplified from pMT-puro- *CHKov1*^{tr0}-V5 (Figure 3.5)) using primers CHKovtopubnotag_FR for pUB-*CHKov1*^{tr0} and CHKovtopubtag_FR for pUB- *CHKov1*^{tr0}-Venus (Appendix 11). For pUB- *CHKov1*^{tr0}-Venus we added a second insert: the Venus tag (including a 3xFLAG sequence, and two flanking Strep-Tag II sequences) which we amplified from a Venus Tag DNA sequence (using primers Venuschkov_FR, Appendix 10). Sequences were inserted into pUb-attB (provided by Jinghua Gui (Karam-Teixeira lab)) which has been digested with NotI-HF.

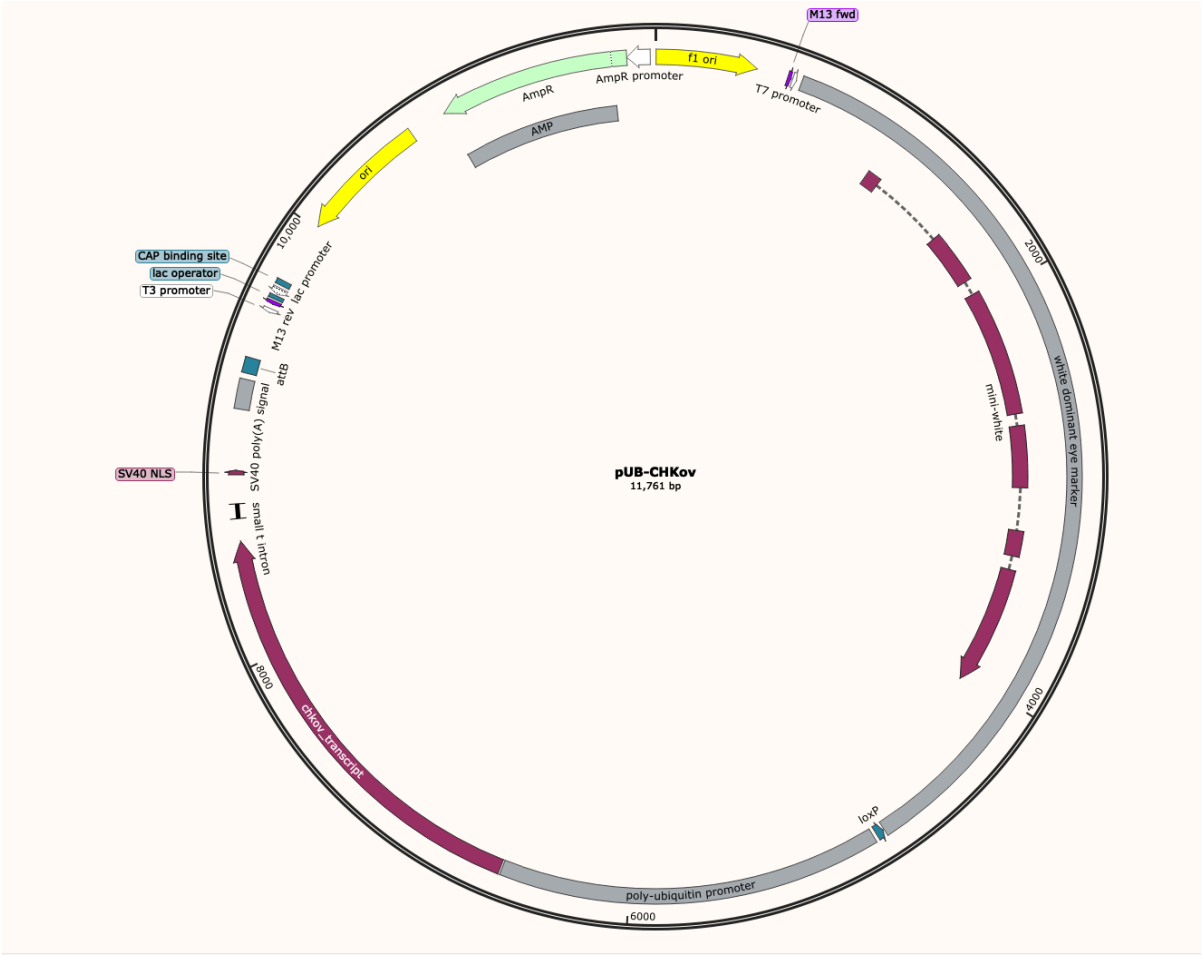


Figure 3.7 Plasmid pUB- CHKov1^{tr0} Expresses CHKov1^{tr0}, the sequence of *CHKov1* truncated just upstream of the first in-frame stop codon (occurring within the *Doc* element) under the control of the ubiquitin promoter. Contains attB sequence for phiC31 insertion into a fly line. We digested pUb-attb (provided by Jinghua Gui (Karam-Teixeira lab)) with NotI-HF, then used NEBuilder (see main methods) to insert CHKov1^{tr0} which we amplified from pMT-puro-CHKov1^{tr0}-V5 DNA using primers in appendix 11.

Figure 3.8 Plasmid pUB-CHKov1^{tr0}-Venus Expresses CHKov1^{tr0}, the sequence of *CHKov1* truncated just upstream of the first in-frame stop codon (occurring within the *Doc* element) attached to a Venus fluorescent reporter under the control of the ubiquitin promoter. Contains attB sequence for phiC31 insertion into a fly line. We digested pUb with NotI-HF, then used NEBuilder (see main methods) to insert CHKov1^{tr0} and the Venus tag (including a 3xFLAG sequence, and two flanking Strep-Tag II sequences) which we amplified from pMT-puro-CHKov1^{tr0}-V5 DNA and a Venus Tag DNA sequence respectively using primers in appendix 11.

Fly Lines

After preparing the plasmids, we sent the pUB- CHKov1^{tr0} and pUB-CHKov1^{tr0}-Venus plasmids to BestGene Inc for microinjection into the Attp40 stock (on chromosome II) which expresses phiC31 driven by *vasa*. We used the resulting ;CHK^{tr0}/+; flies as well as double balancer lines and line w*;If/Cyo;DGRP$_{850}$ (previously produced) to create the line w*; CHK^{trunc0};DGRP$_{850}$ using the

crossing scheme in figure 3.9. We also used the untransformed attP40 line to create the control line w*; attP;DGRP$_{850}$ using the same crossing scheme.

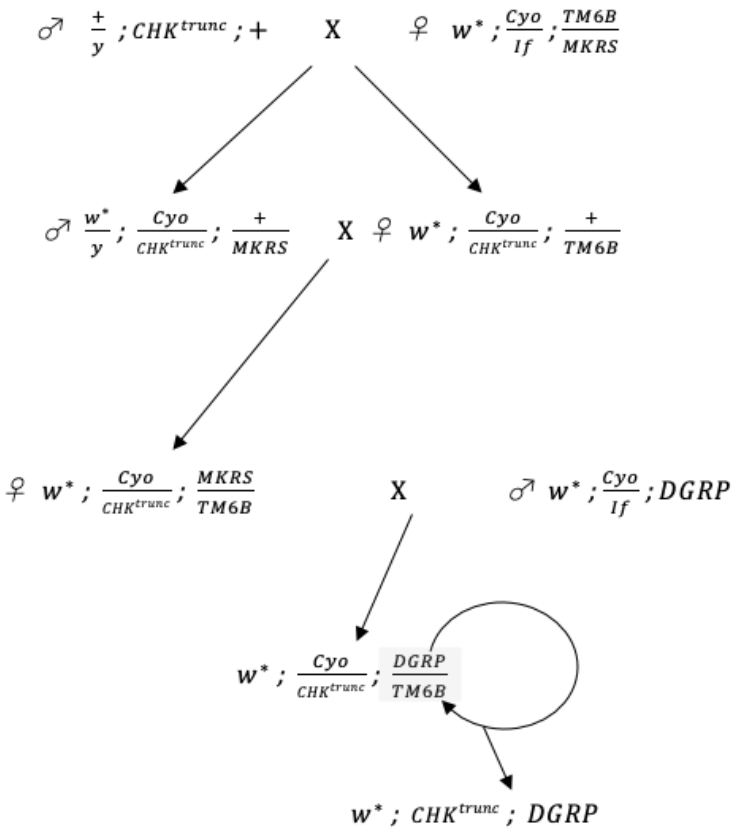


Figure 3.9 Crossing scheme to produce lines w*; CHK^{trunc0};DGRP$_{850}$ and w*; attP; DGRP$_{850}$

Experiment

We infected lines w*; CHK^{trunc0}-Venus; DGRP$_{850}$ and w*; attP40; DGRP$_{850}$ with Hap23 DMelSV then extracted their RNA with TRIzol, reverse transcribed it, and tested it for sigma virus infection using qPCR (see main methods).

Chapter 4 – Genetic Engineering of Virus Resistance Using CRISPR/Cas13

Introduction

Arthropod-borne viruses (arboviruses) such as Zika Virus, Dengue Virus, Yellow Fever Virus and Chikungunya Virus are transmitted to humans by blood feeding arthropods such as the mosquito *Aedes aegypti.* Together these viruses account for a large number of infections leading to adverse effects in humans. For example, the Zika Virus outbreak in 2016 led to more than 500,000 cases of Zika Virus infection in Brazil, a soberingly large number given the virus' association with congenital microcephaly and other neurological abnormalities in foetuses and new-borns when it occurs during pregnancy (Martins et al., 2021). On the other hand, estimates for worldwide infections with Dengue range between 50 million and 500 million infections per year with around 1 in 4 cases resulting in symptoms such as fever, severe muscle aches, headaches, and rashes (Castro et al., 2017). Given these large disease burdens, it is incredibly important to find ways to combat the spread of arboviruses.

One proposed method to limit arbovirus spread is to genetically engineer virus resistant mosquitos: resistant mosquitos are less likely to become infected with arbovirus, and therefore less likely to transmit it to humans. Several methods to do so have been attempted.

One idea is to target endogenous insect antiviral systems, such as RNAi, against arboviruses. Olson et al. transduced female mosquitos with antisense RNA targeting the genome of the virus, thereby making use of the mosquito RNAi system to reduce Dengue type 2 virus replication in the salivary gland of *Aedes aegypti*, and in so doing, limit the virus' ability to transmit to humans (Olson et al., 1996). This was extended further by Franz et al. who expressed the inverted repeat RNA transgenically in the genome of the mosquito, thereby creating Dengue resistant mosquitos capable of passing down their resistance to their offspring (Franz et al., 2006). Jupatanakul et al. made use of a different endogenous system, JAK/STAT, to induce viral resistance (Jupatanakul et

al., 2017). By upregulating that pathway, they were able to cause increased resistance to Dengue Virus, but not Zika Virus or Chikungunya Virus (Jupatanakul et al., 2017).

However, since RNAi and JAK/STAT are endogenous antiviral pathways in arthropods, viruses have had a long time to evolve ways to combat it. For example, the Flock House Virus expresses the B2 protein which suppresses the RNAi pathway in *Drosophila* cells upon infection (Li et al., 2002). Other viruses which infect arthropods exhibit their own RNAi pathway silencing, such as Drosophila C Virus (van Rij et al., 2006) and Cricket Paralysis Virus (X.-H. Wang et al., 2006).

It is therefore a good idea to attempt to engineer viral resistance in flies using a mechanism which is not endogenous to arthropods. One such system is CRISPR-Cas, which is an antiviral system native to prokaryotes (Barrangou et al., 2007). Several attempts have been made to use CRISPR-Cas9, which targets DNA, to engineer resistance against DNA viruses such as Hepatitis B Virus (X. Liu et al., 2015) and retroviruses such as HIV-1 (W. Zhu et al., 2015). Some variants of Cas9 which target RNA have also been used to target RNA viruses such as Hepatitis C Virus (Price et al., 2015).

Here we explore using Cas13, which specifically targets RNA based on a target sequence specified by CRISPR RNA (crRNA) (Y. Zhu et al., 2018) to engineer arbovirus resistance in mosquitos. To test the feasibility of this system, we conduct an exploratory experiment in *Drosophila melanogaster* to see if we can induce resistance to Drosophila A Virus (DAV) using CRISPR-Cas13. However, we were unable to find evidence that this system is feasible against DAV, and find further experimentation to be required.

Results

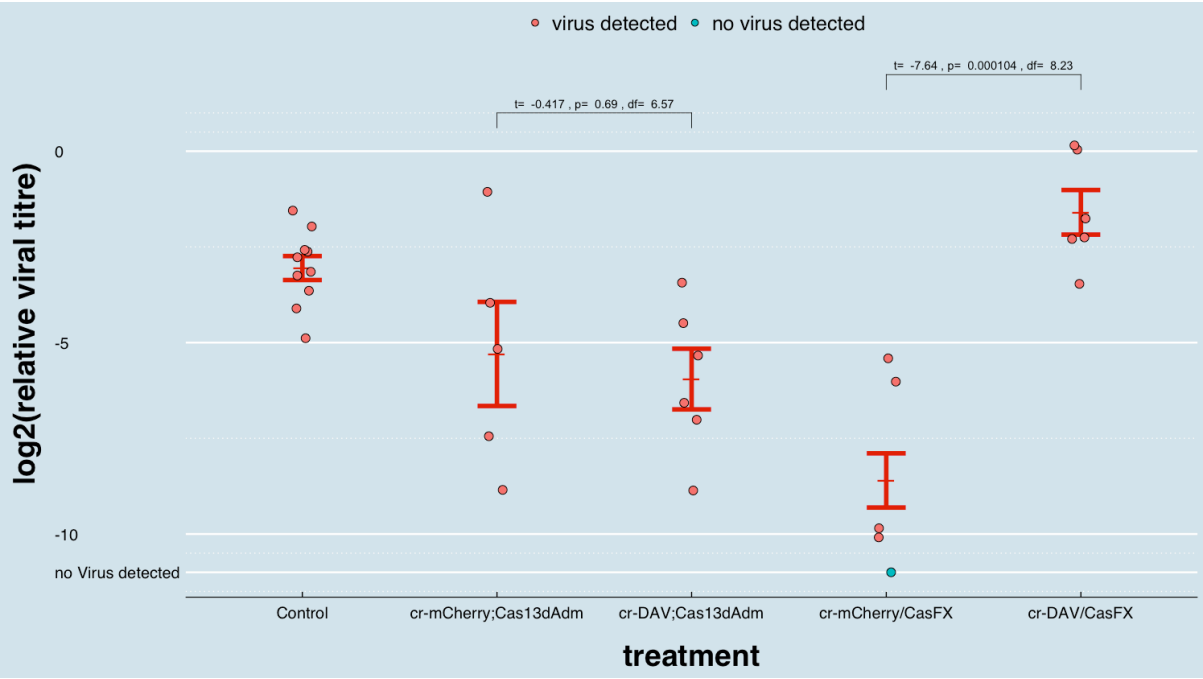
Cas13 targeting DAV

To test for whether Cas13 is a viable tool for targeting viruses in insects, we checked whether the system could be used to engineer DAV resistance in *D. melanogaster.* To do so, we created

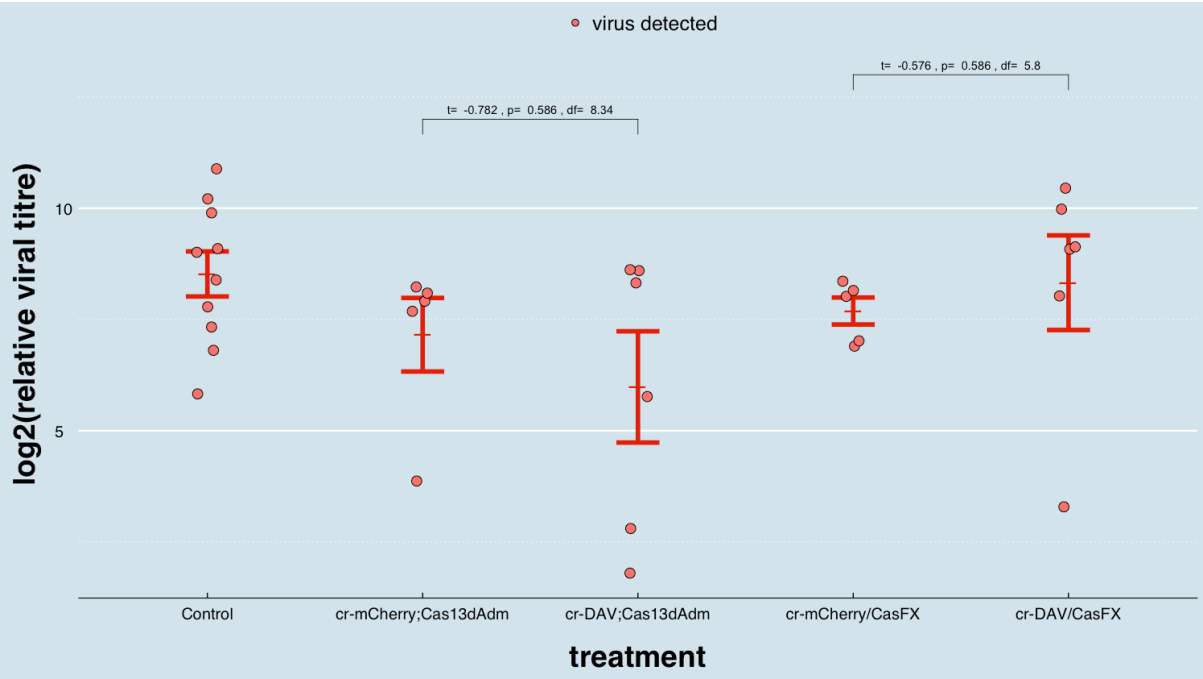
transgenic flies expressing the Cas13 variant Cas13dAdm (from the anaerobic digester metagenome (Adm)) (Konermann et al., 2018) as well as the fluorescent protein mCherry connected by T2A peptide (which is cleaved off after translation) and both driven by a ubiquitin promoter. The flies also expressed CRISPR RNA (crRNA) targeting 4 sites in either DAV or mCherry, separated by direct repeats specific to Cas13d, and driven ubiquitously by a *U6:3* promoter, a strong type III RNA Polymerase III promoter. We also used a fly line expressing CasFX (kindly provided by Kirst King Jones) to create a fly line expressing the Cas13 variant CasFX (Huynh et al., 2020) driven by an Actin promoter, and a crRNA targeting either DAV or mCherry (though no mCherry is expressed by this line). We infected with DAV these flies expressing the CRISPR-Cas13 system as well as flies expressing only one part of the system (Cas13 only or crRNA only), then froze them after 1 day (24 ± 1 hours), 3 days (72 ± 3 hours), and 25 days (600 ± 3 hours). We extracted their RNA and measured viral titre using qPCR (using DAV_FR primers in Appendix 19 which do not amplify the crRNA).

We found no significant decrease in viral titre in the lines expressing CRISPR-Cas13 targeting DAV compared to the line expressing CRISPR-Cas13 targeting mCherry for either of the Cas13 variants and for any of the infection times (Figure 4.1A-C). Perplexingly, there was a significant *increase* in viral titre in the line expressing the CRISPR-CasFX system targeting DAV compared to the line expressing the CRISPR-CasFX system targeting mCherry 1 dpi (Figure 4.1A).

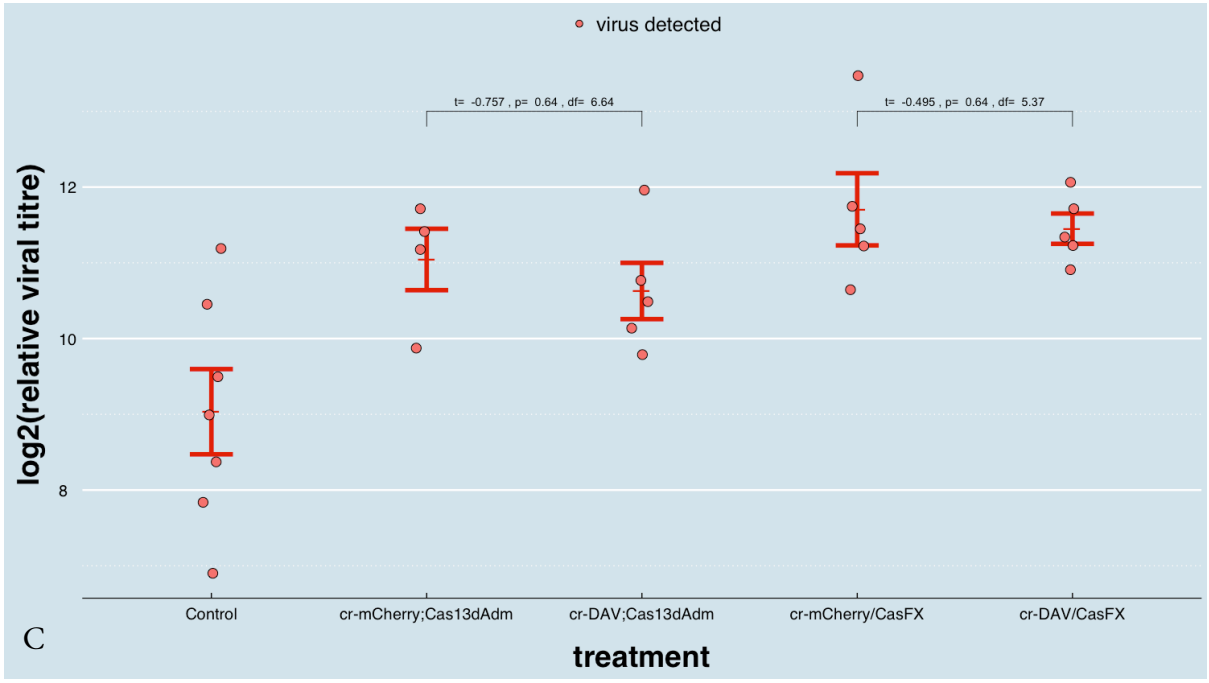
We also imaged the fly lines with CRISPR-Cas13dAdm systems (which also express mCherry) under a fluorescent light to check for mCherry expression compared to controls expressing only one component of the system (either Cas13dAdm or cr-mCherry). We found that both flies with CRISPR-Cas13dAdm targeting either mCherry or DAV did not look qualitatively different from one another (Figure 4.1d), indicating that mCherry expression wasn't abolished by the CRISPR-Cas13 system.

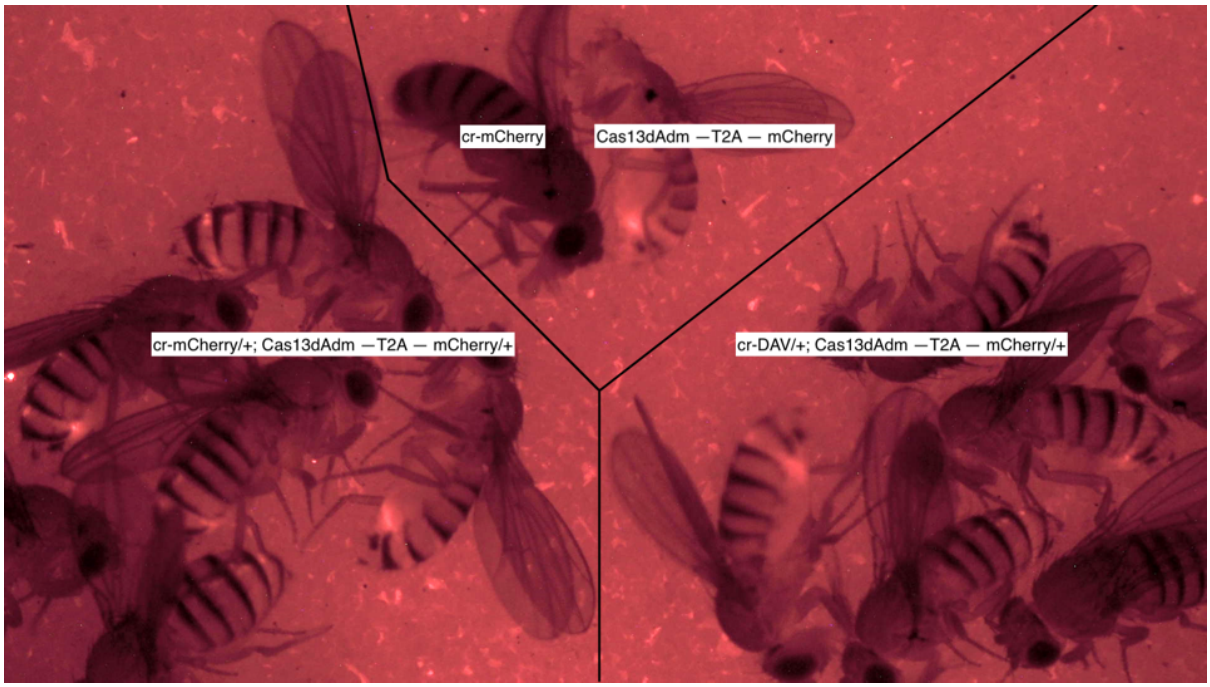

virus detected
no virus detected
t= -7.64 , p= 0.000104 , df= 8.23
t= -0.417 , p= 0.69 , df= 6.57
log2(relative viral titre)
0
-5
-10
no Virus detected
Control
cr-mCherry;Cas13dAdm
cr-DAV;Cas13dAdm
cr-mCherry/CasFX
cr-DAV/CasFX
treatment

A 1 dpi


virus detected
t= -0.576 , p= 0.586 , df= 5.8
t= -0.782 , p= 0.586 , df= 8.34
log2(relative viral titre)
10
5
Control
cr-mCherry;Cas13dAdm
cr-DAV;Cas13dAdm
cr-mCherry/CasFX
cr-DAV/CasFX
treatment

B 3 dpi





D

Figure 4.1. CRISPR-Cas13 Experiment Failed to Produce Antiviral Resistance to DAV
A,B,C) Flies expressing either a CRISPR-Cas13dAdm or a CRISPR-CasFX system targeting either DAV or mCherry as well as controls expressing only Cas13dAdm, only CasFX, only cr-DAV or only cr-mCherry were infected with DAV. Viral titre was measured with qPCR 1dpi **(A)**, 3 dpi **(B)** and 25 dpi **(C)**, and is shown here relative to Rpl32. Controls were combined. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. P-values calculated using Welch's t-test with and FDR adjustment for multiple tests with n=2. Flies expressing Cas13dAdm also expressed mCherry. There is no significant decrease in viral titre in flies expressing CRISPR-Cas13 targeting DAV compared to flies expressing CRISPR-Cas13 targeting mCherry.
D) Flies expressing mCherry as well as a CRISPR-Cas13dAdm system targeting either mCherry or DAV were imaged under a fluorescent microscope and compared to control flies either expressing only Cas13dAdm (with mCherry), or only crRNA targeting mCherry. It does not appear, qualitatively, that CRISPR-Cas13 targeting mCherry abolished mCherry expression to levels below that of the line expressing CRISPR-Cas13 targeting DAV.

Cas13 targeting Zika

To test whether the CRISPR-Cas13 system is viable against arboviruses which can infect humans, we created *Drosophila melanogaster* lines expressing crRNA targeting Zika virus. The idea is to use these flies to create lines with CRISPR-Cas13 systems targeting Zika, similar to our DAV experiment, then to attempt infection with Zika Virus to test whether the system causes Zika resistance. Unfortunately, this has not been completed in time for this thesis.

Discussion

We were unable to show that CRISPR-Cas13 is an effective system to target RNA viruses in insects. At the varying time points we tested, none of our results show a significant decrease in viral titre resulting from the simultaneous expression of the Cas13 protein and the CRISPR RNA targeting DAV. Given that this is a negative result, it is difficult to explain the underlying reason behind our data. One explanation is that the CRISPR-Cas13 system is incompatible with antiviral function. In Chapter 1 we have shown that RNAi is not involved in DAV resistance, so it is possible that the viral RNA of DAV is protected somehow from effector molecules. However, we do not know based on our data whether that is the case or whether our results can be explained in a different way, perhaps by experimental error which can be fixed in future iterations of this experimental design.

One argument in the favour of this latter explanation is the apparent failure of Cas13dAdm to target mCherry. Though it is qualitative data, the fluorescent imaging of the flies expressing Cas13dAdm – T2A – mCherry and cr-mCherry did not show that the Cas13 system reduced mCherry expression to below the levels seen in flies expressing Cas13dAdm – T2A – mCherry and cr-DAV. This may be seen as evidence that we were unable to use the CRISPR-Cas13 system successfully to target mCherry RNA. However, given that this data is qualitative this evidence is not definitive. Cas13 has been deployed successfully in flies, with CasFX reducing expression of a target by 80-90% (Huynh et al., 2020). We have used similar promoters as Huynh et al. to express

our crRNA; however, they used tRNAs (which are cleaved off by the fly's endogenous processes) to separate their multiple crRNA, whereas we relied on Cas13's ability to process crRNAs immature crRNA (composed of multiple crRNAs separated by direct repeats) into mature crRNAs. One possibility is that this processing did not work, in which case, repeating the experiment using tRNAs in a manner similar to Huynh et al. may allow it to work.

Another explanation which is consistent with our results is that insufficient expression of either component of the CRISPR-Cas system has resulted in a failure to produce antiviral function. One future direction to troubleshoot this work is to check whether both the Cas13 and crRNA are indeed being expressed using qPCR, something which can be done relatively easily with our collected samples. Another potential issue is that these components might not be expressed in the correct tissue. We have used ubiquitous promoters, but this does not guarantee that expression will actually occur to a high level in every cell. A small number of cells in which the CRISPR-Cas system is ineffectual may be sufficient in producing high viral titres. One way of testing this is to measure the expression of the CRISPR-Cas components in different tissue compared it to viral titre.

Another possibility which is consistent with our results is that CRISPR-Cas13 is only effective against viruses in so far that viruses do not escape by evolving differences in the parts of their genome targeted by CRISPR-Cas13, and that in our case the flies have indeed done so. In that case, escape mutants are not targeted by the CRISPR-Cas13 system, and would account for the viral titre being measured. This possibility might even help explain the apparently paradoxical result of the cr-DAV/CasFX flies having a higher viral titre 1dpi than the cr-mCherry/CasFX flies: the escape mutants may simply be able to reach higher titres more quickly than the original variant. One way to test this is by using Ilumina sequencing to sequence the virus RNA in the regions to check for escape mutants, which would allow us to see if the regions are indeed mutated in the flies with CRISPR-Cas13 systems targeting DAV.

Given that this experiment was exploratory, more steps should be taken to explore the reasons behind its failure. A first step would be to carry out the qPCR and Ilumina, which at the time of

writing this thesis we have not yet done. More diagnostic steps would be taken depending on the results of those experiments. Once the reason the experiment failed is identified, it could be repeated with modified conditions and then perhaps extended to Zika Virus using the fly lines we have prepared.

Methods

Cas13

We used NEBuilder (See main methods) to prepare plasmid pUb-mCherry-Adm-miniwhite (Figure 4.2). We digested pUb-attB (provided by Jinghua Gui) using NotI-HF, then inserted the mCherry-Adm sequence which we amplified from plasmid Fused-pACT-AdmCas13d-STABLE-Blasticidin (provided by Grant Hughes) using primers Adm_F and Adm_R in appendix 16.

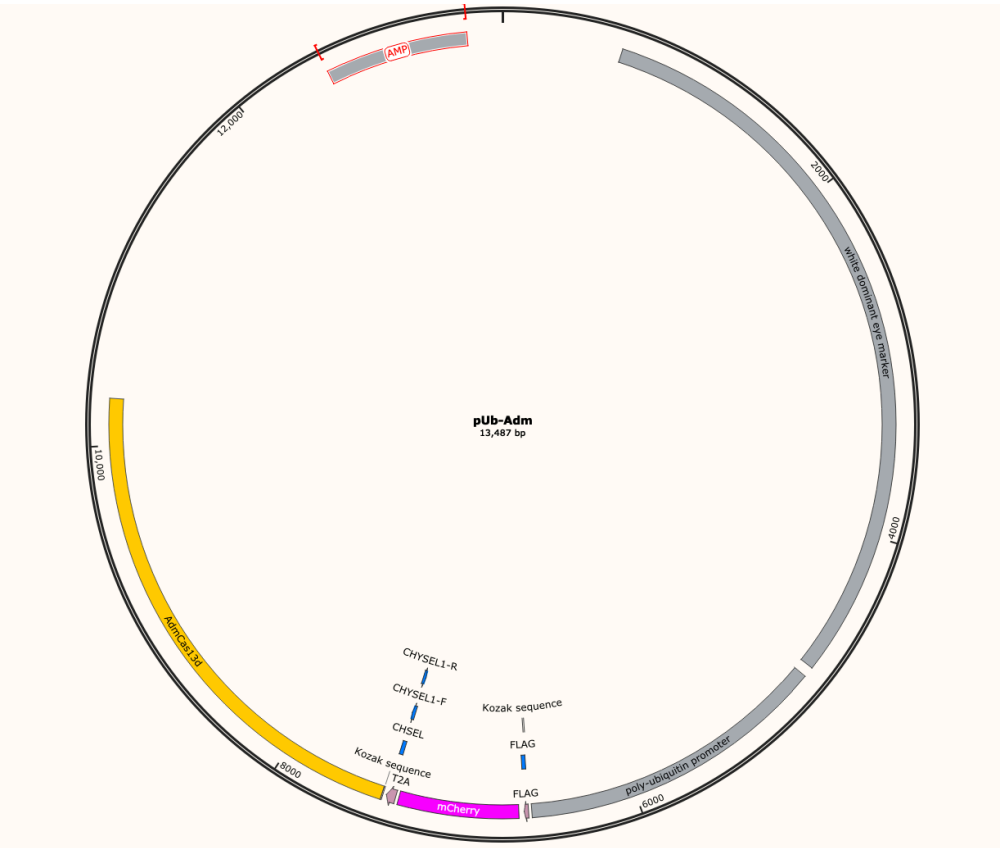


Figure 4.2 pUb-mCherry-Adm-miniwhite Expresses fluorescent protein mCherry and Cas13dAdm under the control of the Ubiquitin promoter. The two proteins are separated by a T2A sequence coding for a peptide sequence which is cleaved after translation to produce the two separate proteins. We digested pUb-attB (provided by Jinghua Gui) using NotI-HF, then used NEBuilder (see main methods) to insert the mCherry-Adm sequence which we amplified from plasmid Fused-pACT-AdmCas13d-STABLE-Blasticidin (provided by Grant Hughes) using primers in appendix 16. (Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

crRNA

Cas13 target sequences were chosen by our collaborators (Ian Patterson and Grant Hughes at the Liverpool School of Tropical Medicine) to maximize Cas13 efficiency. 4 targets were chosen for each of DAV, Zika, and mCherry, and the crRNA sequences were flanked by Cas13 direct repeats. We ordered ultramer DNA oligonucleotides (IDT) of the resulting crRNA-direct repeat sequences (two ultramers per virus/gene, Appendix 14). We amplified the ultramers using PCR reactions to create the inserts for the preparation of plasmids using primers in Appendix 15.

We then used NEBuilder to create plasmids pCFD5-crRNA-DAV-vermillion (Figure 4.3) and pCFD5-crRNA-mCherry-vermillion (Figure 4.4), and pCFD5-crRNA-Zika-vermillion (Figure 4.5): We digested pCFD5 (Addgene #73914; (Port & Bullock, 2016) using BbsI and inserted the amplified crRNA inserts using NEBuilder (see main methods). The plasmids express a vermillion eye colour selectable marker, and 4 crRNAs each, driven by U6:3, separated by direct repeats, and targeting DAV, mCherry, and Zika respectively.

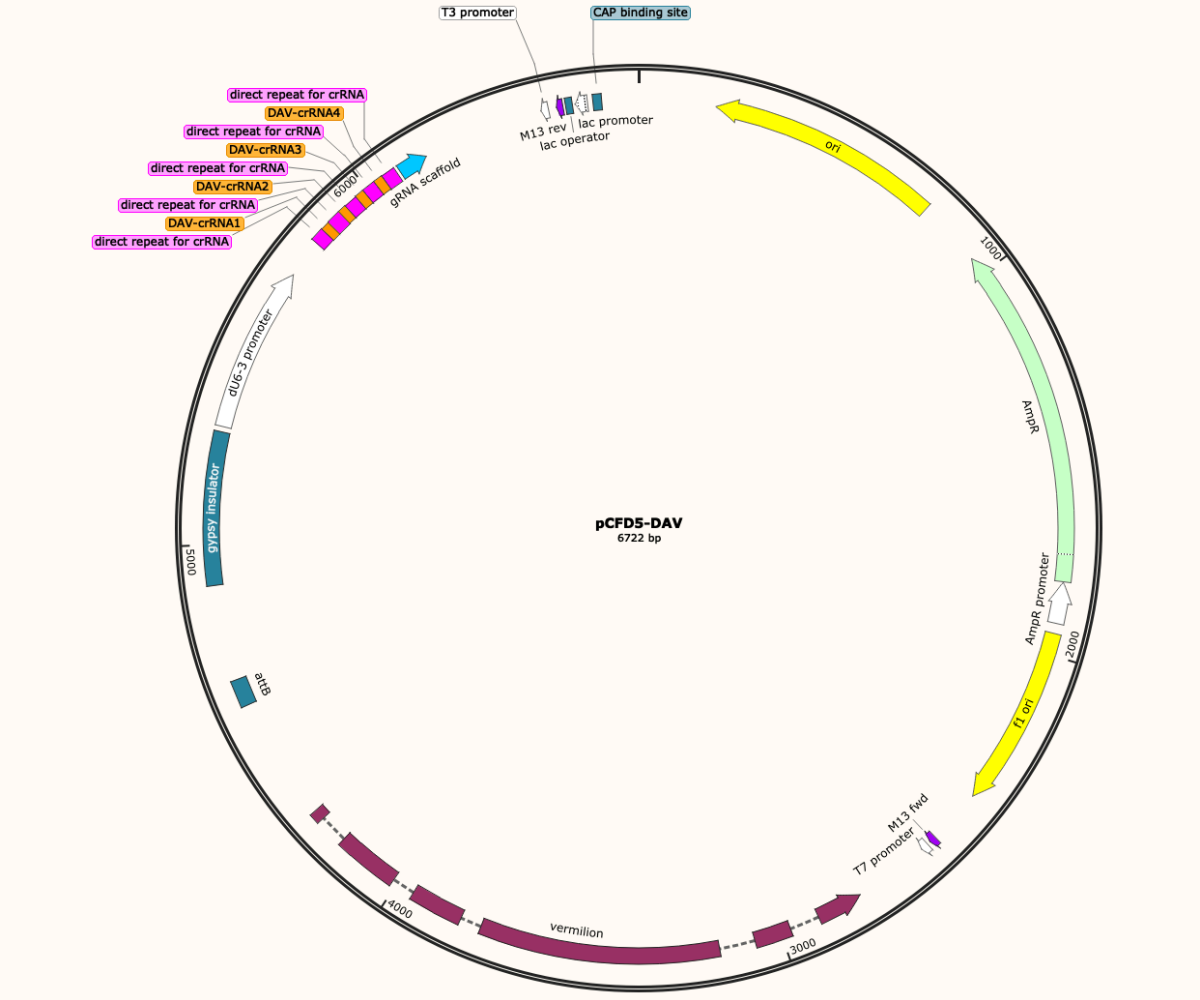


Figure 4.3 pCFD5-crRNA-DAV-vermillion Expresses DAV crRNA under the control of the U6:3 promoter. We digested pCFD5 (Addgene #73914) using BbsI, then used NEBuilder (see main methods) to insert the crRNA DAV sequence (which we amplified from the ultramer sequences) using primers in appendix 15.
(Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

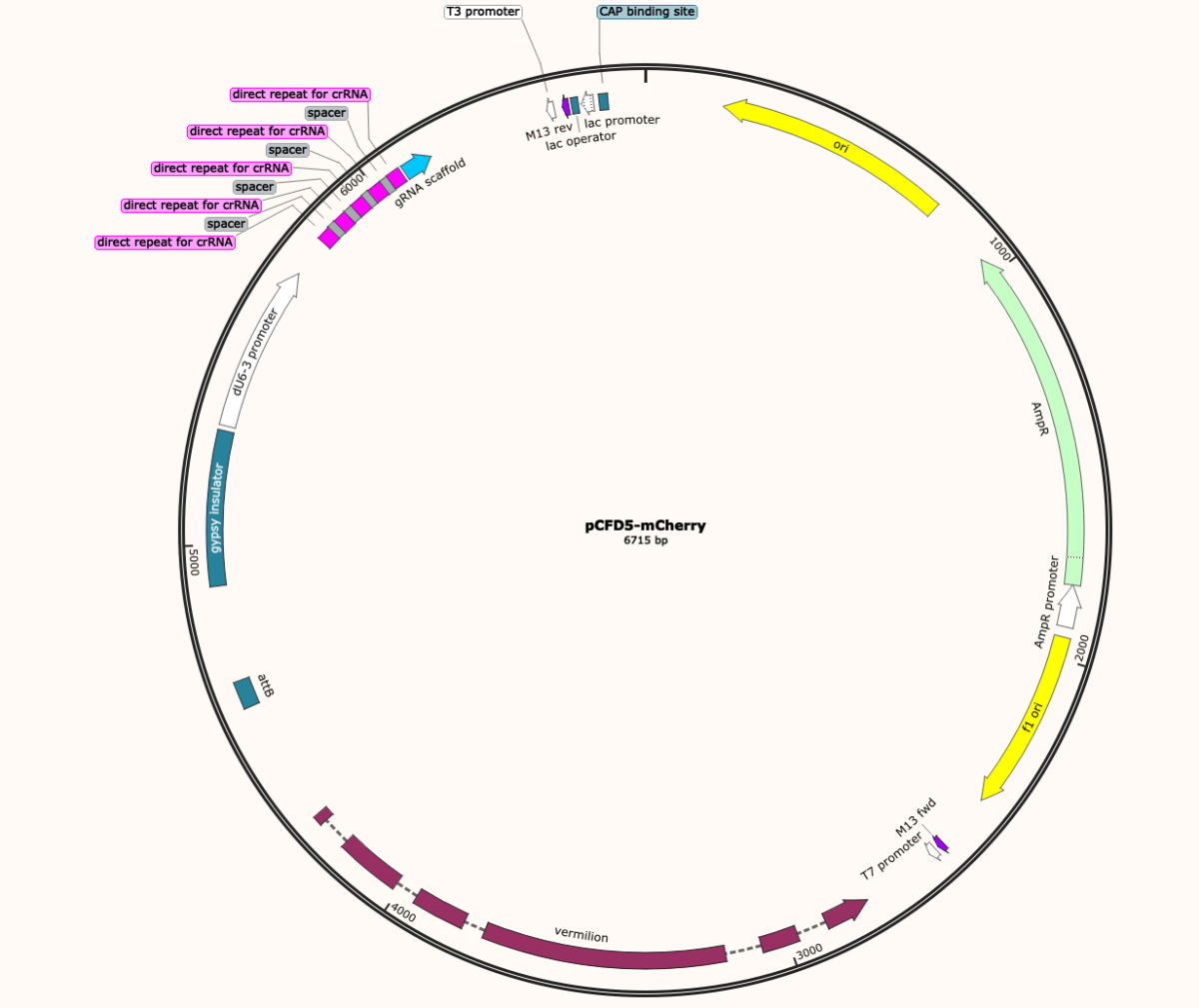


Figure 4.4 pCFD5-crRNA-mCherry-vermillion Expresses mCherry crRNA under the control of the U6:3 promoter. We digested pCFD5 (Addgene #73914) using BbsI, then used NEBuilder (see main methods) to insert the crRNA mCherry sequence which we amplified from the ultramer sequences using primers in appendix16. (Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

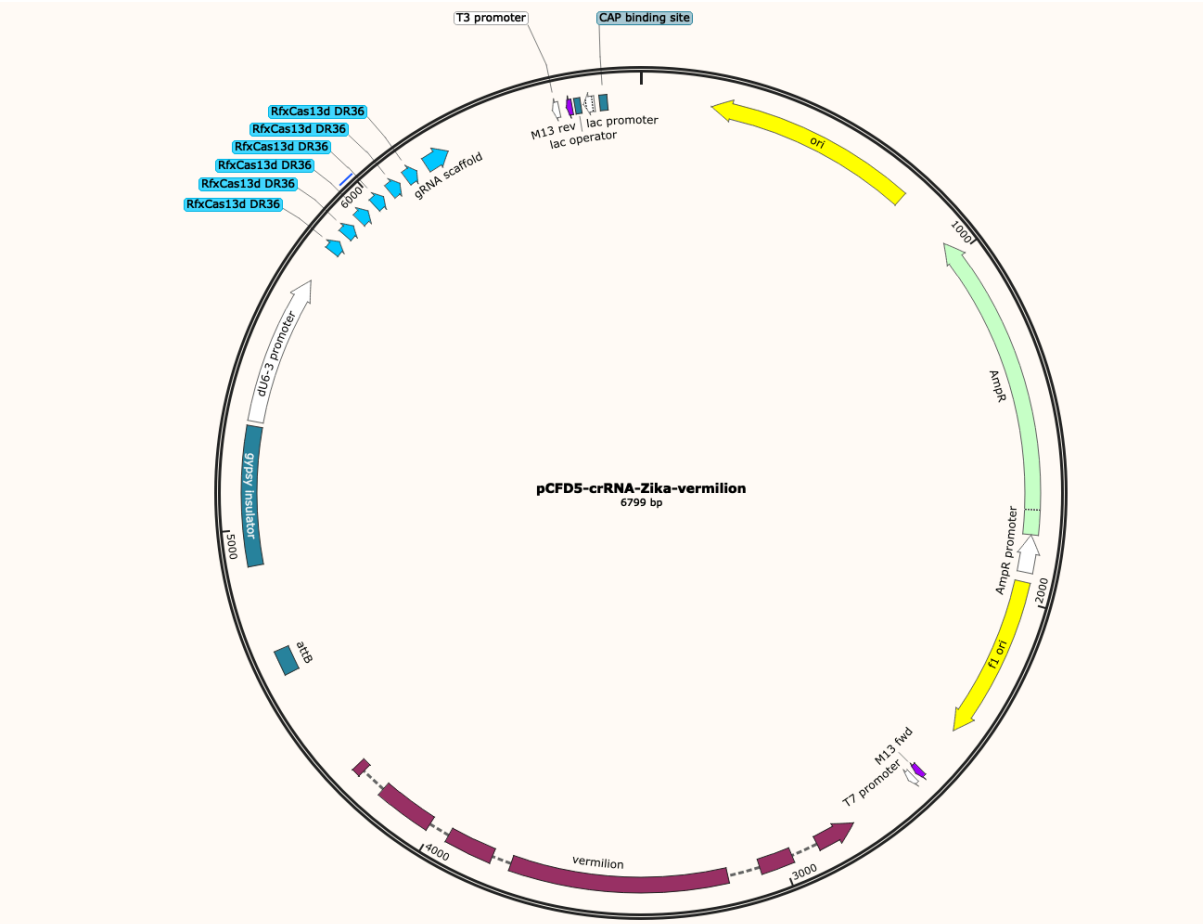


Figure 4.5 pCFD5-crRNA-Zika-vermillion Expresses Zika crRNA under the control of the U6:3 promoter. We digested pCFD5 (Addgene #73914) using BbsI, then used NEBuilder (see main methods) to insert the crRNA Zika sequence which we amplified from the ultramer sequences using primers in appendix 15.
(Plasmid sequences on benchling: https://benchling.com/obrosh/f_/zERxdTn2-thesis-plasmids/)

Fly Stocks

After preparing the plasmids, we sent the pUb-mCherry-Adm-miniwhite plasmid to BestGene Inc for microinjection into their VK13 stock (which expresses *vasa* driven phiC31, and is a phiC31 P[acman] (Venken et al., 2009) strain with an attP-9A insertion site on chromosome III, Bloomington 9732), the DAV and mCherry pCFD5 plasmids to the University of Cambridge Department of Genetics Fly Facility for microinjection into y, v, nos-int;attP40 (Bloomington 25709) (on chromosome II), and the Zika pCFD5 plasmid to BestGene Inc for microinjection onto chromosome 2 of their attP40 stock (which expresses *vasa* driven phiC31).

We crossed the Cas13 line with the crRNA line (cross scheme in Figure 4.6) and infected F1 generation flies with DAV (see main methods). We froze flies after 24±1 hours, 72±3 hours and 25 days using liquid nitrogen and stored them at -80°C in TRIzol until RNA extraction, reverse transcription, and measurement of viral titre using qPCR (see main methods).

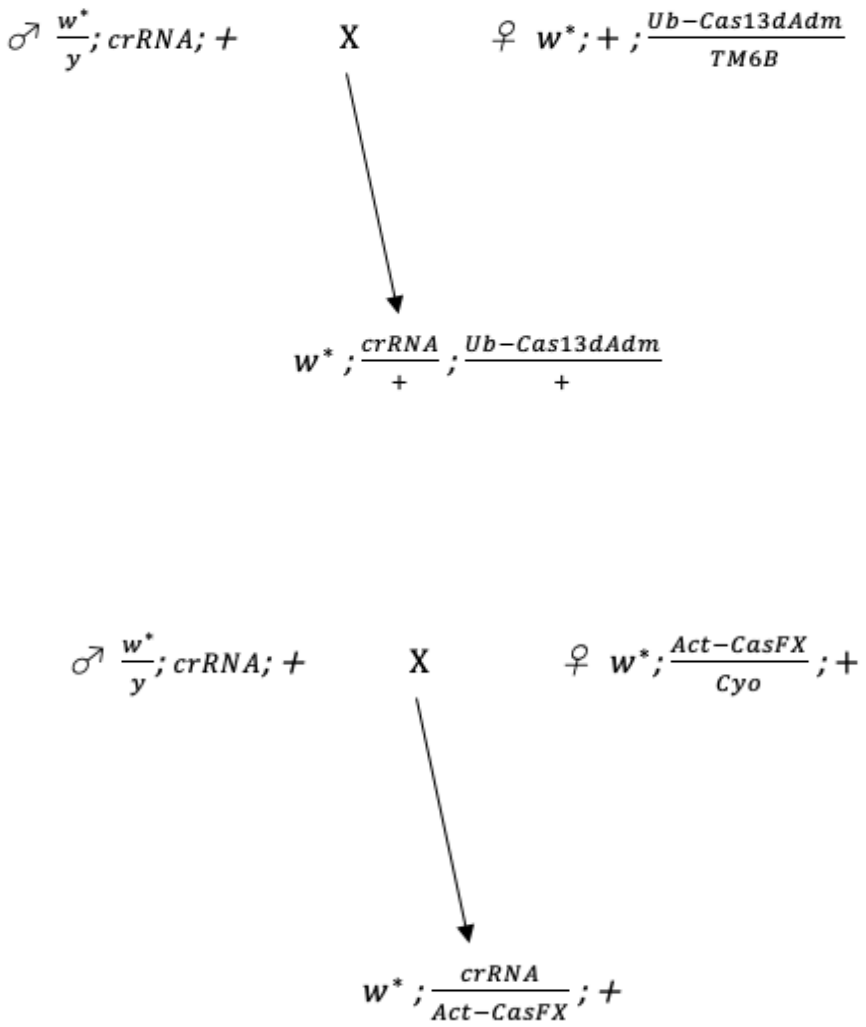


Figure 4.6 Crossing schemes for Cas13 Experiment. We used these crosses to produce flies expressing Cas13 and crRNA targeting DAV, mCherry, or ZIka.

Conclusion

We have set out to investigate the role that transposable elements play in the evolution of virus resistance. When TEs insert into the coding sequence of a gene, they can allow that gene to gain an antiviral function by altering its transcript sequence. We have explored two different ways this can occur.

In Chapter 1, we describe a *Doc* element insertion into the coding sequence of *Veneno,* which results in the gene gaining an antiviral function. This function seems to rely on the sequence of *Veneno* upstream of the insertion, and results in a truncated transcript which includes the coding sequence upstream of the insertion as well as 137 base pairs of *Doc* element origin (18 of which are translated). This truncated transcript is sufficient to produce DAV resistance.

In Chapter 2 we explore how exactly this occurs. We find that resistance does not rely on sequences from the *Doc* element, and that the truncation without the *Doc* element insertion is sufficient to cause resistance. Furthermore, we show that resistance does not require the Tudor domain or the MYND Zn finger domain expressed in the truncated transcript encoded by Ven^{Doc} to be functional, nor does it require the siRNA pathway (the main antiviral pathway in *Drosophila*) to be functional. We also find no evidence that it requires any of the several tested immune pathways (including piRNA).

In Chapter 3, we look into another *Doc* element insertion, this time into the coding sequence of *CHKov1,* resulting in allele $CHKov^{Doc1420}$. This second insertion causes resistance to DMelSV by also causing a truncated transcript. However, the *Doc* insertion into *CHKov1* results in a truncated transcript tr3 which encodes a gene-TE chimera by expressing the entire ORF1 of the *Doc* element, predicted to encode a nucleic acid (possibly RNA) binding function (O'Hare et al., 1991). The translated sequence from tr3 is sufficient to cause DMelSV resistance. We have not yet shown that the sequence of the ORF1 from the *Doc* element in tr3 is required for DMelSV resistance, but we will test this in a future experiment. If that is the case, the gene-TE chimera would have a novel

antiviral function which may make use of the *CHKov1* sequences upstream of the insertion, and the RNA binding function of ORF1.

We therefore propose two ways transcript truncations by TEs can cause novel functions and thereby affect the evolution of virus resistance: chimeric truncations which produce transcripts that contain sequences of TE origin, and rely on these sequences to cause their phenotype (such as *SETMAR* in primates (Lee et al., 2005)); non-chimeric truncations which either produce transcripts that do not include sequences of TE origin, or do not rely on these sequences to produce their phenotype, (such as the truncation in Ven^{Doc}).

Transposable elements are an important source of mutations, and are therefore frequently important in recent adaptations. This may especially be the case when novel selective pressures create the need for rapid adaptations (Schrader & Schmitz, 2019; Stapley et al., 2015); for example, transposable element insertions seem to be very important in the evolution of insecticide resistance (Rostant et al., 2012). Since viruses adapt quickly to host defences, which allows them to incur significant fitness costs on the host through infection, they provide a continuous source of novel selective pressure. Our results have shown that transposable element insertion can be useful in adapting to this pressure. However, in contrast to insertions causing insecticide resistance, which are often due to regulatory changes (such as the *Doc* insertion altering the expression of *Cyp6g1* leading to DDT resistance (Schlenke & Begun, 2004)) or loss of function (such as the *Hel-1* insertion in the HevCalP coding sequence which leads to Bt resistance (Gahan et al., 2001)), the insertions causing virus resistance seemed (in both cases) to do so through gain of function mutations which occur by altering the transcript of a host gene, leading to a novel transcription factor. It is possible that this is due to a mechanistic difference between virus resistance and insecticide resistance; for instance, viral resistance might arise through alterations in the protein-protein interactions of host proteins which interact with the virus during replication, whereas resistance to insecticides might occur from altering the expression of certain detoxification enzymes. However, it is also possible that this difference is simply due to chance.

Our discovery of a novel restriction factor encoded by *Ven^{Doc}* also adds to the growing list of restriction factors which cause virus resistance in *Drosophila*. These include *CHKov^{Doc1420}* (Magwire et al., 2011), as well as *Vago* (Deddouche et al., 2008), *ref(2)P* (Bangham et al., 2008), *pastrel* (Magwire et al., 2012), and *Ge-1* (Cao et al., 2016). Although a lot is known about conserved antiviral pathways such as RNAi (see Introduction: Antiviral Immunity in *Drosophila*), little is known about the mechanisms of these restriction factors, which seem to account for a large amount of the variation in virus resistance in natural populations (Magwire et al., 2012). *Ven^{Doc}* (which also has a very large effect size of ~19,000-fold reduction in viral titre) is yet another case which suggests that there might be many idiosyncratic restriction factors which target specific viruses, and that therefore the mechanisms of virus resistance are still largely unclear. Further investigation into these restriction factors is thus important to clarify virus resistance mechanisms in *Drosophila* (and perhaps in other species as well).

References

Abad, J. P., de Pablos, B., Osoegawa, K., de Jong, P. J., Martín-Gallardo, A., & Villasante, A. (2004a). Genomic analysis of Drosophila melanogaster telomeres: full-length copies of HeT-A and TART elements at telomeres. *Molecular Biology and Evolution*, *21*(9), 1613–1619. https://doi.org/10.1093/molbev/msh174

Abad, J. P., de Pablos, B., Osoegawa, K., de Jong, P. J., Martín-Gallardo, A., & Villasante, A. (2004b). TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. *Molecular Biology and Evolution*, *21*(9), 1620–1624. https://doi.org/10.1093/molbev/msh180

Agrawal, A., Eastman, Q. M., & Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*, *394*(6695), 744–751. https://doi.org/10.1038/29457

Almeida, L. M., Silva, I. T., Silva, W. A., Castro, J. P., Riggs, P. K., Carareto, C. M., & Amaral, M. E. J. (2007). The contribution of transposable elements to Bos taurus gene structure. *Gene*, *390*(1–2), 180–189. https://doi.org/10.1016/j.gene.2006.10.012

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., & LIPMAN, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ambrose, R. L., Lander, G. C., Maaty, W. S., Bothner, B., Johnson, J. E., & Johnson, K. N. (2009). Drosophila A virus is an unusual RNA virus with a T=3 icosahedral core and permuted RNA-dependent RNA polymerase. *The Journal of General Virology*, *90*(Pt 9), 2191–2200. https://doi.org/10.1099/vir.0.012104-0

Aminetzach, Y. T., Macpherson, J. M., & Petrov, D. A. (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science*, *309*(5735), 764–767. https://doi.org/10.1126/science.1112699

Anxolabéhère, D., Kidwell, M. G., & Periquet, G. (1988). Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. *Molecular Biology and Evolution*, *5*(3), 252–269. https://doi.org/10.1093/oxfordjournals.molbev.a040491

Aravind, L. (2000). The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends in Biochemical Sciences*, *25*(9), 421–423. https://doi.org/10.1016/s0968-0004(00)01620-0

Avadhanula, V., Weasner, B. P., Hardy, G. G., Kumar, J. P., & Hardy, R. W. (2009). A Novel System for the Launch of Alphavirus RNA Synthesis Reveals a Role for the Imd Pathway in Arthropod Antiviral Response. *PLoS Pathogens*, *5*(9). https://doi.org/10.1371/journal.ppat.1000582

Bae, J.-H., Ahn, K., Nam, G.-H., Lee, C.-E., Park, K.-D., Lee, H.-K., Cho, B.-W., & Kim, H.-S. (2011). Molecular characterization of alternative transcripts of the horse BMAL1 gene. *Zoological Science*, *28*(9), 671–675. https://doi.org/10.2108/zsj.28.671

Bangham, J., Kim, K.-W., Webster, C. L., & Jiggins, F. M. (2008). Genetic variation affecting host-parasite interactions: Different genes affect different aspects of sigma virus replication and transmission in Drosophila melanogaster. *Genetics*, *178*(4), 2191–2199. https://doi.org/10.1534/genetics.107.085449

Bangham, J., Knott, S. A., Kim, K.-W., Young, R. S., & Jiggins, F. M. (2008). Genetic variation affecting host-parasite interactions: major-effect quantitative trait loci affect the transmission of sigma virus in Drosophila melanogaster. *Molecular Ecology*, *17*(17), 3800–3807. https://doi.org/10.1111/j.1365-294X.2008.03873.x

Bangham, J., Obbard, D. J., Kim, K.-W., Haddrill, P. R., & Jiggins, F. M. (2007). The age and evolution of an antiviral resistance mutation in Drosophila melanogaster. *Proceedings of the Royal Society B-Biological Sciences*, *274*(1621), 2027–2034. https://doi.org/10.1098/rspb.2007.0611

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Battlay, P., Schmidt, J. M., Fournier-Level, A., & 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), 2573–2581. https://doi.org/10.1534/g3.116.031054

Beck, C. R., Garcia-Perez, J. L., Badge, R. M., & Moran, J. v. (2011). LINE-1 elements in structural variation and disease. *Annual Review of Genomics and Human Genetics*, *12*(1), 187–215. https://doi.org/10.1146/annurev-genom-082509-141802

Belbin, O., Carrasquillo, M. M., Crump, M., Culley, O. J., Hunter, T. A., Ma, L., Bisceglio, G., Zou, F., Allen, M., Dickson, D. W., Graff-Radford, N. R., Petersen, R. C., Morgan, K., & Younkin, S. G. (2011). Investigation of 15 of the top candidate genes for late-onset Alzheimer's disease. *Human Genetics*, *129*(3), 273–282. https://doi.org/10.1007/s00439-010-0924-2

Belenkaya, T., Soldatov, A., Nabirochkina, E., Biryukova, I., Birjukova, I., Georgieva, S., & Georgiev, P. (1998). P-Element insertion at the polyhomeotic gene leads to formation of a novel chimeric protein that negatively regulates yellow gene expression in P-element-induced alleles of Drosophila melanogaster. *Genetics*, *150*(2), 687–697. /pmc/articles/PMC1460360/?report=abstract

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., Hoskins, R. A., & Spradling, A. C. (2011). The Drosophila gene disruption project: progress using transposons with distinctive site specificities. *Genetics*, *188*(3), 731–743. https://doi.org/10.1534/genetics.111.126995

Berezikov, E., Bucheton, A., & Busseau, I. (2000). A search for reverse transcriptase-coding sequences revealsnew non-LTR retrotransposons in the genome of Drosophilamelanogaster. *Genome Biology*, *1*(6). http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcAuth=mekentosj&SrcApp=Papers&DestLinkType=FullRecord&DestApp=WOS&KeyUT=000207583400007

Bitto, A., Lerner, C. A., Nacarelli, T., Crowe, E., Torres, C., & Sell, C. (2014). p62/SQSTM1 at the interface of aging, autophagy, and disease. *Age*, *36*(3). https://doi.org/10.1007/s11357-014-9626-3

Boden, D., Pusch, O., Silbermann, R., Lee, F., Tucker, L., & Ramratnam, B. (2004). Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. *Nucleic Acids Research*, *32*(3), 1154–1158. https://doi.org/10.1093/nar/gkh278

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1), 112–199. https://doi.org/10.1186/s13059-018-1577-z

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, *128*(6), 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*, *322*(5906), 1387–1392. https://doi.org/10.1126/science.1165171

Britten, R. (2006). Transposable elements have contributed to thousands of human proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(6), 1798–1803. https://doi.org/10.1073/pnas.0510007103

Bronkhorst, A. W., & van Rij, R. P. (2014). The long and short of antiviral defense: small RNA-based immunity in insects. *Current Opinion in Virology*, *7*, 19–28. https://doi.org/10.1016/j.coviro.2014.03.010

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., Wan, K. H., Yu, C., Zhang, D., Carlson, J. W., Cherbas, L., Eads, B. D., Miller, D., Mockaitis, K., Roberts, J., … Celniker, S. E. (2014). Diversity and dynamics of the Drosophila transcriptome. *Nature*, *512*(7515). https://doi.org/10.1038/nature12962

Brun, G., & Plus, N. (1980). The viruses of Drosophila. In M. Ashburner & T. R. F. Wright (Eds.), *The Genetics and Biology of Drosophila* (pp. 625–702). Academic Press.

Bureau, T. E., Ronald, P. C., & Wessler, S. R. (1996). A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(16), 8524–8529. https://doi.org/10.1073/pnas.93.16.8524

BURT, A., & Trivers, R. (2009). *Genes in Conflict*. Harvard University Press. http://books.google.co.uk/books?id=C-aSjMj8FK8C&printsec=frontcover&dq=intitle:Genes+in+Conflict&hl=&cd=1&source=gbs_api

Cao, C., Cogni, R., Barbier, V., & Jiggins, F. M. (2017). Complex Coding and Regulatory Polymorphisms in a Restriction Factor Determine the Susceptibility of Drosophila to Viral Infection. *Genetics*, *206*(4), 2159–2173. https://doi.org/10.1534/genetics.117.201970

Cao, C., Magwire, M. M., Bayer, F., & Jiggins, F. M. (2016). A Polymorphism in the Processing Body Component Ge-1 Controls Resistance to a Naturally Occurring Rhabdovirus in Drosophila. *PLoS Pathogens*, *12*(1), e1005387. https://doi.org/10.1371/journal.ppat.1005387

Carpenter, J. A., Obbard, D. J., Maside, X., & Jiggins, F. M. (2007). The recent spread of a vertically transmitted virus through populations of Drosophila melanogaster. *Molecular Ecology*, *16*(18), 3947–3954. https://doi.org/10.1111/j.1365-294X.2007.03460.x

Carrau, T., Hiebert, N., Vilcinskas, A., & Lee, K.-Z. (2018). Identification and characterization of natural viruses associated with the invasive insect pest Drosophila suzukii. *Journal of Invertebrate Pathology*, *154*, 74–78. https://doi.org/10.1016/j.jip.2018.04.001

Castro, M. C., Wilson, M. E., & Bloom, D. E. (2017). Disease and economic burdens of dengue. *The Lancet. Infectious Diseases*, *17*(3), e70–e78. https://doi.org/10.1016/S1473-3099(16)30545-X

Chain, A. C., Zollman, S., Tseng, J. C., & Laski, F. A. (1991). Identification of a cis-acting sequence required for germ line-specific splicing of the P element ORF2-ORF3 intron. *Molecular and Cellular Biology*, *11*(3), 1538–1546. https://doi.org/10.1128/mcb.11.3.1538

Chao, J. A., Lee, J. H., Chapados, B. R., Debler, E. W., Schneemann, A., & Williamson, J. R. (2005). Dual modes of RNA-silencing suppression by Flock House virus protein B2. *Nature Structural & Molecular Biology*, *12*(11), 952–957. https://doi.org/10.1038/nsmb1005

Charlesworth, B., & Langley, C. H. (1986). The evolution of self-regulated transposition of transposable elements. *Genetics*, *112*(2). https://doi.org/10.1093/genetics/112.2.359

Chintapalli, V. R., Wang, J., & Dow, J. A. T. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. In *Nature Genetics* (Vol. 39, Issue 6). https://doi.org/10.1038/ng2049

Chiu, Y.-L., & Greene, W. C. (2008). The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annual Review of Immunology*, *26*(1), 317–353. https://doi.org/10.1146/annurev.immunol.26.021607.090350

Cogni, R., Cao, C., Day, J. P., Bridson, C., & Jiggins, F. M. (2016). The genetic architecture of resistance to virus infection in Drosophila. *Molecular Ecology*, *25*(20), 5228–5241. https://doi.org/10.1111/mec.13769

CONTURSI, C., MINCHIOTTI, G., & DINOCERA, P. P. (1995). Identification of Sequences Which Regulate the Expression of Drosophila-Melanogaster Doc Elements. *Journal of Biological Chemistry*, *270*(44), 26570–26576. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=7592878&retmode=ref&cmd=prlinks

Cordaux, R., Udit, S., Batzer, M. A., & Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8101–8106. https://doi.org/10.1073/pnas.0601161103

Costa, A., Jan, E., Sarnow, P., & Schneider, D. (2009). The Imd pathway is involved in antiviral immune responses in Drosophila. *PloS One*, *4*(10), e7436. https://doi.org/10.1371/journal.pone.0007436

Cross, S. T., Maertens, B. L., Dunham, T. J., Rodgers, C. P., Brehm, A. L., Miller, M. R., Williams, A. M., Foy, B. D., & Stenglein, M. D. (2020). Partitiviruses Infecting Drosophila melanogaster and Aedes aegypti Exhibit Efficient Biparental Vertical Transmission. *Journal of Virology*, *94*(20). https://doi.org/10.1128/JVI.01070-20

Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J., & Brennecke, J. (2008). An endogenous small interfering RNA pathway in Drosophila. *Nature*, *453*(7196), 798–802. https://doi.org/10.1038/nature07007

Daborn, P. J., Boundy, S., Yen, J., Pittendrigh, B., & Ffrench-Constant, R. (2001). DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the

neonicotinoid imidacloprid. *Molecular Genetics and Genomics : MGG*, *266*(4), 556–563. https://doi.org/10.1007/s004380100531

Daborn, P. J., Yen, J. L., Bogwitz, M. R., le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., Feyereisen, R., Wilson, T. G., & ffrench-Constant, R. H. (2002). A single P450 allele associated with insecticide resistance in Drosophila. *Science*, *297*(5590), 2253–2256. https://doi.org/10.1126/science.1074170

Dawson, A., Hartswood, E., Paterson, T., & Finnegan, D. J. (1997). A LINE-like transposable element in Drosophila, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. *Embo Journal*, *16*(14), 4448–4455. https://doi.org/10.1093/emboj/16.14.4448

Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D., Dostert, C., Antoniewski, C., Hoffmann, J. A., & Imler, J.-L. (2008). The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in drosophila. *Nature Immunology*, *9*(12), 1425–1432. https://doi.org/10.1038/ni.1664

Derr, L. K., & Strathern, J. N. (1993). A role for reverse transcripts in gene conversion. *Nature*, *361*(6408), 170–173. https://doi.org/10.1038/361170a0

Devine, S. E., & Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes & Development*, *10*(5), 620–633. https://doi.org/10.1101/gad.10.5.620

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., & Dickson, B. J. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature*, *448*(7150). https://doi.org/10.1038/nature05954

Dimmock, N. J., Easton, A. J., & Leppard, K. N. (2007). *Introduction to Modern Virology* (6th ed.). Blackwell Publishing Ltd.

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1077-+. https://doi.org/10.1126/science.1258096

Driver, A., Lacey, S. F., Cullingford, T. E., Mitchelson, A., & O'Hare, K. (1989). Structural analysis of Doc transposable elements associated with mutations at the white and suppressor of forked loci of Drosophila melanogaster. *Molecular & General Genetics : MGG*, *220*(1), 49–52. https://doi.org/10.1007/BF00260854

Duxbury, E. M., Day, J. P., Maria Vespasiani, D., Thüringer, Y., Tolosana, I., Smith, S. C., Tagliaferri, L., Kamacioglu, A., Lindsley, I., Love, L., Unckless, R. L., Jiggins, F. M., & Longdon, B. (2019). Host-pathogen coevolution increases genetic variation in susceptibility to infection. *ELife*, *8*. https://doi.org/10.7554/eLife.46440

Eggleston, W. B., Johnson-Schlitz, D. M., & Engels, W. R. (1988). P-M hybrid dysgenesis does not mobilize other transposable element families in D. melanogaster. *Nature*, *331*(6154), 368–370. https://doi.org/10.1038/331368a0

Elrod-Erickson, M., Mishra, S., & Schneider, D. (2000). Interactions between the cellular and humoral immune responses in Drosophila. *Current Biology*, *10*(13), 781–784. https://doi.org/10.1016/s0960-9822(00)00569-8

Engels, W. R. (1992). The origin of P elements in Drosophila melanogaster. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *14*(10), 681–686. https://doi.org/10.1002/bies.950141007

Engels, W. R., & Preston, C. R. (1984). Formation of chromosome rearrangements by P factors in Drosophila. *Genetics*, *107*(4), 657–678. /pmc/articles/PMC1202383/?report=abstract

FEDOROFF, N. V. . (1983). Controlling Elements in Maize. In J. A. . Shapiro (Ed.), *Mobile Genetic Elements* (pp. 1–63). Academic Press.

Ferreira, Á. G., Naylor, H., Esteves, S. S., Pais, I. S., Martins, N. E., & Teixeira, L. (2014). The Toll-dorsal pathway is required for resistance to viral oral infection in Drosophila. *PLoS Pathogens*, *10*(12), e1004507. https://doi.org/10.1371/journal.ppat.1004507

Ferrigno, O., Virolle, T., Djabari, Z., Ortonne, J. P., White, R. J., & Aberdam, D. (2001). Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nature Genetics*, *28*(1), 77–81. https://doi.org/10.1038/ng0501-77

ffrench-Constant, R., Daborn, P., & Feyereisen, R. (2006). Resistance and the jumping gene. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *28*(1), 6–8. https://doi.org/10.1002/bies.20354

Franz, A. W. E., Sanchez-Vargas, I., Adelman, Z. N., Blair, C. D., Beaty, B. J., James, A. A., & Olson, K. E. (2006). Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(11), 4198–4203. https://doi.org/10.1073/pnas.0600479103

Gahan, L. J., Gould, F., & Heckel, D. G. (2001). Identification of a gene associated with Bt resistance in Heliothis virescens. *Science (New York, N.Y.)*, *293*(5531). https://doi.org/10.1126/science.1060949

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A., & Imler, J.-L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. *Nature Immunology*, *7*(6), 590–597. https://doi.org/10.1038/ni1335

García Guerreiro, M. P. (2012). What makes transposable elements move in the Drosophila genome? *Heredity*, *108*(5), 461–468. https://doi.org/10.1038/hdy.2011.89

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L. W., Zapp, M. L., Weng, Z., & Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science*, *320*(5879), 1077–1081. https://doi.org/10.1126/science.1157396

Gilbert, M. K., Tan, Y. Y., & Hart, C. M. (2006). The Drosophila boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics*, *173*(3), 1365–1375. https://doi.org/10.1534/genetics.106.056002

Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199–202. https://doi.org/10.1038/nature04917

Goodier, J. L., Ostertag, E. M., & Kazazian, H. H. (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Human Molecular Genetics*, *9*(4), 653–657. https://doi.org/10.1093/hmg/9.4.653

Gotea, V., & Makałowski, W. (2006). Do transposable elements really contribute to proteomes? *Trends in Genetics : TIG*, *22*(5), 260–267. https://doi.org/10.1016/j.tig.2006.03.006

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G., van Baren, M. J., Boley, N., Booth, B. W., Brown, J. B., Cherbas, L., Davis, C. A., Dobin, A., Li, R., Lin, W., Malone, J. H., Mattiuzzo, N. R., Miller, D., ... Celniker, S. E. (2011). The

developmental transcriptome of Drosophila melanogaster. *Nature*, *471*(7339). https://doi.org/10.1038/nature09715

Gray, Y. H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends in Genetics : TIG*, *16*(10), 461–468. https://doi.org/10.1016/s0168-9525(00)02104-1

Guruharsha, K. G., Rual, J.-F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D. Y., Cenaj, O., McKillip, E., Shah, S., Stapleton, M., Wan, K. H., Yu, C., Parsa, B., Carlson, J. W., Chen, X., Kapadia, B., … Artavanis-Tsakonas, S. (2011). A protein complex network of Drosophila melanogaster. *Cell*, *147*(3), 690–703. https://doi.org/10.1016/j.cell.2011.08.047

Ha, H.-S., Moon, J.-W., Gim, J.-A., Jung, Y.-D., Ahn, K., Oh, K.-B., Kim, T.-H., Seong, H.-H., & Kim, H.-S. (2012). Identification and characterization of transposable element-mediated chimeric transcripts from porcine Refseq and EST databases. *Genes & Genomics*, *34*(4), 409–414. https://doi.org/10.1007/s13258-011-0212-0

Habayeb, M. S., Cantera, R., Casanova, G., Ekström, J.-O., Albright, S., & Hultmark, D. (2009). The Drosophila Nora virus is an enteric virus, transmitted via feces. *Journal of Invertebrate Pathology*, *101*(1), 29–33. https://doi.org/10.1016/j.jip.2009.02.003

Haig, D. (2016). Transposable elements: Self-seekers of the germline, team-players of the soma. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *38*(11), 1158–1166. https://doi.org/10.1002/bies.201600125

Han, J. S. (2010). Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mobile DNA*, *1*(1), 12–15. https://doi.org/10.1186/1759-8753-1-15

Herold, N., Will, C. L., Wolf, E., Kastner, B., Urlaub, H., & Lührmann, R. (2009). Conservation of the Protein Composition and Electron Microscopy Structure of Drosophila melanogaster and Human Spliceosomal Complexes . *Molecular and Cellular Biology*, *29*(1). https://doi.org/10.1128/mcb.01415-08

Hiom, K., Melek, M., & Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell*, *94*(4), 463–470. https://doi.org/10.1016/s0092-8674(00)81587-1

Hoffmann, J. A. (2003). The immune response of Drosophila. *Nature*, *426*(6962), 33–38. https://doi.org/10.1038/nature02021

Hohjoh, H., & Singer, M. F. (1997). Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *Embo Journal*, *16*(19), 6034–6043. https://doi.org/10.1093/emboj/16.19.6034

Houle, D., & Nuzhdin, S. v. (2004). Mutation accumulation and the effect of copia insertions in Drosophila melanogaster. *Genetical Research*, *83*(1), 7–18. https://doi.org/10.1017/s0016672303006505

Huang, Z., Kingsolver, M. B., Avadhanula, V., & Hardy, R. W. (2013). An antiviral role for antimicrobial peptides during the arthropod response to alphavirus replication. *Journal of Virology*, *87*(8), 4272–4280. https://doi.org/10.1128/JVI.03360-12

Huszar, T., & Imler, J.-L. (2008). Drosophila viruses and the study of antiviral host-defense. *Advances in Virus Research*, *72*, 227–265. https://doi.org/10.1016/S0065-3527(08)00406-5

Huynh, N., Depner, N., Larson, R., & King-Jones, K. (2020). A versatile toolkit for CRISPR-Cas13-based RNA manipulation in Drosophila. *Genome Biology*, *21*(1), 29. https://doi.org/10.1186/s13059-020-02193-y

Jiang, Y. W. (2002). Transcriptional cosuppression of yeast Ty1 retrotransposons. *Genes & Development*, *16*(4), 467–478. https://doi.org/10.1101/gad.923502

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Joosten, J., Miesen, P., Pennings, B., Jansen, P. W. T. C., Huynen, M. A., Vermeulen, M., & van Rij, R. P. (2018). The Tudor protein Veneno assembles the ping-pong amplification complex that produces viral piRNAs in Aedesmosquitoes. *BioRxiv*, 1–25. https://doi.org/10.1101/242305

Jordan, I. K., Rogozin, I. B., Glazko, G. v, & Koonin, E. v. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics : TIG*, *19*(2), 68–72. https://doi.org/10.1016/s0168-9525(02)00006-9

Juhasz, G., & Neufeld, T. P. (2006). Autophagy: a forty-year search for a missing membrane source. *PLoS Biology*, *4*(2), e36. https://doi.org/10.1371/journal.pbio.0040036

Jupatanakul, N., Sim, S., Anglero-Rodriguez, Y. I., Souza-Neto, J., Das, S., Poti, K. E., Rossi, S. L., Bergren, N., Vasilakis, N., & Dimopoulos, G. (2017). Engineered Aedes aegypti JAK/STAT Pathway-Mediated Immunity to Dengue Virus. *Plos Neglected Tropical Diseases*, *11*(1). https://doi.org/10.1371/journal.pntd.0005187

Kemp, C., Mueller, S., Goto, A., Barbier, V., Paro, S., Bonnay, F., Dostert, C., Troxler, L., Hetru, C., Meignin, C., Pfeffer, S., Hoffmann, J. A., & Imler, J.-L. (2013). Broad RNA interference-mediated antiviral immunity and virus-specific inducible responses in Drosophila. *Journal of Immunology (Baltimore, Md. : 1950)*, *190*(2), 650–658. https://doi.org/10.4049/jimmunol.1102486

Kennedy, E. M., & Cullen, B. R. (2017). Gene Editing: A New Tool for Viral Disease. *Annual Review of Medicine*, *68*, 401–411. https://doi.org/10.1146/annurev-med-051215-031129

Khurana, J. S., Xu, J., Weng, Z., & Theurkauf, W. E. (2010). Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. *PLoS Genetics*, *6*(12), e1001246. https://doi.org/10.1371/journal.pgen.1001246

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M. G., & Katsuma, S. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature*, *509*(7502), 633–636. https://doi.org/10.1038/nature13315

Kofler, R., Nolte, V., & Schlötterer, C. (2015). Tempo and Mode of Transposable Element Activity in Drosophila. *PLoS Genetics*, *11*(7), e1005406. https://doi.org/10.1371/journal.pgen.1005406

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., & Hsu, P. D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell*, *173*(3), 665-+. https://doi.org/10.1016/j.cell.2018.02.033

Koonin, E. v, & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *374*(1772), 20180087. https://doi.org/10.1098/rstb.2018.0087

Kronhamn, J., & Rasmuson-Lestander, A. (1999). Genetic organization of the ci-M-pan region on chromosome IV in Drosophila melanogaster. *Genome*, *42*(6), 1144–1149. https://doi.org/10.1139/g99-085

Lamiable, O., Arnold, J., de Faria, I. J. da S., Olmo, R. P., Bergami, F., Meignin, C., Hoffmann, J. A., Marques, J. T., & Imler, J.-L. (2016). Analysis of the Contribution of Hemocytes and Autophagy to Drosophila Antiviral Immunity. *Journal of Virology*, *90*(11), 5415–5426. https://doi.org/10.1128/JVI.00238-16

Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J., & Robertson, H. M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(20), 11428–11433. https://doi.org/10.1073/pnas.96.20.11428

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. v, Goodman, J. L., Gramates, L. S., Millburn, G., Strelets, V. B., Tabone, C. J., Thurmond, J., Perrimon, N., Gelbart, S. R., Agapite, J., Broll, K., Crosby, M., dos Santos, G., Falls, K., … Lovato, T. (2021). FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Research*, *49*(D1). https://doi.org/10.1093/nar/gkaa1026

Lasko, P. (2010). Tudor domain. *Current Biology*, *20*(16), R666–R667. https://doi.org/10.1016/j.cub.2010.05.056

Lazzaro, B. P., Lazzaro, B. P., Sceurman, B. K., & Clark, A. G. (2004). Genetic Basis of Natural Variation in D. melanogaster Antibacterial Immunity. *Science*, *303*(5665), 1873–1876. https://doi.org/10.1126/science.1092447

Lee, S.-H., Oshige, M., Durant, S. T., Rasila, K. K., Williamson, E. A., Ramsey, H., Kwan, L., Nickoloff, J. A., & Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(50), 18075–18080. https://doi.org/10.1073/pnas.0503676102

Lemaitre, B., & Hoffmann, J. (2007). The host defense of Drosophila melanogaster. *Annual Review of Immunology*, *25*(1), 697–743. https://doi.org/10.1146/annurev.immunol.25.022106.141615

Levin, H. L., & Moran, J. v. (2011). Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, *12*(9), 615–627. https://doi.org/10.1038/nrg3030

Levy, D. E., & Darnell, J. E. (2002). STATs: Transcriptional control and biological impact. *Nature Reviews. Molecular Cell Biology*, *3*(9), 651–662. https://doi.org/10.1038/nrm909

Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., Sharma, P. P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D. H., Zamore, P. D., Miska, E. A., Sarkies, P., & Jiggins, F. M. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nature Ecology & Evolution*, *2*(1), 174–181. https://doi.org/10.1038/s41559-017-0403-4

Li, H., Li, W.-X., & Ding, S.-W. (2002). Induction and suppression of RNA silencing by an animal virus. *Science*, *296*(5571), 1319–1321. https://doi.org/10.1126/science.1070948

Liao, J.-F., Wu, C.-P., Tang, C.-K., Tsai, C.-W., Rouhová, L., & Wu, Y.-L. (2019). Identification of Regulatory Host Genes Involved in Sigma Virus Replication Using RNAi Knockdown in Drosophila. *Insects*, *10*(10), 339. https://doi.org/10.3390/insects10100339

Lipatov, M., Lenkov, K., Petrov, D. A., & Bergman, C. M. (2005). Paucity of chimeric gene-transposable element transcripts in the Drosophila melanogaster genome. *BMC Biology*, *3*(1), 18–24. https://doi.org/10.1186/1741-7007-3-24

Liu, B., Behura, S. K., Clem, R. J., Schneemann, A., Becnel, J., Severson, D. W., & Zhou, L. (2013). P53-mediated rapid induction of apoptosis conveys resistance to viral infection in Drosophila melanogaster. *PLoS Pathogens*, *9*(2), e1003137. https://doi.org/10.1371/journal.ppat.1003137

Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y., & Chalmers, R. (2007). The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. *Molecular and Cellular Biology*, *27*(3), 1125–1132. https://doi.org/10.1128/MCB.01899-06

Liu, H., Wang, J.-Y. S., Huang, Y., Li, Z., Gong, W., Lehmann, R., & Xu, R.-M. (2010). Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes & Development*, *24*(17), 1876–1881. https://doi.org/10.1101/gad.1956010

Liu, X., Hao, R., Chen, S., Guo, D., & Chen, Y. (2015). Inhibition of hepatitis B virus by the CRISPR/Cas9 system via targeting the conserved regions of the viral genome. *The Journal of General Virology*, *96*(8), 2252–2261. https://doi.org/10.1099/vir.0.000159

Liu, Y., Chen, W., Gaudet, J., Cheney, M. D., Roudaia, L., Cierpicki, T., Klet, R. C., Hartman, K., Laue, T. M., Speck, N. A., & Bushweller, J. H. (2007). Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity. *Cancer Cell*, *11*(6), 483–497. https://doi.org/10.1016/j.ccr.2007.04.010

Lockton, S., & Gaut, B. S. (2009). The contribution of transposable elements to expressed coding sequence in Arabidopsis thaliana. *Journal of Molecular Evolution*, *68*(1), 80–89. https://doi.org/10.1007/s00239-008-9190-5

Lohe, A. R., & Hartl, D. L. (1996). Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. *Molecular Biology and Evolution*, *13*(4), 549–555. https://doi.org/10.1093/oxfordjournals.molbev.a025615

Longdon, B., Wilfert, L., & Jiggins, F. M. (2012). The Sigma viruses of Drosophila. In R. Dietzgen & I. Kuzmin (Eds.), *Rhabdoviruses: Molecular Taxonomy, Evolution, Genomics, Ecology, Cytopathology and Control*.

Lorenc, A., & Makałowski, W. (2003). Transposable elements and vertebrate protein diversity. *Genetica*, *118*(2–3), 183–191. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12868608&retmode=ref&cmd=prlinks

Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science*, *302*(5649), 1401–1404. https://doi.org/10.1126/science.1089370

Lyttle, T. W., & Haymer, D. S. (1992). The role of the transposable element hobo in the origin of endemic inversions in wild populations of Drosophila melanogaster. *Genetica*, *86*(1–3), 113–126. https://doi.org/10.1007/bf00133715

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt, R. R. H., Barrón, M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., … Gibbs, R. A. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature*, *482*(7384), 173–178. https://doi.org/10.1038/nature10811

Magwire, M. M., Bayer, F., Webster, C. L., Cao, C., & 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), e1002337-11. https://doi.org/10.1371/journal.pgen.1002337

Magwire, M. M., Fabian, D. K., Schweyen, H., Cao, C., Longdon, B., Bayer, F., & 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), e1003057. https://doi.org/10.1371/journal.pgen.1003057

Majzoub, K. (2013). *The antiviral siRNA interactome in Drosophila melanogaster*.

Martinez, J., Cogni, R., Cao, C., Smith, S., Illingworth, C. J. R., & Jiggins, F. M. (2016). Addicted? Reduced host resistance in populations with defensive symbionts. *Proceedings of the Royal Society B-Biological Sciences*, *283*(1833). https://doi.org/10.1098/rspb.2016.0778

Martins, M. M., Medronho, R. D. A., & Cunha, A. J. L. A. da. (2021). Zika virus in Brazil and worldwide: a narrative review. *Paediatrics and International Child Health*, *41*(1). https://doi.org/10.1080/20469047.2020.1776044

Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z., & Izsvák, Z. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature Genetics*, *41*(6), 753–761. https://doi.org/10.1038/ng.343

Matthews, J. M., & Sunde, M. (2002). Zinc fingers--folds for many occasions. *IUBMB Life*, *54*(6). https://doi.org/10.1080/15216540216035

McCaffrey, A. P., Nakai, H., Pandey, K., Huang, Z., Salazar, F. H., Xu, H., Wieland, S. F., Marion, P. L., & Kay, M. A. (2003). Inhibition of hepatitis B virus in mice by RNA interference. *Nature Biotechnology*, *21*(6), 639–644. https://doi.org/10.1038/nbt824

McCarrey, J. R., & Thomas, K. (1987). Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature*, *326*(6112), 501–505. https://doi.org/10.1038/326501a0

MINCHIOTTI, G., CONTURSI, C., & DINOCERA, P. P. (1997). Multiple downstream promoter modules regulate thetranscription of the Drosophila melanogaster I, Doc and Felements. *Journal of Molecular Biology*, *267*(1), 37–46. https://doi.org/10.1006/jmbi.1996.0860

Mondotte, J. A., & Saleh, M.-C. (2018). Antiviral Immune Response and the Route of Infection in Drosophila melanogaster. *Advances in Virus Research*, *100*, 247–278. https://doi.org/10.1016/bs.aivir.2017.10.006

Morazzani, E. M., Wiley, M. R., Murreddu, M. G., Adelman, Z. N., & Myles, K. M. (2012). Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathogens*, *8*(1), e1002470. https://doi.org/10.1371/journal.ppat.1002470

Morgan, T. H. (1910). SEX LIMITED INHERITANCE IN DROSOPHILA. *Science*, *32*(812), 120–122. https://doi.org/10.1126/science.32.812.120

Moy, R. H., Gold, B., Molleston, J. M., Schad, V., Yanger, K., Salzano, M.-V., Yagi, Y., Fitzgerald, K. A., Stanger, B. Z., Soldan, S. S., & Cherry, S. (2014). Antiviral autophagy restrictsRift Valley fever virus infection and is conserved from flies to mammals. *Immunity*, *40*(1), 51–65. https://doi.org/10.1016/j.immuni.2013.10.020

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G. G., & Münk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *Journal of Biological Chemistry*, *281*(31), 22161–22172. https://doi.org/10.1074/jbc.M601716200

Nakamoto, M., Moy, R. H., Xu, J., Bambina, S., Yasunaga, A., Shelly, S. S., Gold, B., & Cherry, S. (2012). Virus Recognition by Toll-7 Activates Antiviral Autophagy in Drosophila. *Immunity*, *36*(4), 658–667. https://doi.org/10.1016/j.immuni.2012.03.003

Nassif, N., Penney, J., Pal, S., Engels, W. R., & Gloor, G. B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and Cellular Biology*, *14*(3), 1613–1625. https://doi.org/10.1128/mcb.14.3.1613

Nekrutenko, A., & Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics : TIG*, *17*(11), 619–621. https://doi.org/10.1016/s0168-9525(01)02445-3

Nigumann, P., Redik, K., Mätlik, K., & Speek, M. (2002). Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics*, *79*(5), 628–634. https://doi.org/10.1006/geno.2002.6758

Nikitin, A. G., & Woodruff, R. C. (1995). Somatic movement of the mariner transposable element and lifespan of Drosophila species. *Mutation Research*, *338*(1–6), 43–49. https://doi.org/10.1016/0921-8734(95)00010-4

O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M., & Doudna, J. A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*, *516*(7530), 263–266. https://doi.org/10.1038/nature13769

O'Hare, K., Alley, M. R., Cullingford, T. E., Driver, A., & Sanderson, M. J. (1991). DNA sequence of the Doc retroposon in the white-one mutant of Drosophila melanogaster and of secondary insertions in the phenotypically altered derivatives white-honey and white-eosin. *Molecular & General Genetics : MGG*, *225*(1), 17–24. https://doi.org/10.1007/BF00282637

Olson, K. E., Higgs, S., Gaines, P. J., Powers, A. M., Davis, B. S., Kamrud, K. I., Carlson, J. O., Blair, C. D., & Beaty, B. J. (1996). Genetically engineered resistance to dengue-2 virus transmission in mosquitoes. *Science*, *272*(5263), 884–886. https://doi.org/10.1126/science.272.5263.884

Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., & Zamore, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics*, *20*(2), 89–108. https://doi.org/10.1038/s41576-018-0073-3

Page, S. L., Nielsen, R. J., Teeter, K., Lake, C. M., Ong, S., Wright, K. R., Dean, K. L., Agne, D., Gilliland, W. D., & Hawley, R. S. (2007). A germline clone screen for meiotic mutants in Drosophila melanogaster. *Fly*, *1*(3), 172–181. https://doi.org/10.4161/fly.4720

Palmer, W. H., Medd, N. C., Beard, P. M., & Obbard, D. J. (2018). Isolation of a natural DNA virus of Drosophila melanogaster, and characterisation of host resistance and immune responses. *PLOS Pathogens*, *14*(6), e1007050. https://doi.org/10.1371/journal.ppat.1007050

Palmer, W. H., Varghese, F. S., & van Rij, R. P. (2018). Natural Variation in Resistance to Virus Infection in Dipteran Insects. *Viruses*, *10*(3), 118. https://doi.org/10.3390/v10030118

Pardue, M.-L., & DeBaryshe, P. G. (2003). Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annual Review of Genetics*, *37*(1), 485–511. https://doi.org/10.1146/annurev.genet.38.072902.093115

Pasyukova, E. G., & Nuzhdin, S. v. (1993). Doc and copia instability in an isogenic Drosophila melanogaster stock. *Molecular & General Genetics : MGG*, *240*(2), 302–306. https://doi.org/10.1007/BF00277071

Pasyukova, E. G., Nuzhdin, S. v, Morozova, T. v, & Mackay, T. F. C. (2004). Accumulation of transposable elements in the genome of Drosophila melanogaster is associated with a decrease in fitness. *The Journal of Heredity*, *95*(4), 284–290. https://doi.org/10.1093/jhered/esh050

Petit, M., Mongelli, V., Frangeul, L., Blanc, H., Jiggins, F. M., & Saleh, M.-C. (2016). piRNA pathway is not required for antiviral defense in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), E4218-27. https://doi.org/10.1073/pnas.1607952113

Petrov, D. A., Fiston-Lavier, A.-S., Lipatov, M., Lenkov, K., & González, J. (2011). Population genomics of transposable elements in Drosophila melanogaster. *Molecular Biology and Evolution*, *28*(5), 1633–1644. https://doi.org/10.1093/molbev/msq337

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature Methods*, *13*(10). https://doi.org/10.1038/nmeth.3972

Port, F., Chen, H.-M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proceedings of the National Academy of Sciences*, *111*(29). https://doi.org/10.1073/pnas.1405500111

Price, A. A., Grakoui, A., & Weiss, D. S. (2016). Harnessing the Prokaryotic Adaptive Immune System as a Eukaryotic Antiviral Defense. *Trends in Microbiology*, *24*(4), 294–306. https://doi.org/10.1016/j.tim.2016.01.005

Price, A. A., Sampson, T. R., Ratner, H. K., Grakoui, A., & Weiss, D. S. (2015). Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(19), 6164–6169. https://doi.org/10.1073/pnas.1422340112

Qu, J., Ye, J., & Fang, R. (2007). Artificial microRNA-mediated virus resistance in plants. *Journal of Virology*, *81*(12), 6690–6699. https://doi.org/10.1128/JVI.02457-06

Quesneville, H., Nouaud, D., & Anxolabéhère, D. (2005). Recurrent recruitment of the THAP DNA-binding domain and molecular domestication of the P-transposable element. *Molecular Biology and Evolution*, *22*(3), 741–746. https://doi.org/10.1093/molbev/msi064

Radion, E., Morgunova, V., Ryazansky, S., Akulenko, N., Lavrov, S., Abramov, Y., Komarov, P. A., Glukhov, S. I., Olovnikov, I., & Kalmykova, A. (2018). Key role of piRNAs in telomeric chromatin maintenance and telomere nuclear positioning in Drosophila germline. *Epigenetics & Chromatin*, *11*(1), 18–40. https://doi.org/10.1186/s13072-018-0210-4

Rashkova, S., Karam, S. E., & Pardue, M. L. (2002). Element-specific localization of Drosophila retrotransposonGag proteins occurs in both nucleus and cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(6), 3621–3626. https://doi.org/10.1073/pnas.032071999

Reed, L. J., & Muench, H. (1938). *A Simple Method for Estimating Fifty Per Cent Endpoints*. http://books.google.co.uk/books?id=xaiengEACAAJ&dq=intitle:A+simple+method+of+estimating+fifty+per+cent+endpoints&hl=&cd=1&source=gbs_api

Ross, R. J., Weiner, M. M., & Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature*, *505*(7483), 353–359. https://doi.org/10.1038/nature12987

Rostant, W. G., Wedell, N., & Hosken, D. J. (2012). Transposable Elements and Insecticide Resistance. *Advances in Genetics, Vol 78*, *78*, 169–201. https://doi.org/10.1016/B978-0-12-394394-1.00002-X

Roulston, A., Marcellus, R. C., & Branton, P. E. (1999). Viruses and apoptosis. *Annual Review of Microbiology*, *53*, 577–628. https://doi.org/10.1146/annurev.micro.53.1.577

Roussigne, M., Kossida, S., Lavigne, A.-C., Clouaire, T., Ecochard, V., Glories, A., Amalric, F., & Girard, J.-P. (2003). The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. *Trends in Biochemical Sciences*, *28*(2), 66–69. https://doi.org/10.1016/S0968-0004(02)00013-0

Ruiz-Vazquez, P., & Silva, F. J. (1999). Aberrant splicing of the Drosophila melanogasterphenylalanine hydroxylase pre-mRNA caused by the insertion of aB104/roo transposable element in the Henna locus. *Insect Biochemistry and Molecular Biology*, *29*(4), 311–318. https://doi.org/10.1016/s0965-1748(99)00002-8

Saha, A., Mitchell, J. A., Nishida, Y., Hildreth, J. E., Ariberre, J. A., Gilbert, W. v, & Garfinkel, D. J. (2015). A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. *Journal of Virology*, *89*(7), 3922–3938. https://doi.org/10.1128/JVI.03060-14

Sarkar, A., Sim, C., Hong, Y. S., Hogan, J. R., Fraser, M. J., Robertson, H. M., & Collins, F. H. (2003). Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. *Molecular Genetics and Genomics : MGG*, *270*(2), 173–180. https://doi.org/10.1007/s00438-003-0909-0

Sarkar, A., Volff, J.-N., & Vaury, C. (2017). piRNAs and their diverse roles: a transposable element-driven tactic for gene regulation? *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *31*(2), 436–446. https://doi.org/10.1096/fj.201600637RR

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., & Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. *Genes & Development*, *20*(3), 345–354. https://doi.org/10.1101/gad.370206

Schlenke, T. A., & Begun, D. J. (2004). Strong selective sweep associated with a transposoninsertion in Drosophila simulans. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1626–1631. https://doi.org/10.1073/pnas.0303793101

Schnettler, E., Donald, C. L., Human, S., Watson, M., Siu, R. W. C., McFarlane, M., Fazakerley, J. K., Kohl, A., & Fragkoudis, R. (2013). Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *The Journal of General Virology*, *94*(Pt 7), 1680–1689. https://doi.org/10.1099/vir.0.053850-0

Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Molecular Ecology*, *28*(6), 1537–1549. https://doi.org/10.1111/mec.14794

Sela, N., Kim, E., & Ast, G. (2010). The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biology*, *11*(6), R59-13. https://doi.org/10.1186/gb-2010-11-6-r59

Shelly, S., Lukinova, N., Bambina, S., Berman, A., & Cherry, S. (2009). Autophagy is an essential component of Drosophila immunity against vesicular stomatitis virus. *Immunity*, *30*(4), 588–598. https://doi.org/10.1016/j.immuni.2009.02.009

Sinzelle, L., Izsvák, Z., & Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cellular and Molecular Life Sciences : CMLS*, *66*(6), 1073–1093. https://doi.org/10.1007/s00018-009-8376-3

Smith, M. C. M., Brown, W. R. A., McEwan, A. R., & Rowley, P. A. (2010). Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochemical Society Transactions*, *38*(2), 388–394. https://doi.org/10.1042/BST0380388

Smith, N. A., Singh, S. P., Wang, M. B., Stoutjesdijk, P. A., Green, A. G., & Waterhouse, P. M. (2000). Gene expression - Total silencing by intron-spliced hairpin RNAs. *Nature*, *407*(6802), 319–320. https://doi.org/10.1038/35030305

Spellberg, M. J., & Marr, M. T. (2015). FOXO regulates RNA interference in Drosophila and protects from RNA virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), 14587–14592. https://doi.org/10.1073/pnas.1517124112

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, *24*(9), 2241–2252. https://doi.org/10.1111/mec.13089

Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichiyanagi, K., Adachi, N., Koyama, H., Takeda, S., & Okada, N. (2009). Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genetics*, *5*(4), e1000461. https://doi.org/10.1371/journal.pgen.1000461

Svoboda, Y. H., Robson, M. K., & Sved, J. A. (1995). P-element-induced male recombination can be produced in Drosophila melanogaster by combining end-deficient elements in trans. *Genetics*, *139*(4), 1601–1610. /pmc/articles/PMC1206487/?report=abstract

Tellier, M., & Chalmers, R. (2019). Human SETMAR is a DNA sequence-specific histone-methylase with a broad effect on the transcriptome. *Nucleic Acids Research*, *47*(1), 122–133. https://doi.org/10.1093/nar/gky937

Teninges, D., Bras, F., & Dezélée, S. (1993). Genome organization of the sigma rhabdovirus: six genes and a gene overlap. *Virology*, *193*(2), 1018–1023. https://doi.org/10.1006/viro.1993.1219

Thomas-Orillard, M., Jeune, B., & Cusset, G. (1995). Drosophila-host genetic control of susceptibility to Drosophila C virus. *Genetics*, *140*(4), 1289–1295. /pmc/articles/PMC1206694/?report=abstract

Tinsley, M. C., Blanford, S., & Jiggins, F. M. (2006). Genetic variation in Drosophila melanogaster pathogen susceptibility. *Parasitology*, *132*(Pt 6), 767–773. https://doi.org/10.1017/S0031182006009929

Tipney, H. J., Hinsley, T. A., Brass, A., Metcalfe, K., Donnai, D., & Tassabehji, M. (2004). Isolation and characterisation of GTF2IRD2, a novel fusion gene and member of the TFII-I family of transcription factors, deleted in Williams-Beuren syndrome. *European Journal of Human Genetics : EJHG*, *12*(7), 551–560. https://doi.org/10.1038/sj.ejhg.5201174

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, *302*(5909), 575–581. https://doi.org/10.1038/302575a0

Vagin, V. v, Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, *313*(5785), 320–324. https://doi.org/10.1126/science.1129333

van Beest, M., Clevers, H., & Mortin, M. (1998). Drosophila RpS3a, a novel Minute gene situated between the segment polarity genes cubitus interruptus and dTCF. *Nucleic Acids Research*, *26*(19), 4471–4475. https://doi.org/10.1093/nar/26.19.4471

van de Lagemaat, L. N., Landry, J.-R., Mager, D. L., & Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics : TIG*, *19*(10), 530–536. https://doi.org/10.1016/j.tig.2003.08.004

van Rij, R. P., Saleh, M.-C., Berry, B., Foo, C., Houk, A., Antoniewski, C., & Andino, R. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. *Genes & Development*, *20*(21), 2985–2995. https://doi.org/10.1101/gad.1482006

Vasiliauskaitė, L., Berrens, R. v, Ivanova, I., Carrieri, C., Reik, W., Enright, A. J., & O'Carroll, D. (2018). Defective germline reprogramming rewires the spermatogonial transcriptome. *Nature Structural & Molecular Biology*, *25*(5), 394–404. https://doi.org/10.1038/s41594-018-0058-0

Vaury, C., Chaboissier, M. C., Drake, M. E., Lajoinie, O., Dastugue, B., & Pélisson, A. (1994). The Doc transposable element in Drosophila melanogaster and Drosophila simulans: genomic distribution and transcription. *Genetica*, *93*(1–3), 117–124. https://doi.org/10.1007/BF01435244

Venken, K. J. T., Carlson, J. W., Schulze, K. L., Pan, H., He, Y., Spokony, R., Wan, K. H., Koriabine, M., de Jong, P. J., White, K. P., Bellen, H. J., & Hoskins, R. A. (2009). Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. *Nature Methods*, *6*(6). https://doi.org/10.1038/nmeth.1331

Wang, S. H., & Elgin, S. C. R. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(52), 21164–21169. https://doi.org/10.1073/pnas.1107892109

Wang, X.-H., Aliyari, R., Li, W.-X., Li, H.-W., Kim, K., Carthew, R., Atkinson, P., & Ding, S.-W. (2006). RNA interference directs innate immunity against viruses in adult Drosophila. *Science*, *312*(5772), 452–454. https://doi.org/10.1126/science.1125694

Wayne, M. L., Blohm, G. M., Brooks, M. E., Regan, K. L., Brown, B. Y., Barfield, M., Holt, R. D., & Bolker, B. M. (2011). The prevalence and persistence of sigma virus, a biparentally transmitted parasite of Drosophila melanogaster. *Evolutionary Ecology Research*, *13*(4), 323–345. /pmc/articles/PMC5313041/

Webster, A., Li, S., Hur, J. K., Wachsmuth, M., Bois, J. S., Perkins, E. M., Patel, D. J., & Aravin, A. A. (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. *Molecular Cell*, *59*(4). https://doi.org/10.1016/j.molcel.2015.07.017

Webster, C. L., Waldron, F. M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J. F., Brouqui, J.-M., Bayne, E. H., Longdon, B., Buck, A. H., Lazzaro, B. P., Akorli, J., Haddrill, P. R., & Obbard, D. J. (2015). The Discovery, Distribution, and Evolution of Viruses Associated with

Drosophila melanogaster. *PLoS Biology*, *13*(7), e1002210. https://doi.org/10.1371/journal.pbio.1002210

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. https://doi.org/10.1038/nrg2165

Wilfert, L., & Jiggins, F. M. (2010). Disease association mapping in Drosophila can be replicated in the wild. *Biology Letters*, *6*(5), 666–668. https://doi.org/10.1098/rsbl.2010.0329

Wu, M., Li, L., & Sun, Z. (2007). Transposable element fragments in protein-coding regions and their contributions to human functional proteins. *Gene*, *401*(1–2), 165–171. https://doi.org/10.1016/j.gene.2007.07.012

Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E. C., Li, W.-X., & Ding, S.-W. (2010). Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(4), 1606–1611. https://doi.org/10.1073/pnas.0911353107

Xu, H., & Boeke, J. D. (1990). Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. *Molecular and Cellular Biology*, *10*(6), 2695–2702. https://doi.org/10.1128/mcb.10.6.2695

Xu, J., & Cherry, S. (2014). Viruses and antiviral immunity in Drosophila. *Developmental and Comparative Immunology*, *42*(1), 67–84. https://doi.org/10.1016/j.dci.2013.05.002

Yampolsky, L. Y., Webb, C. T., Shabalina, S. A., & Kondrashov, A. S. (1999). Rapid accumulation of a vertically transmitted parasite triggered by relaxation of natural selection among hosts. *Evolutionary Ecology Research*, *1*(5), 581–589. http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcAuth=mekentosj&SrcApp=Papers&DestLinkType=FullRecord&DestApp=WOS&KeyUT=000082728800006

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics : TIG*, *13*(8), 335–340. https://doi.org/10.1016/s0168-9525(97)01181-5

Zambon, R. A., Nandakumar, M., Vakharia, V. N., & Wu, L. P. (2005). The Toll pathway is important for an antiviral response in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7257–7262. https://doi.org/10.1073/pnas.0409181102

Zhao, D., & Bownes, M. (1998). The RNA product of the Doc retrotransposon is localized on the Drosophila oocyte cytoskeleton. *Molecular & General Genetics : MGG*, *257*(5), 497–504. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9563835&retmode=ref&cmd=prlinks

Zhu, W., Lei, R., le Duff, Y., Li, J., Guo, F., Wainberg, M. A., & Liang, C. (2015). The CRISPR/Cas9 system inactivates latent HIV-1 proviral DNA. *Retrovirology*, *12*(1), 22–27. https://doi.org/10.1186/s12977-015-0150-z

Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J., & Staals, R. H. J. (2018). Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Bioscience Reports*, *38*(3). https://doi.org/10.1042/BSR20170788

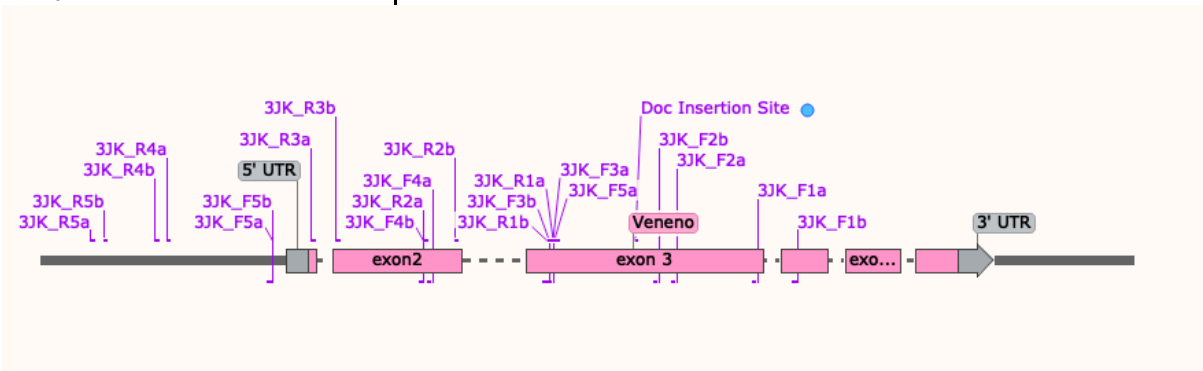
Zirin, J., Hu, Y., Liu, L., Yang-Zhou, D., Colbeth, R., Yan, D., Ewen-Campen, B., Tao, R., Vogt, E., VanNest, S., Cavers, C., Villalta, C., Comjean, A., Sun, J., Wang, X., Jia, Y., Zhu, R., Peng, P., Yu, J., ... Perrimon, N. (2020). Large-Scale Transgenic Drosophila Resource Collections for Loss- and Gain-of-Function Studies. *Genetics*, *214*(4). https://doi.org/10.1534/genetics.119.302964

Zitzmann, J., Schreiber, C., Eichmann, J., Bilz, R. O., Salzig, D., Weidner, T., & Czermak, P. (2018). Single-cell cloning enables the selection of more productive Drosophila melanogaster S2 cells for recombinant protein expression. *Biotechnology Reports (Amsterdam, Netherlands)*, *19*, e00272. https://doi.org/10.1016/j.btre.2018.e00272

Appendices

Appendix 1 - Primers used in locating *Doc* element

Oligo Name	Sequence
3JK_F1a	ACGATTCCATTTGTTTGGGCAT
3JK_R1a	AGAAGCTCCGTGAGGTTTCC
3JK_F1b	CCACAACGATGAGATCAATGCCT
3JK_R1b	GTGCGAGCGAAGATGTGTAAG
3JK_F2a	CATTCTGTACCTCTGGTTTCGTT
3JK_R2a	GATATTCCCGTTTCTCCGCCA
3JK_F2b	TCAATATGTGAAGGTTGCGGA
3JK_R2b	GCTGAAAGAGCGTTTGACGA
3JK_F3a	TTACACATCTTCGCTCGCACATA
3JK_R3a	AAAAGTCAGTTGCCGGTGCT
3JK_F3b	GCACATACATGCCATTAAAGTGGA
3JK_R3b	AAGCCAAAGGAAAATGCGTCC
3JK_F4a	ATGATCGCTTTTGGCGGAGA
3JK_R4a	TGCAAGGTACACAGGTCCAC
3JK_F4b	TCTTTGGCTGCAATGATGCTG
3JK_R4b	GCGATGAGAAGCTCGTTGGT
3JK_F5a	ATGAAATGCCGTTTTCCCACTTATT
3JK_R5a	CCGACGCAGTACTTCCGATAG
3JK_F5b	AATGAAATGCCGTTTTCCCACTTA
3JK_R5b	AGTTCCAGCGTGAAACTCCAA



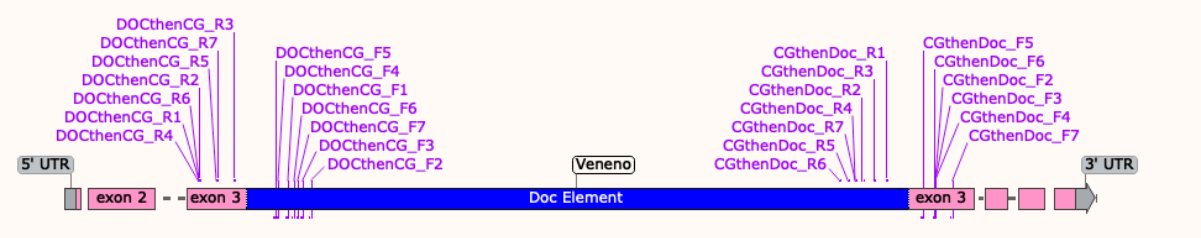
Appendix 2 - Primers for sequencing *Doc* insertion and flanking sequences

3' junction

Oligo Name	Sequence
CGthenDoc_F1	TAGGCTTTGGTGCAGTGTCC
CGthenDoc_R1	CCTCCCAGCCCAGCAATAAA
CGthenDoc_F2	TGTCCGGAGTTCTGTTGCTT
CGthenDoc_R2	CCGTGGTACCTTCGAAACGA
CGthenDoc_F3	TTGTCCGGAGTTCTGTTGCT
CGthenDoc_R3	AACGAAAAGCTGACCACCCA
CGthenDoc_F4	TTTGTCCGGAGTTCTGTTGCT
CGthenDoc_R4	TCAGCGAGCACAGTCAAGAA
CGthenDoc_F5	GCAGGAATTGCAATGGGTTGA
CGthenDoc_R5	CTTATGGCTCCGAGCTGTGG
CGthenDoc_F6	GTCCGGAGTTCTGTTGCTTGA
CGthenDoc_R6	TCGCTCTCCACTTAGTCTGGA
CGthenDoc_F7	TTAGGCTTTGGTGCAGTGTC
CGthenDoc_R7	ATTCAGCGAGCACAGTCAAGA

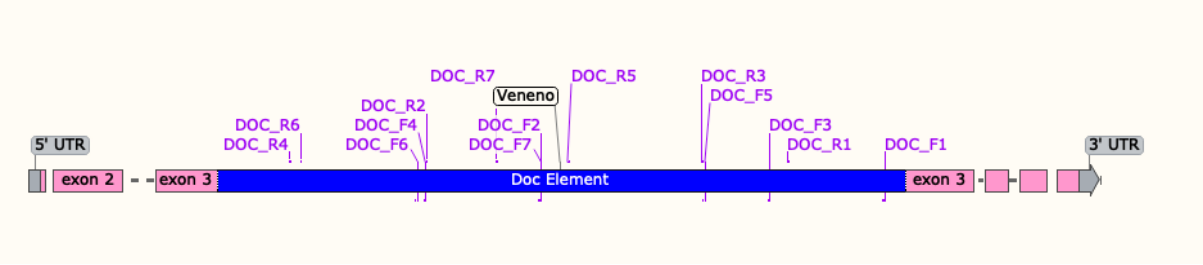
5' junction

Oligo Name	Sequence
DOCthenCG_F1	GTGCTCGATCACCGATTTGC
DOCthenCG_R1	GCATGTATGTGCGAGCGAAG
DOCthenCG_F2	GTTACAGTTACCCGGGACGG
DOCthenCG_R2	TGTGCGAGCGAAGATGTGTA
DOCthenCG_F3	GTCTCTTCGCTCCACGACTT
DOCthenCG_R3	GTGCCCTGGATTACACCTCG
DOCthenCG_F4	GCGCTCGTCTTGTTCACATT
DOCthenCG_R4	AATGGCATGTATGTGCGAGC
DOCthenCG_F5	TCGCTGAGAACGTATGTCGT
DOCthenCG_R5	AAAGCATTGCCCTGTTCCAG
DOCthenCG_F6	TCGCTGAGGGCAAAAGTTTC
DOCthenCG_R6	TGTATGTGCGAGCGAAGATGT
DOCthenCG_F7	GGCAGAGAGGGAGAGCAAGA
DOCthenCG_R7	CCTGTTCCAGCCAACTGGT



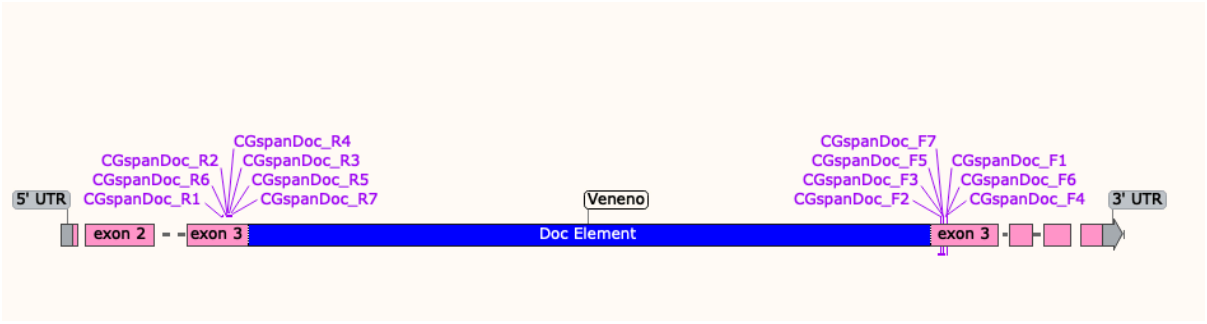
***Doc* only primers**

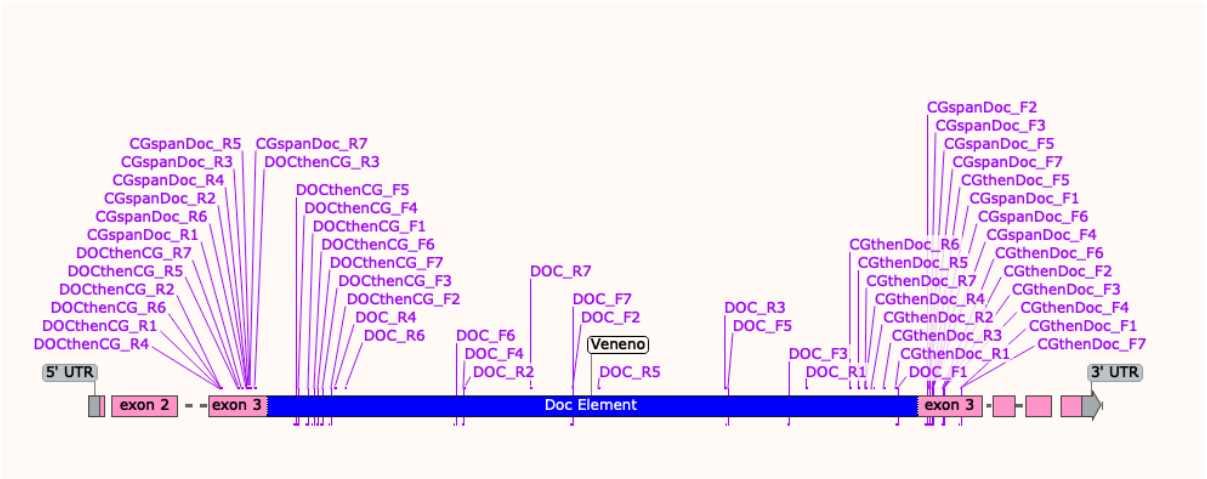
Oligo Name	Sequence
DOC_F1	TTTATTGCTGGGCTGGGAGG
DOC_R1	GCGATACTCAGTCGCTCGAA
DOC_F2	GCAGTAGACTGCGGCTAGAG
DOC_R2	CTGCCCCAAAAACAAGGACG
DOC_F3	CGCGTGAAGTGCTTGTATCG
DOC_R3	TGTTCAACGCAATCGCCAAG
DOC_F4	TGCGGTAGTATGAGGTTCGC
DOC_R4	CTGGCAACTGCCACAACATC
DOC_F5	CTTGGCGATTGCGTTGAACA
DOC_R5	ATCCCCGGGTAAGCCTACAT
DOC_F6	CCTTGGTGTGGCCGTATTCT
DOC_R6	GCACTGACCGTTTGCAACAA
DOC_F7	TAGACTGCGGCTAGAGTCGT
DOC_R7	CTCTGTGGAACGCAAATGGC



Spanning primers

Oligo Name	Sequence
CGspanDoc_F1	TTGCAGGAATTGCAATGGGTT
CGspanDoc_R1	AGCTCAAGAATGTATTGAATTGTGA
CGspanDoc_F2	TGCGGACCTTCTAAGGAACA
CGspanDoc_R2	TTCAACTCGAAATTCGTGGGC
CGspanDoc_F3	ATATGTGAAGGTTGCGGACC
CGspanDoc_R3	CAACTCGAAATTCGTGGGCAAG
CGspanDoc_F4	ATTGCAGGAATTGCAATGGGTT
CGspanDoc_R4	TTCAACTCGAAATTCGTGGGCA
CGspanDoc_F5	TCAATATGTGAAGGTTGCGGAC
CGspanDoc_R5	TCGTGGGCAAGAAATATATGGCA
CGspanDoc_F6	ATTGCAGGAATTGCAATGGGTTG
CGspanDoc_R6	ATGTCATTGCCTTCAACTCGAAAT
CGspanDoc_F7	ATCAATATGTGAAGGTTGCGGA
CGspanDoc_R7	GTGGGCAAGAAATATATGGCAAC





Appendix 3 - Oligos Used to make plasmid expressing RNAi targeting *Doc* Element

DocBegin

Target:	GGCTCTCTTTACAACGCAACA
Oligo Name	Sequence
DocBegin (Top Strand)	**CTAGCAGTTGTTGCGTTGTAAAGAGAGCC**TAGTTATATT CAAGCATAGGCTCTCTTTACAACGCAACAGCG
DocBegin (Bottom Strand)	**AATTCGCTGTTGCGTTGTAAAGAGAGCCTATGCTTGAATAT AACTA**GGCTCTCTTTACAACGCAACAA**CTG**

DocEnd

Target:	GGAGTTCAAAGCTCTTCTATA
Oligo Name	Sequence
DocEnd (Top Strand)	**CTAGCAGTTATAGAAGAGCTTTGAACTCC**TAGTTATATT CAAGCATAGGAGTTCAAAGCTCTTCTATAGCG
DocEnd (Bottom Strand)	**AATTCGCTATAGAAGAGCTTTGAACTCCTATGCTTGAATA TAACTA**GGAGTTCAAAGCTCTTCTATAA**CTG**

Appendix 4 - Oligos used for CRISPR/Cas9 gRNAs

upDoc

Target:	TTCCCTGTCGCTCTGGAACA
Oligo Name	Sequence
upDoc_sense	GTCGTTCCCTGTCGCTCTGGAACA
upDoc_anti-sense	AAACTGTTCCAGAGCGACAGGGAA

downDoc

Target:	TGTAATCGAGGTACATTACC
Oligo Name	Sequence
downDoc_sense	GTCGTGTAATCGAGGTACATTACC
upDoc_anti-sense	AAACGGTAATGTACCTCGATTACA

Doc_RNAgene

Target:	AATGTGAACAAGACGAGCGC
Oligo Name	Sequence
Doc_RNAgene_sense	GTCGAATGTGAACAAGACGAGCGC
Doc_RNAgene_anti-sense	AAACGCGCTCGTCTTGTTCACATT

DocRT

Target:	CTTCCCTACGGATATCTCTG
Oligo Name	Sequence
Doc_RT_sense	GTCGCTTCCCTACGGATATCTCTG
Doc_RT_anti-sens	AAACCAGAGATATCCGTAGGGAAG

Appendix 5 - Illumina primers for Sequencing CRISPR target sites:

DownDoc 1 ($DGRP_{59}$)

Before overhangs:	
Forward	TGGACCTTCAAAACCTCAGCCA
Reverse	CAGCACCTACGCCTCCGAAT
With overhangs:	
Oligo Name	Sequence
Ilumina_DownDoc_F1:	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTGGACCTTCAAAACCTCAGCCA
Ilumina_DownDoc_R1:	GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC AGCAGCACCTACGCCTCCGAAT

DownDoc 2 ($DGRP_{362}$)

Before overhangs:	
Forward	TCAGCCAGATAGGAAACCTGTGC
Reverse	GGGGCCTTTGACAGCACCTA
With overhangs:	
Oligo Name	Sequence
Ilumina_DownDoc_F2:	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG TCAGCCAGATAGGAAACCTGTGC
Ilumina_DownDoc_R2:	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GGGGCCTTTGACAGCACCTA

UpDoc ($DGRP_{59}$ and $DGRP_{362}$)

Before overhangs:	
Forward primer	GGATGATCGCTTTTGGCGGA
Reverse primer	GCACAAAAGTCAGTTGCCGGTG
With overhangs:	
Oligo Name	Sequence
Ilumina_UpDoc_F1:	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGG ATGATCGCTTTTGGCGGA
Ilumina_UpDoc_R1:	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGGCACAAAAGTCAGTTGCCGGTG

Appendix 6 – Primers for pMT-*Ven*^{trunc}–puro plasmid

Oligo Name	Sequence
Ven_exon1_F	cagatatccagcacagtggcACCGTGGGGCTTCAGCTG
Ven_exon1_R	cttgcgcaagTTCATCAGCACCGGCAACTG
Ven_exon2_F	tgctgatgaaCTTGCGCAAGCCAAAGGAAAATG
Ven_exon2_R	tggtattaacCTTCTTGCAAACTTCGTCAAACG
Ven_exon3_F	ttgcaagaagGTTAATACCATAGCTACGTTATTG
Ven_exon3_R_stop	agggccctctagactcgagcTATTCGTAGCTTGAAAGAAAC

Appendix 7 – Primers for pMT-*Ven*^{trunc}-V5–puro plasmid

Oligo Name	Sequence
Ven_exon1_F	cagatatccagcacagtggcACCGTGGGGCTTCAGCTG
Ven_exon1_R	cttgcgcaagTTCATCAGCACCGGCAACTG
Ven_exon2_F	tgctgatgaaCTTGCGCAAGCCAAAGGAAAATG
Ven_exon2_R	tggtattaacCTTCTTGCAAACTTCGTCAAACG
Ven_exon3_F	ttgcaagaagGTTAATACCATAGCTACGTTATTG
Ven_exon3_R_V5	agggccctctagactcgagcATTTCGTAGCTTGAAAGAAAC

Appendix 8 – Primers for Q5 Site Directed Mutatagenesis

Oligo Name	Sequence
ZnCont_F	GGCTACGTCAcCGTTACATCTG
ZnCont_R	AATCTTGCCGCTGACAATG
ZnMut_F	TTGGCTACGTgcTCGTTACATCTGCATC
ZnMut_R	TCTTGCCGCTGACAATGC
PremStop_F	CATGCTAGCAtAAAAGTCAGTTG
PremStop_R	TTGTCACACTTTATTTCCG
PremCont_F	TGCTAGCACAgAAGTCAGTTG
PremCont_R	TGTTGTCACACTTTATTTCC
noDoc-F	ATGCTCGAGTCTAGAGGG
noDoc_R	CAACTTCTTACTGTTGAAAAATGC
TudMut_F	TAATGGCATGgcTGTGCGAGCG
TudMut_R	AAGTGGACTAGGCACAATC
TudCont_F	ATGGCATGTAcGTGCGAGCGA
TudCont_R	TAAAGTGGACTAGGCACAATC

Appendix 9 - Primers for Proteomics Experiment Plasmids

GFP-V5

Oligo Name	Sequence
GFP_F	tatccagcacagtggcATGGTGAGCAAGGGCGAG
GFP_R	ccctctagactcgagccACTTGTACAGCTCGTCCATGCCGAGAG

VenSus:

Oligo Name	Sequence
Ventrunc_F	tatccagcacagtggcATGCTAGCACAAAAGTCAGTTGCC
Ventrunc_R	GGTTTCGTAGCAACTTCTTACTGTTGAAAAATGCGTTTAACTGTTCATTTACC
VenCompleting_F	TAAGAAGTTGCTACGAAACCCTAATTCTGC
VenCompleting_R	ccctctagactcgagcgACAAGCTCAACAGCTTCTGAATAAGAC

Appendix 10 Primers for pMT-CHKov1_tr0-V5-puro

Oligo Name	Sequence
pMTchkov_F1	tccagtgtggtggaattctgcagatatccagcacagtggcCATCA GTCTACGCAAGACCGT
pMTchkov_R1	CATTGTCGCTGAGAACGTATGTTGTGCTTC ATCATTTCCTTGTTGAT
pMTchkov_F2	AGGAAATGATGAAGCACAACATACGTTCTC AGCGACAATGTGAAC
pMTchkov_R2	agggataggcttaccttcgaagggccctctagactcgagcATTGG GGATTTTGAAGATACAAGAAGTTGAATCAG

Appendix 11 - Primers for pUb-*CHKov1_tr0*

Oligo Name	Sequence
CHKovtopubtag_F	Ttctgcccgcagaataatccaagggcggatccag atctgcATGGCGGCTGGCAAAA
CHKovtopubtag_R	CCGTCATGGTCTTTGTAGT CTGGGGATTTTGAAGATACAAGAAGTTGAATCAG
CHKovtopubnotag_F (same as CHKovtopubtag_F)	Ttctgcccgcagaataatccaagggcggatccaga tctgcATGGCGGCTGGCAAAA
CHKovtopubnotag_R	GttccttcacaaagatcctctagaggtaccctcgagccgcTATGGG GATTTTGAAGATACAAGAAGTTGAATCAGGA
Venuschkov_F	TTGTATCTTCAAAATCCCCAGA CTACAAAGACCATGACGGTGATTATAAAGA
Venuschkov_R	Gttccttcacaaagatcctctagaggtaccctcgag ccgcTTACTTTTCGAACTGGGGATGGCT

Appendix 12 – Primers for pUAST-CHKov1

pUAST-CHKov1_tr0-GFP

Oligo Name	Sequence
UAS-tr0_F	CTCTGAATAGGGAATTGGGATGGCGGCTGGCAAAATAC
UAS-tr0_R	TCACCATAGATCTGTTAACGATTGGGGATTT TGAAGATACAAGAAGTTGAAT

pUAST-CHKov1^{tr1}-GFP

Oligo Name	Sequence
UAS-tr1_F	CTCTGAATAGGGAATTGGGATGGCGGCTGGCAAAA
UAS-tr1_R	TCACCATAGATCTGTTAACGTTACGT GCATTATTGTTGTTGCAAACGG

pUAST-CHKov1^{tr2}-GFP

Oligo Name	Sequence
UAS-tr2_F	CTCTGAATAGGGAATTGGGATGGCGGCTGGCAAAA
UAS-tr2_R	TCACCATAGATCTGTTAACGAATTTA AGTGGGATAATGTGGAATTTGCTGTCAC

Appendix 13 - Oligo used for RNAi on *CHKov1*

Sense Sequence	CTAATATGTTGCGAGTTAATA
Antisense Sequence	TATTAACTCGCAACATATTAG
top strand oligo	ctagcagtCTAATATGTTGCGAGTTAATAtagttatattcaagcataTATT AACTCGCAACATATTAGgcg
Bottom strand oligos	aattcgcCTAATATGTTGCGAGTTAATAtatgcttgaatataactaTATTAACTCGC AACATATTAGactg

Appendix 14 Ultramers for Cas13 crRNA

Oligo Name	Sequence
DAV 1	CGG GTT CGA TTC CCG GCC GAT GCA CAA GTA AAC CCC TAC CAA CTG GTC GGG GTT TGA AAC TCA TTC AAA TTC GCC TCA GGA GCA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CTG CTG TAT TTA TAT CGT GGG GCA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CAA GTT TGG GTT GTT TGA TGT TA
DAV 2	AAG TTT GGG TTG TTT GAT GTT ACA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CTG TTT GAT ATA CGC TTT GTG GTC AAG TAA ACC CCT ACC AAC TGG TCG GGG TTT GAA ACT TTT TTT TGT TTT AGA GCT AGA AAT AGC AAG TT
mCherry 1	CCG GGT TCG ATT CCC GGC CGA TGC ACA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CGC AGG CTG CTG TCC TGG GTC CAA GTA AAC CCC TAC CAA CTG GTC GGG GTT TGA AAC GTC TTT TTC TGC ATC ACG GGC AAG TAA ACC CCT ACC AAC TGG TCG GGG TTT GAA ACT CGT AGT GGC CGC CAT CCT T
mCherry 2	TCG TAG TGG CCG CCA TCC TTC AAG TAA ACC CCT ACC AAC TGG TCG GGG TTT GAA ACA TGG TGT AGT CCT CGT TGT GCA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CTT TTT TTT GTT TTA GAG CTA GAA ATA GCA AGT T
Zika 1	TCG ATT CCC GGC CGA TGC ACA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CTC CCC TTG TTT CTT TTC TCT TTC AAG TAA ACC CCT ACC AAC TGG TCG GGG TTT GAA ACT TAG GGT CAG GGG TGT TAA TTG CAA GTA AAC CCC TAC CAA CTG GTC GGG GTT TGA AAC CGT TTC TTG TTA CT
Zika 2	TCT TGT TAC TGT GTA GAT CAA GTA AAC CCC TAC CAA CTG GTC GGG GTT TGA AAC TGT TGT AAC CAG CTC TAT GTC GCA AGT AAA CCC CTA CCA ACT GGT CGG GGT TTG AAA CCC ACT CTT GAT GTT TTG TTT TCC AAG TAA ACC CCT ACC AAC TGG TCG GGG TTT GAA ACT GTG GTC ATT CTC TTC TTC AGC TTT TTT TT

Appendix 15 – Primers to Amplify Cas13 crRNA Ultramers

Oligo Name	Sequence
mcherry1F	ccgggttcgattcccggccgat
mcherry1R	AAGGATGGCGGCCACTACGAg
mCherry2F	TCGTAGTGGCCGCCATC
mCherry2R	aacttgctatttctagctctaaaac
DAV1F	cgggttcgattcccgg
DAV1R	TAACATCAAACAACCCAAACTTgt
DAV2F	AAGTTTGGGTTGTTTGATGTTAca
DAV2R	aacttgctatttctagctctaaaaca
Zika 1F	tcgattcccggccgatgcacaagtaaa
Zika 1R	AGTAACAAGAAACGgtttcaaacccc
Zika 2F	tgaaacCGTTTCTTGTTACTGTGTAGATcaagtaaacccc
Zika 2R	tttaacttgctatttctagctctaaaaaaaaaGCTGAAGAAGAGAATGAccacag

Appendix 16 – Primers for Cas13dAdm plasmid

Oligo Name	Sequence
Adm_F	cggatccagatctgccgccaccATGGACTACAAAGAC
Adm_R	gtaccctcgagccgctaATCATTCGTACTTAAGGGATTTACCTCTGGACG

Appendix 17 - Primers targeting junction for sequencing plasmids and screening colonies:

pcfd5

Oligo Name	Sequence
pcfd5-mcherry _F	tttgaaacTCGTAGTGGCCG
pcfd5-mcherry _R	cctgcaggttgttggttgg
pcfd5-dav F	TATCGTGGGGcaagtaaacc
pcfd5-dav_R	gccgagcacaattgtctaga

pUB

Oligo Name	Sequence
pUBF	gcaacggaacaggtttctca
pUbCHKov_R	AATCGGGTATTTTGCCAGCC
pUbCas13dAdm_R	TGATGGCCATGTTGTCCTCT
Venus_F	CGACAACCACTACCTGAGCT
Venus_R	accactgctcccattcatca

Appendix 18 – Primers surrounding insert for sequencing plasmids and screening colonies:

pCFD5

Oligo Name	Sequence
U63seqfwd	ACGTTTTATAACTTATGCCCCTAAG
pCFDseqrev	GCACAATTGTCTAGAATGCATAC

pMT

Oligo Name	Sequence
pMT_For	CATCTCAGTGCAACTAAA
BGH_Rev	TAGAAGGCACAGTCGAGG

Appendix 19 - qPCR Primers and Probes

Virus and Housekeeping

Oligo Name	Primer Sequence	Probe Sequence (one step RT-qPCR only)
DAV_F	AAGGCATACTTGATGGTA	[6FAM]TAGCACAACAGG
DAV_R	GGGTTCCTTCCTTTATATG	AAGTC[BHQ1]
DMelSV_F1	TTCAATTTTGTACGCGGAATC	
DMelSV_R1	TGATCAAACCGCTAGCTTCA	
Actin5C_F	GACGAAGAAGTTGCTGCTCTGGTTG	
Actin5C_R	TGAGGATACCACGCTTGCTCTGC	
Rpl32_F	TGCCAAGTTGTCGCACAAA	
Rpl32_R	TGCGCTTGTTGGAGCCATAAC	
Ef1alpha100E_F	ACGTCTACAAGATCGGAG	[FAM]CATCGGAACCGTA
Ef1alpha100E_R	CAGACTTTACTTCGGTGAC	CCAGTAGGT[BHQ2]

Ven^{trunc} Expression (in transfected cells)

Oligo Name	Primer Sequence
VentruncExpF	GCCCTGGATTACACCTCGT
VentruncExpR	ggagagggttagggataggc

Veneno Expression

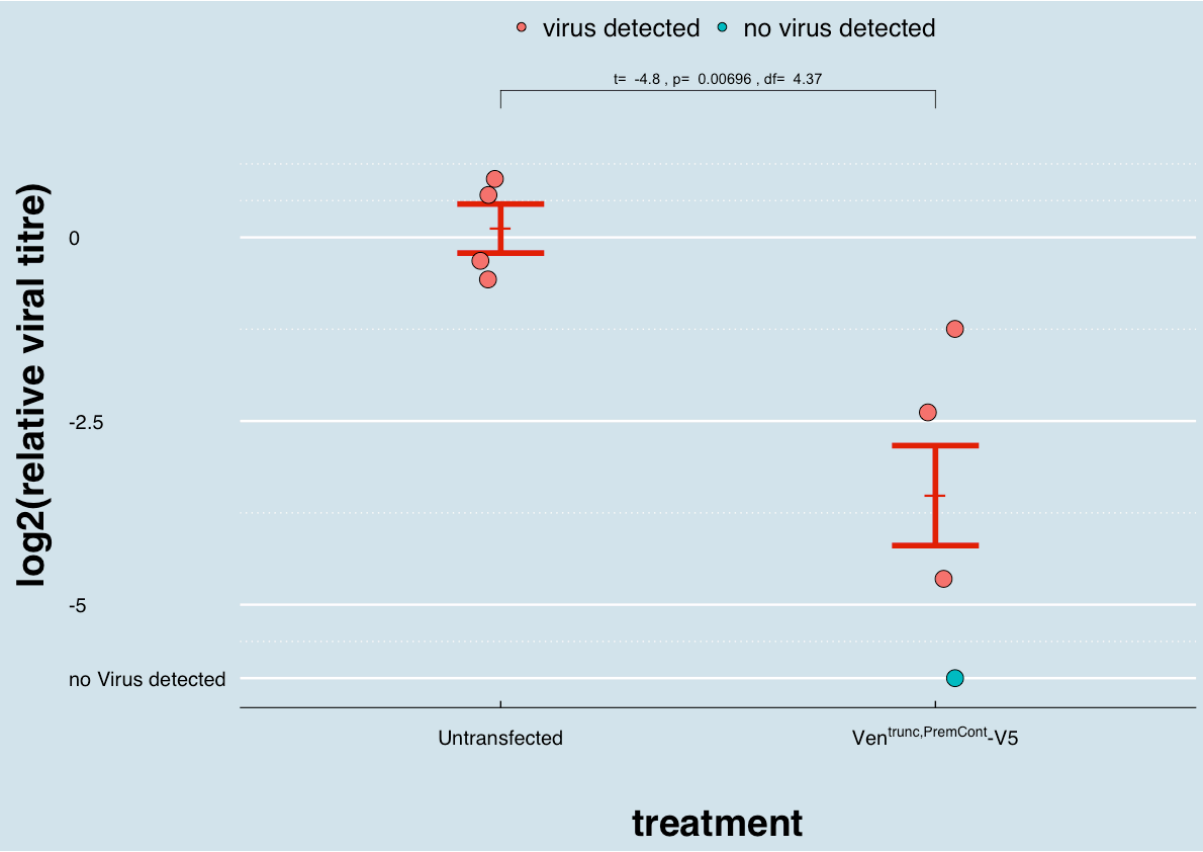
Oligo Name	Primer Sequence
VenExpF	AGGAAAATGCGTCCATTGCG
VenExpR	TTTAACGAACAGGCCCCCTC

***Doc* Expression**

Oligo Name	Primer Sequence
BeginExpF	ACGTTCTCAGCGACAATGTG
BeginExpR	GCAGAGAGGGAGAGCAAGAA
EndExpF	ACAAGCCGACGAGGTAACAT
EndExpR	CTCGGAGCCATAAGTCCAGA

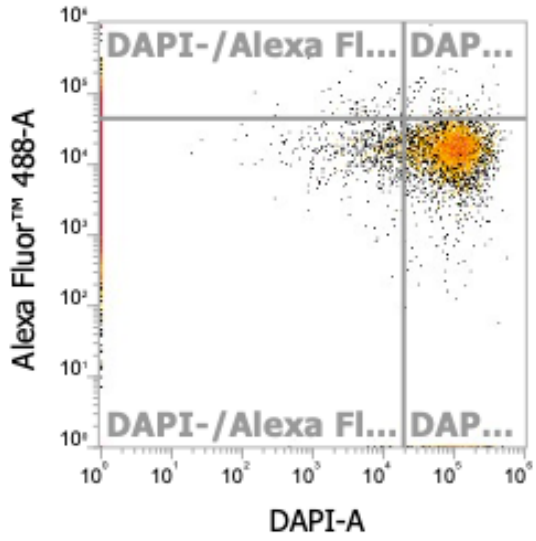
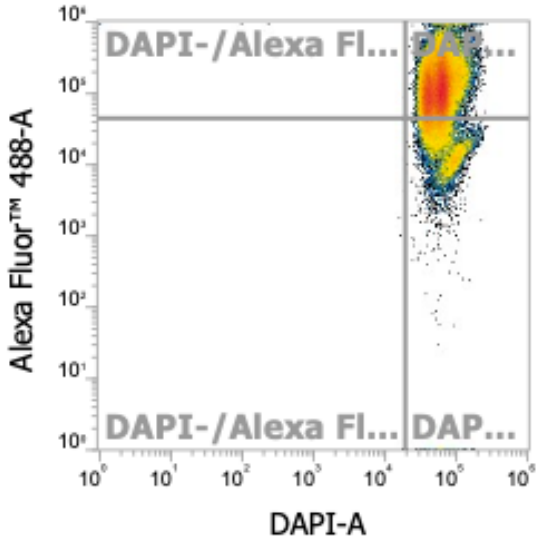
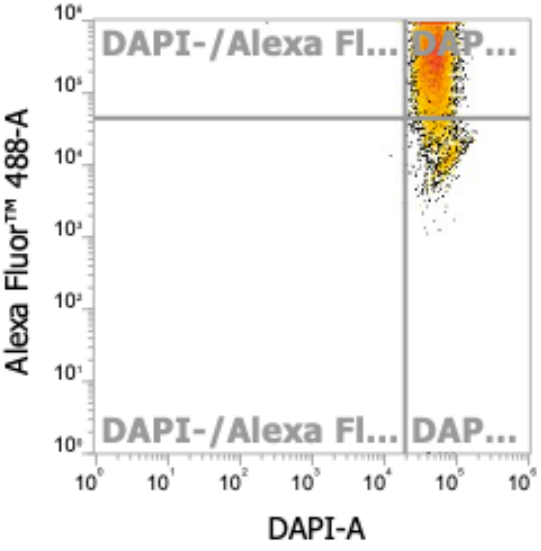
Appendix 20 – Synonymous mutation (Premature Control) and Cloning Does Not Abolish Virus Resistance

We infected cells carrying the pMT-Puro-*Ven*$^{trunc,\ Premature\ Control}$-V5 plasmid (which has an attached V5 tag and a synonymous mutation in its sequence) with DAV which have been cloned from a single cell, and found them to have significantly lower viral titre than control untransfected cells (Figure App21.1, Welch's t-test, t=-4.8, p= 0.00696, df=4.37), indicating they are resistant.



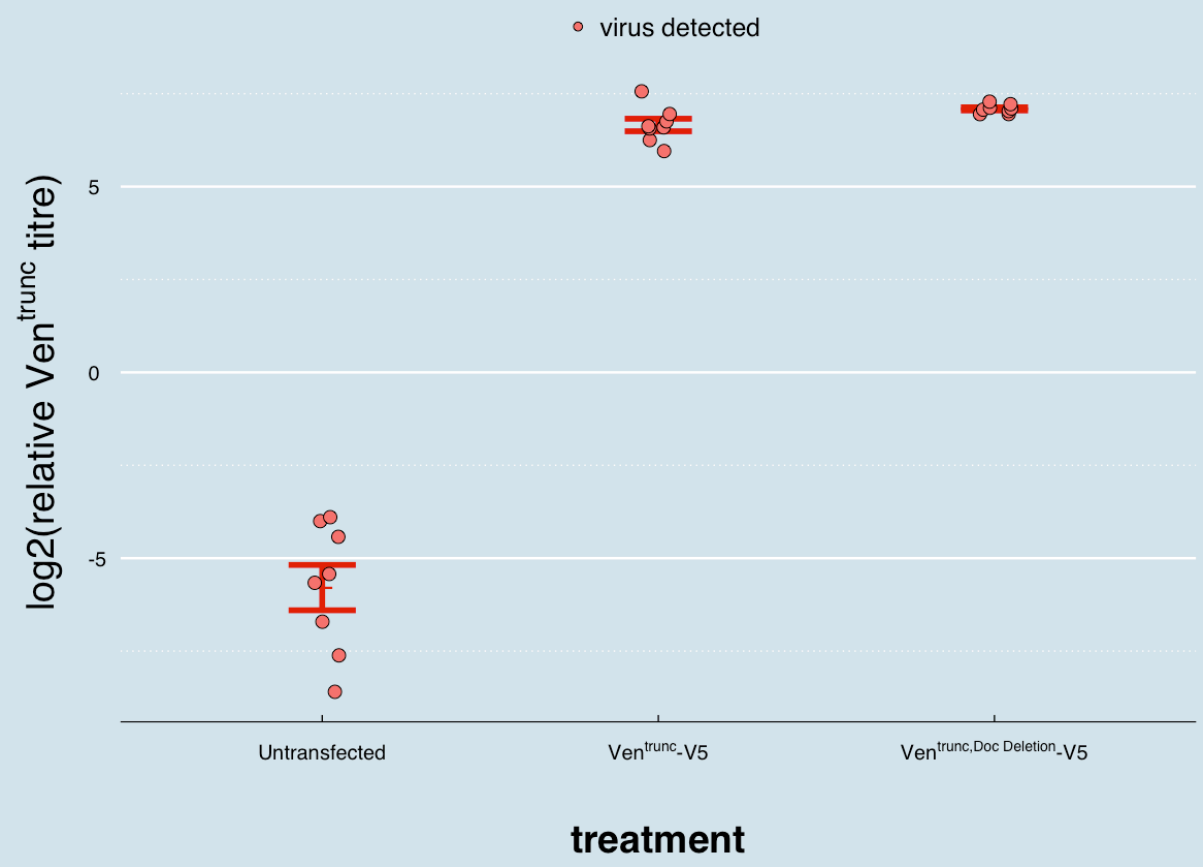
***Ven*$^{trunc,\ Premature\ Control}$ results in lower DAV titre in DL2 cells.** DL2 cells with and without the pMT-Puro-*Ven*$^{trunc,\ Premature\ Control}$ were infected with DAV. Viral titre was measured 3 days post infection. Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the estimated mean assuming a left censored Gaussian distribution. pMT-Puro-*Ven*^{trunc} causes a significant decrease in viral titre from the control, which indicates that the transcript is sufficient to increase resistance in cells. P-values calculated using Welch's t-test. Transfected cells a transgene fused to a V5 tag and are clones of a single cell.

Appendix 21 - Flow cytometry data showing *Ven*^{trunc} expression in Cells from Fig 2.2

Line	Alexa Fluor 488 vs DAPI plot	DAPI+/AF 488A +
untransfected		168 / 65,036 (0.23%)
Ven^{trunc}-V5		58,441 / 69,417 (84%)
Ven$^{trunc,\ Doc\ Deletion}$-V5		10,686 / 12,067 (89%)

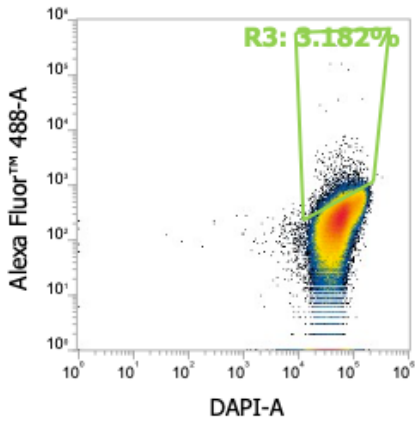
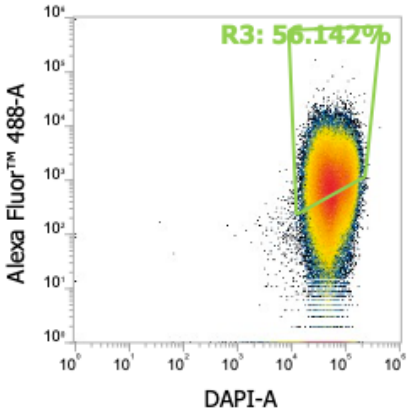
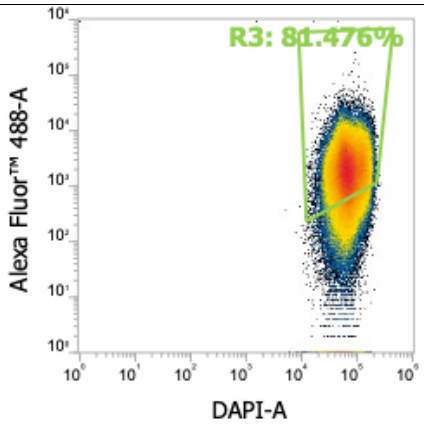
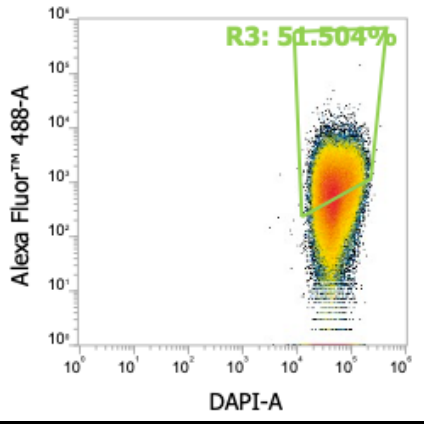
Cells were stained and assayed as described in main methods.

Appendix 22- qPCR data showing *Ven*^{trunc} expression in Cells from Fig 2.2



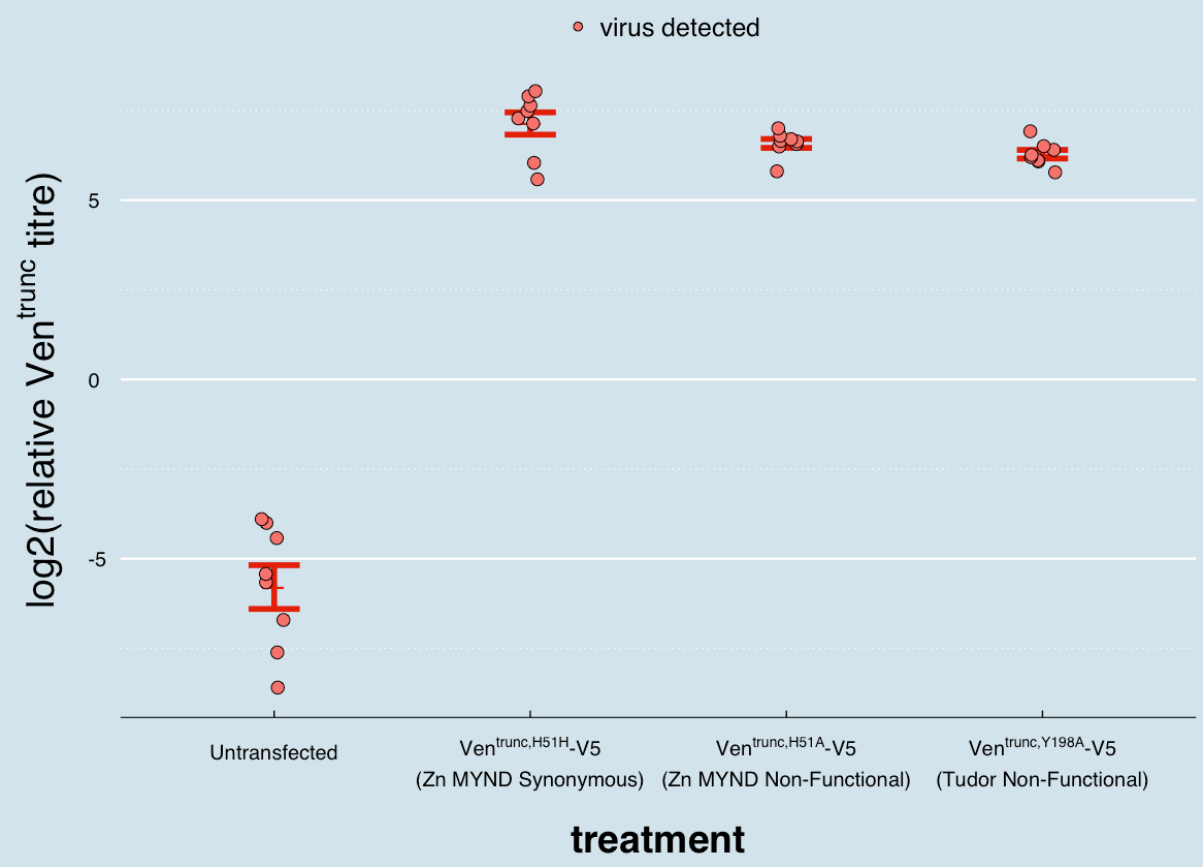
Ven^{trunc} expression was measured in cells with and without a deletion of the *Doc* element sequence from *Ven*^{trunc} (3 days post infection with DAV). Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the mean. Transfected cells are clones of a single cell.

Appendix 23- Flow cytometry data showing *Ven*^{trunc} expression in cells from Fig 2.3

Line	Alexa Fluor 488 vs DAPI plot	DAPI+/AF 488A+
Untransfected		5,555 / 174,564 (3%)
Ven^{trunc,H51H}-V5 (Zn-MYND synonymous)		104,142 / 185,498 (56%)
Ven^{trunc,H51A}-V5 (Zn-MYND non-functional)		165,498 / 203,124 (81%)
Ven$^{trunc,\ Y198A}$ - V5 (Tudor non-functional		61,919 / 120,222 (52%)

Cells were stained and assayed as described in main methods.

Appendix 24- qPCR data showing *Ven^{trunc}* expression in cells from Fig 2.3



Ven^{trunc} expression was measured in cells with synonymous and non-synonymous mutations in the Tudor and Zn MYND domains of *Ven^{trunc}* (3 days post infection with DAV). Y-axis shows -ΔCT relative to Rpl32. Each dot represents a well of cells in a 96 well plate. Red bars show the standard error of -ΔCT centred around the mean. Transfected cells are clones of a single cell.