A novel transposable element-mediated mechanism causes antiviral resistance in *Drosophila* through truncating the Veneno protein

Osama Brosh*[1], Daniel K. Fabian*[1], Rodrigo Cogni*[1,2], Ignacio Tolosana[1,3], Jonathan P Day[1], Francesca Olivieri[1], Manon Merckx[1], Nazli Akilli[1], Piotr Szkuta[1,4] and Francis M. Jiggins**[1]

* contributed equally
** corresponding author

[1] Department of Genetics, University of Cambridge, Cambridge, United Kingdom
[2] Department of Ecology, University of São Paulo, São Paulo, Brazil
[3] Current Address: Department of Life Sciences, Imperial College, London, UK
[4] Current Address: Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, UK

Competing Interests Statement

The authors declare no competing interest.

Abstract

Hosts are continually selected to evolve new defences against an ever-changing array of pathogens. To understand this process, we examined the genetic basis of resistance to the virus DAV in *Drosophila melanogaster.* In a natural population, we identified a polymorphic transposable element (TE) insertion that was associated with a ~19,000-fold reduction in viral titres, allowing flies to largely escape the harmful effects of infection by this virulent pathogen. The insertion occurs in the protein-coding sequence of the gene *Veneno,* which encodes a Tudor domain protein. By mutating *Veneno* with CRISPR-Cas9 in flies and expressing it in cultured cells, we show that the ancestral allele of the gene has no effect on viral replication. Instead, the TE insertion is a gain-of-function mutation that creates a gene encoding a novel resistance factor. Viral titres remained reduced when we deleted the TE sequence from the transcript, indicating that resistance results from the TE truncating the Veneno protein. This is a novel mechanism of virus resistance and a new way by which TEs can contribute to adaptation.

Significance Statement

Pathogens can drive rapid evolution in the species they infect, providing a model for how adaptations arise. We found that some genotypes of the fruit fly *Drosophila melanogaster* are resistant to a common viral pathogen called DAV. Resistance is caused by a transposable element that has inserted into the gene *Veneno*, resulting in the gene encoding a shortened protein. The original form of the protein has no effect on virus resistance, but when truncated it gains a new and potent antiviral function, providing a large fitness advantage to infected flies. This demonstrates a novel mechanism by which transposable elements can generate adaptive phenotypes.

Introduction

Pathogen infection is an important determinant of fitness in natural populations, and the nature of this selection pressure is continually changing as new pathogens appear or existing pathogens evolve to escape host defences. This can result in rapid evolutionary change and continual innovation in the immune defences of animals. Identifying these new defences can therefore provide insights into the genetic basis of adaptation and the mechanisms by which hosts counter infection in nature.

Transposable elements (TEs) can be an important source of genetic variation that natural selection can act on to generate new adaptations. In many species, including *Drosophila melanogaster,* TEs are highly dynamic within populations (1). Studies of both mutation accumulation lines and the frequency of TEs in populations have provided compelling evidence that TE insertions tend to be deleterious (2–4). However, TEs are also an important source of beneficial mutations that give rise to adaptations (5). A common mechanism of this involves the TE inserting upstream of genes and altering their expression. This is especially prevalent in the evolution of insecticide resistance, where TEs can lead to the upregulation of detoxification enzymes (6–8). In other cases, TEs or retroviruses have been 'domesticated' and play important roles in host biology (9). For example, *Syncytin* genes, which play a role in nutrient transfer across the placenta, have arisen through multiple independent domestication events across mammals and a species of placental lizard (10, 11). In mammals, TEs and endogenous retroviruses have played an important role in the evolution of the immune system, giving rise to enhancers regulating the expression of genes in response to interferon, and the RAG proteins that cleave DNA during V(D)J recombination (12, 13). In *D. melanogaster* TEs affect the expression of immunity genes (14), and we have reported a *Doc* element insertion that is associated with resistance to the virus DMelSV (15, 16).

In insects and other invertebrates, the absence of an adaptive immune system means infection must be controlled by innate immune defences (17), with the RNAi pathway being a key defence against viruses (18). While these core immune pathways evolve rapidly (19, 20), it is unclear what role they play in the evolution of resistance. In humans, a key component of antiviral defences is provided by a diverse collection of proteins that can inhibit viral replication by targeting almost any step of the viral replication cycle (21). Many of these are dominantly-acting cell-autonomous molecules known as restriction factors. Viruses have frequently evolved mechanisms to escape restriction factors and the restriction factors are often under positive selection, suggesting they are involved in an evolutionary arms race (21). Restriction factors have been little-studied in insects, but we have described several polymorphic genes in *Drosophila* that have large effects on viral replication (15, 16, 22–24). It is likely these play a critical role in host-virus evolution in insects, analogous to restriction factors in mammals.

Despite approximately a third of flies in wild populations of *Drosophila melanogaster* being infected with one or more viruses (25), many studies of antiviral immunity have used viruses that are rare or absent in nature. For this reason, we investigated resistance to DAV (Drosophila A Virus), which typically infects about 5% of flies in natural populations of *D.*

melanogaster (25, 26). DAV is a positive-sense single-stranded RNA virus that is related to the Permutotetraviridae (25, 27). Little is known about its interaction with *Drosophila* or its effects on the health of infected insects. To understand the evolution of virus resistance, we used a panel of inbred fly lines with publicly available whole genome sequences (28) to investigate the genetic factors that cause variation in susceptibility to DAV in nature.

Results

Some genotypes in a natural population are resistant to DAV

We investigated genetic variation in susceptibility to DAV, which is a common virus in natural populations of *D. melanogaster* (25). We used 182 inbred lines of flies from the DGRP collection, that were derived from flies collected from a natural population in North Carolina, USA (28). Across these fly lines, we infected a total of 11,985 female flies, extracted RNA from groups of 15 flies 3 days post infection, and estimated viral titres by quantitative PCR.

We found that there was substantial genetic variation in susceptibility to DAV, with 50% of the variance in viral titre being explained by genetic differences between the lines (Figure 1A; 95% CI: 43%-57%). Two lines were extremely resistant, having a mean viral titre ~19,000 times lower than the rest of the lines (Figure 1A). These lines explain the majority of the genetic variance in DAV resistance, and if they are excluded from the dataset only 20% of the variance in viral titre is explained by genetic effects. Over a time-course the reduction in viral titre was greatest shortly after infection, and by 28 days post infection the viral titre in the resistant flies was only slightly lower than in the susceptible flies (Figure 1B; ANOVA, days-post-infection x resistance-status: F=14.7, d.f.=1,226, p=0.0002).

We have previously measured the susceptibility of these lines to three other viruses—the positive strand RNA viruses DCV (*Dicistroviridae*) and FHV (*Nodaviridae*), and the negative strand virus DMelSV (*Rhabdoviridae*) (16). However, the two lines that are highly resistant to DAV are not resistant to these viruses (Figure S1A; DMelSV, linear model including *CHKov1* and *ref(2)P* genotype as cofactor: F=1.13,d.f.=1,185, p=0.29; DCV: F=0.70, d.f.=1,153, p=0.41; FHV: F=0.002, d.f.=1,180, p=0.96). Therefore, this mechanism of resistance is likely virus-specific.

Increased fecundity of resistant genotypes upon infection

As DAV is common in nature, we investigated how resistance affects the fitness of infected flies. It is known that different mechanisms protect flies against viral infection through different routes (29). Therefore, we first mimicked natural infection by feeding adults with live yeast paste that was contaminated with the virus. In an attempt to obtain an ecologically realistic dose, the virus was extracted from infected flies, mixed with yeast and water, and added to the fly vials such that the virus extracted from a single fly would be split between five vials. Seven days post infection, the median viral titre in the susceptible flies was ~10,260 times greater than in the resistant flies (Figure 2A; t=17.04, d.f.=307, $p<10^{-16}$). Therefore, resistance protects flies against oral infection with the virus.

Little is known about the effects of DAV on the fitness of *Drosophila,* so we measured the effects of DAV infection on fecundity. To avoid the effects of inbreeding depression, we crossed the inbred lines to a standard susceptible laboratory line and inoculated the F1 progeny with DAV or saline solution. Four days post infection, females from two susceptible lines laid 58%-63% fewer eggs than uninfected controls. In contrast, there was no reduction in the number of eggs laid by a resistant line (Figure 2B; Quasipoisson GLM, resistance status x infection status interaction: $\chi^2=25.3$, d.f.=1, $p=5 \times 10^{-7}$). The results were similar 11 days post-infection (dpi; Figure 2B; Quasipoisson GLM, resistance status x infection status interaction: $\chi^2=12.2$, d.f.=1, $p=0.0005$).

To investigate the effects of infection through a natural route, we examined the fecundity of flies after they had been fed food contaminated with DAV. DAV infection caused a 28% reduction in the fecundity of the susceptible flies but no reduction in the fecundity of the resistant genotypes (Figure 2C; Quasipoisson GLM, resistance status x infection status interaction: $\chi^2=3.8$, d.f.=1, $p=0.05$).

To investigate the the survival of flies we inoculated flies with two different doses of the virus and followed their survival for 65 days. DAV substantially reduced lifespan, but regardless of the dose mortality only increased from ~30 dpi (Figure S1). Despite this large effect on survival, there was no consistent difference in the survival of resistant and susceptible flies after infection at either dose (Figure S1). A possible explanation is that at the time when infected flies start to die there is little difference in the viral load of the resistant and susceptible flies (Figure 1B and Figure S1).

Increased resistance to infection is sometimes genetically correlated with reduced fitness in uninfected animals (30). As a measure of reproductive success, we measured the number of adult offspring produced by homozygous resistant and susceptible flies (data obtained by allowing the eggs laid over 20 hours by 72 uninfected females in Figure 2C to develop into adults). We found there was no significant difference between the reproductive success of resistant and susceptible flies (Poisson GLM: $\chi^2=1.63$, d.f.=1, $p=0.20$). In a separate experiment we measured survival from first instar larvae to adulthood in a subset of these lines, and again there was no difference between resistant and susceptible genotypes (N=185 larvae, Quasibinomial GLM: $\chi^2=0.001$, d.f.=1, $p=0.97$). Furthermore, we found that there were no significant correlations between DAV titre (Figure 1A) and published measurements of lifespan and fecundity made on these inbred lines (DGRP panel; Table S1).

Resistance is caused by a single dominant major-effect locus

Our next goal was to characterize the genetic basis of resistance to DAV. We found that resistance is genetically dominant, with the F1 progeny of a cross having the phenotype of the resistant parent (Figure 3A; Tukey HSD Test, susceptible vs heterozygote: $p<0.00001$, resistant vs heterozygote: $p=0.59$). To identify which chromosomes affect susceptibility, we generated flies which carry a single chromosome from a resistant line and two chromosomes from a susceptible line (Figure 3B). When these flies were infected, viral titres

were strongly reduced in lines where chromosome 3 came from the resistant parent (Figure 3B; Tukey HSD Test, chr 2 vs 3: $p<0.00001$, chr X vs 3: $p=0.00001$). There was also a small effect of the X chromosome (Figure 3B; Tukey HSD Test, chr 2 vs X: $p=0.04$).

To map the region of chromosome 3 that controls resistance, we crossed two pairs of resistant and susceptible lines and created two panels of lines carrying recombinant third chromosomes. We genotyped molecular markers across the chromosome, and inferred genotype probabilities between these markers (31). In both crosses we identified a single quatitative trait locus (QTL) associated with resistance, and in both cases this mapped to the same region (Figure 3C; line 239 cross: 47-49cM; line 362 cross: 46-50cM).

The genetic mapping results suggest that the same allele might be responsible for DAV resistance in the two different lines. We therefore searched the genome sequences of the 182 inbred lines that we used in our infection experiments (Figure 1A) for consistent single nucleotide polymorphism (SNPs) differences between resistant and susceptible lines. We found 26 such polymorphisms, 12 of which were on chromosome 3 (Figure 3D). All of these fell within the QTL we identified in our genetic mapping experiments (Figures 3C and 3D). Together these results demonstrate that a single major-effect polymorphism on chromosome 3 controls susceptibility to DAV.

High resolution genetic mapping identifies the region controlling resistance

The QTL controlling resistance contained many genes, and the recombination rate in this region of the genome made it impractical to continue using simple genetic crosses to identify the causative gene. To overcome this, we devised a genetic cross to identify recombinants within the QTL based on eye color (Figure 4A). In this cross, non-recombinant flies either had white eyes, or died as a result of a homozygous recessive lethal allele or the expression of a lethal gene controlled by a heat shock promoter. Using this approach, we generated 643 lines that were recombinant within the QTL.

We scored five molecular markers across the QTL, which allowed us to identify six genotypes with different recombination breakpoints (Figure 4B). We retained 219 lines for phenotyping, and found that they varied considerably in susceptibility to infection (Figure 4C; ANOVA: $F=129$, d.f.=5,258, $p<0.00001$). Within Genotype B there was a mixture of resistant and susceptible flies (Figure 4C), indicating that the gene was within the recombinant region of these lines (between the dotted lines in Figure 4B).

To identify the region affecting DAV resistance more precisely we carried out additional phenotyping and genotyping of lines that had recombined in this region. First, we genotyped a high density of molecular markers to precisely define recombination breakpoints in the lines (Figure 4D; all recombinants in this region were retained from our panel of 643 lines). Second, we performed additional infection experiments to ensure we had accurate estimates of DAV susceptibility in recombinant lines whose breakpoints defined the location of the gene (Figure 4E). Combining these two datasets, we found that the polymorphism controlling DAV susceptibility was in a region of 33,847bp encompassing 11 genes (Figure 4D; 3R:4163320..4197167 in genome v5).

A transposable element insertion in *Veneno* is associated with resistance

To test which of these genes underlies virus resistance, we knocked down the expression of the allele of these genes found in resistant flies by RNAi. Of the nine genes where this was successful, eight remained resistant to DAV (Figure 5A). However, when we knocked down the expression of the gene *CG9684*, which we named *Veneno* (*Ven*) after its ortholog in *Aedes aegypti* (32), there was a ~2000-fold increase in viral titre compared to the resistant control, resulting in similar viral titres to susceptible flies (Figure 5A; Welch's *t* test, Susceptible vs *Ven* knock-down: $t = 0.77$, d.f. = 11.0, $p=0.46$, Resistant vs *Ven* knock-down: $t = 7.88$, d.f. = 28.2, $p=10^{-8}$). Alongside this experiment we measured the transcription of the candidate genes in adult females. Again, *Veneno* was the only gene where there was a substantial difference in expression between the resistant and susceptible lines (Figure S2A; Welch's *t* test: $t=12.51$, d.f.=5.81, $p=0.00002$). In line with this, using published microarray expression data from the DGRP panel (33), we found that *Ven* expression had a higher correlation with DAV resistance than any other gene in females, but showed no relationship in males (Pearson's $r=0.5$, FDR=$7.5\text{x}10^{-12}$ in females and $r=0.04$, FDR=0.99 in males, Figure S2B, Supplementary Dataset 1 and 2).

The published genome sequences of the resistant and susceptible lines were generated with short read sequencing (28) and therefore may not include structural variants such as insertions. We therefore amplified *Veneno* from genomic DNA by PCR. While the susceptible line yielded PCR products of the expected size, the resistant allele contained a large insertion. Through a series of PCR reactions, we amplified and Sanger-sequenced a 4685bp insertion in the protein-coding sequence of exon 3 of *Veneno* (Figure 5B; GenBank Accession: MZ047782). Using BLAST, we identified the insertion as a *Doc* element, which is a non-LTR retrotransposon. Compared to the published full-length *Doc* element sequence (34), our sequence has a 37bp deletion at the 5' end, three single nucleotide mismatches and a single nucleotide deletion (DGRP-362 sequence; Supplementary Dataset 3). We named the allele of *Veneno* carrying the *Doc* element Ven^{Doc} and the allele without the insertion Ven^{+}.

To test whether the *Doc* insertion is associated with DAV resistance, we checked for the presence of the insertion in 162 DGRP lines. We genotyped these lines using two primers on either side of the *Doc* element insertion and one within, which results in different sized PCR products when amplifying Ven^{Doc} and Ven^{+} (Figure 1C). We found that the *Doc* element insertion is perfectly associated with the DAV resistance (Figure 1A; Fisher's Exact Test: $p=0.00008$). By sequencing both breakpoints between the *Doc* element and *Veneno,* we confirmed that the insertion was in the identical location in the two resistant lines. Therefore, in this experiment Ven^{Doc} is associated with a ~19,000-fold reduction in DAV titre.

Truncation of *Veneno* created a novel resistance factor

Many TE insertions associated with adaptive traits result from the insertion altering gene expression. In the experiments above we found that *Ven* expression is correlated with DAV titre in females. This correlation is entirely caused by the Ven^{Doc} allele having reduced

expression (Figure S2A), as when these lines are removed there is no correlation between *Ven*$^{+}$ expression and DAV titre (Pearson's correlation in females: r=0.06, p=0.43). Furthermore, using quantitative PCR we found that *Ven*^{Doc} expression was only reduced in in female abdomens–in males and the female thorax the two alleles were expressed at similar levels or the resistant allele had slightly higher expression (Figure S2C). The epigenetic silencing of TE insertions in the *Drosophila* genome can reduce the expression of nearby genes in the female germline (35), which may explain the reduced expression of *Ven*^{Doc} in female abdomens. Regardless of its causes, because *Ven*^{Doc} confers DAV resistance in both males and females (Figure S2D), it is unlikely that resistance results from the *Doc* element insertion altering gene expression.

This is the second *Doc* element insertion that is associated with virus resistance in *Drosophila* – we previously reported a *Doc* insertion that is associated with resistance to the rhabdovirus DMelSV (15, 16). This led us to hypothesise that *Doc* elements may have intrinsic antiviral activity. To test this hypothesis, we created transgenic flies expressing RNAi constructs targeting two different regions of *Doc.* However, despite successfully knocking down *Doc* expression, this did not affect DAV titres (Figure S3).

To investigate the effect of *Ven*$^{+}$ and *Ven*^{Doc} on susceptibility to DAV, we mutated the sequence upstream and downstream of the *Doc* insertion using CRISPR/Cas9. We created flies that express Cas9 and either *Ven*$^{+}$ or *Ven*^{Doc}. Alongside this, we generated two transgenic fly lines which expressed a CRISPR/Cas9 guide RNA molecule targeting either a *Veneno* exon before the *Doc* insertion or after the insertion. Again, each of these transgenes was crossed into a line carrying either *Ven*$^{+}$ or *Ven*^{Doc}. We crossed these lines, producing F1 progeny that express Cas9 and guide RNAs throughout somatic cells, with the aim of producing somatic mutations. By amplifying the Cas9 cutting sites by PCR and Illumina sequencing the products, we estimated that this strategy mutated 82.3-99.7% of the chromosomes (Figure 6A; Supplementary Dataset 4). These flies were then infected by DAV. Mutation of *Ven*^{Doc} upstream of the insertion resulted in a ~200,000 fold higher viral titre than control *Ven*^{Doc} flies (Figure 6B, Tukey's HSD Test: $p<10^{-7}$). Mutation of the same region of *Ven*$^{+}$ had no effect on viral titre, indicating that the *Doc* element insertion is a gain-of-function mutation (Figure 6B, Tukey's HSD Test: p=0.95). In contrast, when either allele of *Veneno* was mutated downstream of the insertion there was no effect on viral titres (Figure 6B; Tukey's HSD Tests: p>0.95).

These results suggest that *Ven*^{Doc} encodes a recently-evolved antiviral molecule. As the *Doc* element is found in locations throughout the genome, we were unable to reconstruct transcripts encoded by *Ven*^{Doc} from short sequence reads. We therefore used the long-read Oxford Nanopore Technologies platform to sequence the transcriptome of a line carrying *Ven*^{Doc}. This allows us to sequence the full-length transcripts in a single read, and we mapped 8.8 million reads (NCBI SRA: SRR15541957) to a genome sequence which we had modified to include the *Doc* insertion in *Ven*. We identified three polyadenylated transcripts: Transcript 1 (4 reads), Transcript 2 (39 reads), and Transcript 3 (4 reads) (Figure 6C). We orientated the reads using the poly-A tail sequence, and all the reads were transcribed from the sense strand of the gene. Only Transcript 1 contained regions which were targeted by the *Veneno* RNAi knock-down that eliminated the resistant phenotype (Figures 6C and 5A). Furthermore, this was the only transcript containing an open reading

frame ending with a stop codon, and with the same intron structure as *Ven*$^{+}$ (Figure 6C). *Ven*^{Doc} Transcript 1 contains the *Veneno* 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 (Figure 6C).

To test whether this transcript encodes an antiviral molecule, we expressed *Ven*^{Doc} Transcript 1 in cultured *Drosophila* cells. DAV readily infects and replicates in DL2 cells (Figure S4A). We therefore stably transfected this cell line with a construct expressing V5-tagged *Ven*^{Doc} Transcript 1 under the control of an inducible promoter. To ensure all the cells expressed the construct, we then established a clonal cell line from the transfected cells. When these cells were infected, expressing *Ven*^{Doc} Transcript 1 led to a ~7 fold reduction in DAV titre 3 dpi (Figure 6D, *Ven*^{Doc} vs untransfected; Tukey's HSD Test: p=0.00003). Similar results were obtained using non-clonal cells expressing *Ven*^{Doc} Transcript 1 without the tag (Figure S4B). Both a Western blot and flow cytometry confirmed that these cells were expressing the truncated protein (Figure S4C-E).

The antiviral effects of *Ven*^{Doc} could either result from the *Doc* insertion truncating the Veneno protein, or from a new function that requires a chimeric Veneno-Doc protein. To distinguish these possibilities, we established a clonal cell line that was stably transfected with a construct expressing *Ven*^{Doc} Transcript 1 from which we had deleted the sequence derived from the *Doc* element. When we induced expression of this transgene, these cells were resistant to DAV (Figure 6D, *Ven*^{truncated} vs untransfected; Tukey's HSD Test: p=0.0003). Therefore, the *Doc* insertion has created a new antiviral molecule by truncating Veneno, and resistance does not require a chimeric Veneno-Doc protein.

Resistance does not require the RNAi pathway or the Tudor or MYND domains in *Veneno*

The *Doc* element insertion alters the domain structure of Veneno. The susceptible allele of *Veneno* (*Ven*$^{+}$) encodes a protein with two Tudor domains, which are predicted to bind methylated arginine or lysine residues (Figure 7A) (36). There is also an MYND Zinc finger domain, which is another domain normally involved in protein-protein interactions (37). These domains are also present in the ortholog of this gene in the mosquito *Aedes aegypti,* and here the protein acts as an adaptor protein that interacts with proteins in the antiviral piRNA pathway found in this species (32). The resistant allele (*Ven*^{Doc}) encodes a molecule that has lost the Tudor domain at the carboxyl-terminal end of the protein (Figure 7A).

One hypothesis is that *Veneno* has an adaptor function which facilitates the formation of a protein complex, and resistance results from changes to this complex. We tested whether the Tudor or MYND Zinc Finger domains were necessary for DAV resistance. We used site directed mutagenesis to target residues essential for the function of these domains in the constructs used to transfect cells. In the MYND Zinc Finger domain we mutated the histidine at residue 51 in the protein to an alanine (His51Ala). This histidine is required to chelate zinc, and is therefore required for the correct functioning and folding of this domain (38). In the Tudor domain we mutated the tyrosine at position 198 to alanine (Tyr198Ala). The tyrosine is involved in forming the binding pocket for the symmetric demethylation of arginine, which allows protein-protein interactions (32, 39). We then stably transfected cells

with these plasmids, and generated clonal cell lines to ensure all cells were transfected. Neither the Tudor nor MYND Zinc Finger domains are necessary for resistance, as cell lines expressing both $Ven^{Doc,\ His51Ala}$ and $Ven^{Doc,\ Tyr198Ala}$ had substantially reduced DAV titres (Figure 7B; Tukey's HSD Tests, $p<10^{-7}$).

The main antiviral defence of insects is RNAi, and Tudor domain proteins frequently function as adaptor proteins in small RNA pathways. Furthermore, Veneno physically interacts with Hen1, which methylates siRNAs, (40) and R2D2 (41), which loads siRNAs into the RISC complex to guide the enzymatic shearing of viral RNA. We therefore tested the hypothesis that resistance requires the RNAi pathway by combining mutations in RNAi pathway genes *Dicer-2* ($Dcr\text{-}2^{R416X}$) and *Hen1* ($Hen1^{f00810}$) with the two alleles of *Veneno*. We found that Ven^{Doc} still made flies resistant to DAV when combined with these mutations, indicating that resistance caused by *Veneno* does not require the siRNA pathway (Figure 7C and 7D). We also found that knocking out the RNAi pathway did not increase the viral titre, which was unexpected as this pathway is thought to protect flies against a broad spectrum of viruses (Figure 7C and 7D). We confirmed this by infecting three further RNAi pathway mutant lines ($Dcr\text{-}2^{L811fsx}$, $Ago2^{51B}$ and $Ago2^{414}$), again finding no increase in viral titre (Figure S5A). Together these results demonstrate that RNAi is not an important defence against DAV, and the antiviral effects of Ven^{Doc} do not rely on this pathway.

In *Aedes aegypti*, Veneno acts as an adaptor protein that assembles a protein complex involved in the production of piRNAs from viral RNA (32). To test whether the mechanism of Ven^{Doc} resistance depends on the piRNA pathway, we knocked down *Ago3, vret* and *vas* expression by RNAi in flies carrying both the resistant and susceptible allele of *Veneno.* In all three cases there was no effect on DAV titres (Figure S5B). This is consistent with previous results showing that in *Drosophila* the piRNA pathway is restricted to the germline and plays no role in antiviral immunity (42, 43). Together our results demonstrate that neither small RNA pathways nor the domains involved in protein-protein interactions are necessary for resistance to DAV.

Discussion

As new pathogens appear in populations and existing pathogens evolve to escape immunity, there is continual natural selection favouring novel host defences. We have found that a TE insertion into the protein-coding sequence of the Tudor domain protein *Veneno* has resulted in the gene gaining an antiviral function. The resistant allele of the gene encodes a truncated protein that acts as a potent resistance factor that massively reduces titres of DAV, while the ancestral susceptible form of the protein has no effect on the virus. As DAV is common in nature (25) and causes large reductions in the fecundity of susceptible flies, this allele protects flies against a virulent pathogen. This adds to a growing body of evidence from multiple species of animals that much of the genetic variation in susceptibility to naturally-occurring pathogens is explained by a small number of major-effect polymorphisms (16, 52, 53). This contrasts with most quantitative traits which tend to be controlled by many variants with small phenotypic effects (52, 53). The simple genetic architecture likely results from the evolutionary arms race between hosts and their pathogens driving major-effect alleles up in frequency (52).

The study of viral immunity in invertebrates has been dominated by investigations of broad-spectrum and conserved antiviral defences such as RNAi and autophagy. However, our discovery of *Ven^{Doc}* adds to the list of major-effect polymorphisms which cause virus resistance in *Drosophila*, such as *CHKov1*, *ref(2)P*, *pastrel* and *Ge-1* (15, 16, 22–24). The mechanism by which these genes protect flies against viral infection is unclear, but they may encode restriction factors, analogous to those that play a critical role in defending mammals against viruses. Regardless of mechanism, these genes are central to the defences of *Drosophila* against viruses as they can have large effects on susceptibility. They differ from conserved antiviral pathways such as RNAi in two ways. First, they mostly protect their hosts against a narrow range of viral taxa. Therefore, despite their central importance to antiviral defence, they may be missed by studies using a single virus that may have been isolated from a different species. Second, they are mostly recent evolutionary innovations and are polymorphic in populations. These antiviral factors therefore arise because natural selection is continually generating new defences against the viruses encountered in nature.

TEs frequently underlie adaptations novel selection pressures (8, 44–46). In many cases these adaptive TE insertions alter gene expression, often by inserting upstream of the gene. For example, insecticide resistance has repeatedly evolved when TE insertions upregulate the expression of detoxification enzymes (8, 45, 46). TE insertions can also generate new adaptations when the element itself is recruited to a host function, a process known as domestication. This can involve the element fusing to another gene in the genome. For example, the gene *SETMAR* in primates is a chimera of the gene *SET* and the transposon *Hsmar1,* which retains functions of both a TE domain and the original gene (47–49). In a striking parallel with *Ven^{Doc},* we have previously reported how a *Doc* element insertion into the *Drosophila* gene *CHKov1* is associated with resistance to a rhabdovirus (15, 16). This raised the possibility that *Doc* element sequences may be recruited to a new antiviral function in these gene-TE chimeras. However, we did not find support for the gene-TE chimera hypothesis, nor for resistance resulting from a change in gene expression.

Resistance is instead caused by the TE-dependent truncation of Veneno, resulting from the TE insertion shortening the transcript and prematurely terminating translation. To our knowledge, this is the first demonstration of such a mechanism giving rise to a protein with a novel function. Other cases of TEs introducing stop codons into the coding sequence of genes are thought to be simple loss-of-function mutations. For example, a *Hel-1* LTR retrotransposon insertion into the *HevCalP* gene in the moth *Heliothis virescens* introduces a stop codon (50). This allele is resistant to *Bacillus thuringiensis* (Bt) toxins used in pest control, but resistance is recessive as the truncation results in a loss of function of the host protein (50). In contrast, DAV resistance is a gain of function mutation.

The mechanism of resistance may rely on *Veneno* playing a role in RNA biology. Ortholog of *Veneno* of *Veneno* (32) plays a role in piRNA biogenesis , while Veneno physically interacts with components of the siRNA pathway (41). However, *Ven^{Doc}* resistance did not require a functional siRNA or piRNA pathway. Veneno also has three domains involved in protein-protein interactions, but when we mutated residues that are essential for the function of these domains, *Ven^{Doc}* still conferred virus resistance. The molecular mechanism by which *Veneno* protects *Drosophila* against viral infection therefore remains uncertain. However, the observation that Veneno interacts with proteins involved in RNA biology suggests that

some currently unknown function of the protein has been recruited to a novel antiviral function.

Despite effectively protecting flies against a virulent pathogen, the resistant allele of *Veneno* is found at a frequency of just over 1%, so only about 1 in 50 flies in this population will be resistant. This could result from resistance having arisen recently, so there is insufficient time for the allele to reach a high frequency. Alternatively, the benefits of resistance may be balanced by costs. In *Drosophila*, the Tudor domain proteins Qin, Krimper and Tudor-SN are all expressed in the germline where they play a role in the piRNA pathway (36, 54), and the high expression of the susceptible allele of *Veneno* in ovaries suggests it may have a related function. As the resistant allele has greatly reduced expression in ovaries and has lost a Tudor domain, it is likely its original function will have been changed. While our data suggests that in infected flies the benefits of resistance likely outweigh the costs, we lack the statistical power to detect more modest costs of Ven^{Doc}. If these exist, they could result in the resistant allele being under balancing selection, either due to heterozygote advantage or negative frequency dependant selection.

The strong and rapidly changing selection pressures that pathogens impose on host populations provide a model to study the genetics of adaptation. Our work demonstrates a new mechanism by which TEs can generate these adaptations. The role of TEs in adaptation to DAV infection mirrors studies of how populations have adapted to the selection pressures that humans impose populations when they use pesticides or alter the environment (8, 44–46). The importance of TEs in evolution likely stems not simply from them being a major cause of mutation, but also because of features of these mutations such as the magnitude of the change in DNA sequence and biases in where the elements insert. It is clear that alongside the harm they cause their hosts, TEs can allow populations to evolve to overcome a diverse range of challenges.

Materials and Methods

Resistant and susceptible lines were identified by infecting the DGRP lines with DAV and measuring viral loads by qPCR. We then mapped resistance by using balancer chromosomes to create chromosome substitution lines, followed by the generation of recombinant inbred lines, and finally using the cross in Figure 4A. Genes within the QTL were knocked down using transgenic RNAi constructs, and *Ven* was mutated by expressing cas9 and gRNAs in transgenic flies. To identify protein domains involved in resistance we transfected *Drosophila* cells with plasmids that express modified alleles of *Ven.* Detailed methods are in Supplementary Information.

Acknowledgements

We thank Karyn Johnson for providing DAV.

Funding

This work was funded by grants from the Natural Environment Research Council (NE/P00184X/1) and the Leverhulme Trust (RPG-2020-236) to FJ. RC is funded by the São Paulo Research Foundation (FAPESP) (2013/25991-0 and 2015/08307-3), the National Council for Scientific and Technological Development (CNPq) (307447/2018-9) and a

Newton Advanced Fellowship from the Royal Society (NAF\R1\180244). OB is funded by the Dr. Herchel Smith Fellowship.

Bibliography

1. R. Kofler, V. Nolte, C. Schlötterer, Tempo and Mode of Transposable Element Activity in Drosophila. *PLoS Genet.* **11** (2015).
2. D. Houle, S. V Nuzhdin, Mutation accumulation and the effect of copia insertions in Drosophila melanogaster. *Genet. Res.* **83**, 7–18 (2004).
3. T. F. Mackay, R. F. Lyman, M. S. Jackson, Effects of P element insertions on quantitative traits in Drosophila melanogaster. *Genetics* **130**, 315–332 (1992).
4. B. Charlesworth, P. Sniegowski, W. Stephan, The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).
5. J. González, D. A. Petrov, The adaptive role of transposable elements in the Drosophila genome. *Gene* **448**, 124–33 (2009).
6. P. J. Daborn, *et al.*, A single p450 allele associated with insecticide resistance in Drosophila. *Science* **297**, 2253–6 (2002).
7. J. M. Schmidt, *et al.*, Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet.* **6**, e1000998 (2010).
8. T. A. Schlenke, D. J. Begun, Strong selective sweep associated with a transposon insertion in Drosophila simulans. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1626–31 (2004).
9. H. S. Malik, S. Henikoff, Positive selection of Iris, a retroviral envelope-derived host gene in Drosophila melanogaster. *PLoS Genet.* **1**, e44 (2005).
10. H. S. Malik, Retroviruses push the envelope for mammalian placentation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2184–5 (2012).
11. G. Cornelis, *et al.*, An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proc. Natl. Acad. Sci. U. S. A.* **114** (2017).
12. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science (80-.).* **351** (2016).
13. L. M. Carmona, D. G. Schatz, New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J.* **284** (2017).
14. A. Ullastres, M. Merenciano, J. González, Regulatory regions in natural transposable element insertions drive interindividual differences in response to immune challenges in Drosophila. *bioRxiv*, 655225 (2021).
15. M. M. M. Magwire, F. Bayer, C. L. C. L. Webster, C. Cao, F. M. F. M. Jiggins, Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a Duplication. *PLoS Genet.* **7**, e1002337 (2011).
16. M. M. M. Magwire, *et al.*, Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in Drosophila melanogaster. *PLoS Genet.* **8**, e1003057 (2012).
17. D. K. Fabian, M. Fuentealba, H. M. Dönertaş, L. Partridge, J. M. Thornton, Functional conservation in genes and pathways linking ageing and immunity. *Immun. Ageing* **18** (2021).
18. D. J. Obbard, K. H. J. Gordon, A. H. Buck, F. M. Jiggins, The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. B Biol. Sci.*

364, 99–115 (2009).
19. D. J. Obbard, F. M. Jiggins, D. L. Halligan, T. J. Little, Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* **16** (2006).
20. D. J. Obbard, J. J. Welch, K.-W. Kim, F. M. Jiggins, Quantifying adaptive evolution in the Drosophila immune system. *PLoS Genet.* **5** (2009).
21. S. F. Kluge, D. Sauter, F. Kirchhoff, SnapShot: Antiviral Restriction Factors. *Cell* **163** (2015).
22. C. Cao, M. M. M. Magwire, F. Bayer, F. M. Jiggins, A Polymorphism in the Processing Body Component Ge-1 Controls Resistance to a Naturally Occurring Rhabdovirus in Drosophila. *PLoS Pathog.* **12** (2016).
23. J. Bangham, D. J. Obbard, K.-W. Kim, P. R. Haddrill, F. M. Jiggins, The age and evolution of an antiviral resistance mutation in Drosophila melanogaster. *Proc. R. Soc. B Biol. Sci.* **274** (2007).
24. C. Cao, R. Cogni, V. Barbier, F. M. Jiggins, Complex coding and regulatory polymorphisms in a restriction factor determine the susceptibility of Drosophila to viral infection. *Genetics* **206** (2017).
25. C. L. Webster, *et al.*, The discovery, distribution, and evolution of viruses associated with drosophila melanogaster. *PLoS Biol.* (2015) https:/doi.org/10.1371/journal.pbio.1002210.
26. R. Cogni, S. D. Ding, A. C. Pimentel, J. P. Day, F. M. Jiggins, Wolbachia reduces virus infection in a natural population of Drosophila. *Commun. Biol.* **4** (2021).
27. R. L. Ambrose, *et al.*, Drosophila A virus is an unusual RNA virus with a T=3 icosahedral core and permuted RNA-dependent RNA polymerase. *J. Gen. Virol.* **90** (2009).
28. T. F. C. Mackay, *et al.*, The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
29. Á. G. Ferreira, *et al.*, The Toll-Dorsal Pathway Is Required for Resistance to Viral Oral Infection in Drosophila. *PLoS Pathog.* (2014) https:/doi.org/10.1371/journal.ppat.1004507.
30. K. A. McKean, C. P. Yourth, B. P. Lazzaro, A. G. Clark, The evolutionary costs of immunological maintenance and deployment. *BMC Evol. Biol.* **8**, 76 (2008).
31. C. S. Haley, S. A. Knott, A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity (Edinb).* **69** (1992).
32. J. Joosten, *et al.*, The Tudor protein Veneno assembles the ping-pong amplification complex that produces viral piRNAs in Aedes mosquitoes. *Nucleic Acids Res.* **47** (2019).
33. W. Huanga, *et al.*, Genetic basis of transcriptome diversity in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **112** (2015).
34. K. O'Hare, M. R. K. Alley, T. E. Cullingford, A. Driver, M. J. Sanderson, 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. *MGG Mol. Gen. Genet.* **225** (1991).
35. Y. C. G. Lee, G. H. Karpen, Pervasive epigenetic effects of drosophila euchromatic transposable elements impact their evolution. *Elife* **6** (2017).
36. M. C. Siomi, T. Mannen, H. Siomi, How does the royal family of tudor rule the PIWI-interacting RNA pathway? *Genes Dev.* **24** (2010).
37. R. Gamsjaeger, C. K. Liew, F. E. Loughlin, M. Crossley, J. P. Mackay, Sticky fingers: zinc-

fingers as protein-recognition motifs. *Trends Biochem. Sci.* **32** (2007).
38. Y. Liu, *et al.*, Structural Basis for Recognition of SMRT/N-CoR by the MYND Domain and Its Contribution to AML1/ETO's Activity. *Cancer Cell* **11** (2007).
39. H. Liu, *et al.*, Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev.* **24** (2010).
40. K. G. Guruharsha, *et al.*, A protein complex network of Drosophila melanogaster. *Cell* **147** (2011).
41. K. Majzoub, "The antiviral siRNA interactome in Drosophila melanogaster.," Universite de Strasbourg. (2013).
42. M. Petit, *et al.*, PiRNA pathway is not required for antiviral defense in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **113** (2016).
43. S. H. Lewis, *et al.*, Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat. Ecol. Evol.* **2** (2018).
44. A. E. V. t. Hof, *et al.*, The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534** (2016).
45. W. G. Rostant, N. Wedell, D. J. Hosken, "Transposable elements and insecticide resistance" in *Advances in Genetics*, (2012).
46. R. Ffrench-Constant, P. Daborn, R. Feyereisen, Resistance and the jumping gene. *BioEssays* **28** (2006).
47. R. Cordaux, S. Udit, M. A. Batzer, C. Feschotte, Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl. Acad. Sci. U. S. A.* **103** (2006).
48. M. Tellier, R. Chalmers, Human SETMAR is a DNA sequence-specific histone-methylase with a broad effect on the transcriptome. *Nucleic Acids Res.* **47** (2019).
49. D. Liu, *et al.*, The Human SETMAR Protein Preserves Most of the Activities of the Ancestral Hsmar1 Transposase . *Mol. Cell. Biol.* **27** (2007).
50. L. J. Gahan, F. Gould, D. G. Heckel, Identification of a gene associated with Bt resistance in Heliothis virescens. *Science (80-.).* **293** (2001).
51. A. L. Zamparini, *et al.*, Vreteno, a gonad-specific protein, is essential for germline development and primary pirna biogenesis in Drosophila. *Development* **138** (2011).
52. E. M. Duxbury, *et al.*, Host-pathogen coevolution increases genetic variation in susceptibility to infection. *Elife* (2019) https:/doi.org/10.7554/elife.46440.
53. A. V. S. Hill, Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 840–9 (2012).
54. K. Sato, Y. W. Iwasaki, H. Siomi, M. C. Siomi, Tudor-domain containing proteins act to make the piRNA pathways more robust in Drosophila. *Fly (Austin).* **9** (2015).

Figure Legends

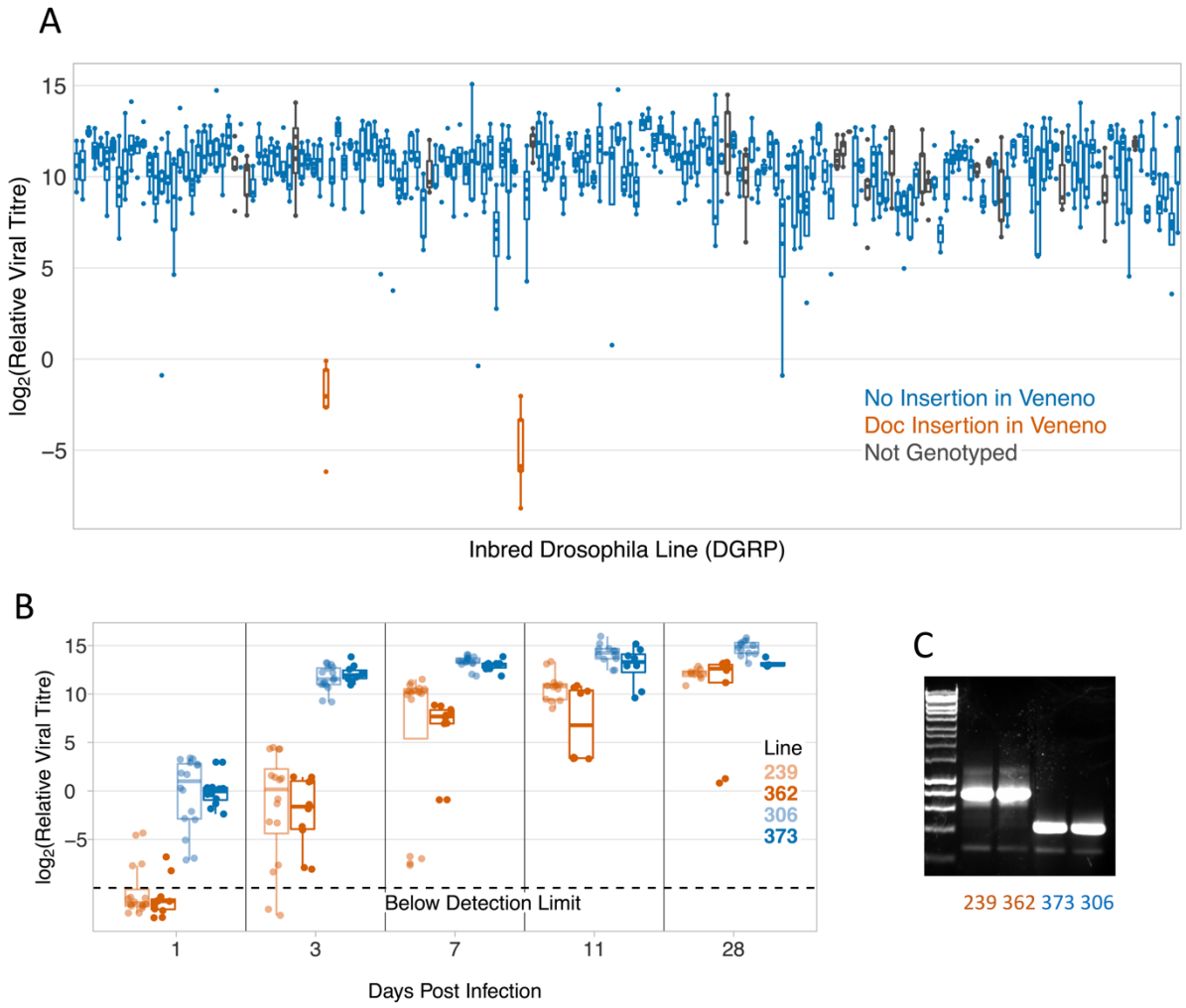

Figure 1. Genetic variation in susceptibility to DAV in a natural population. (A) Viral titre in 182 inbred lines of *D. melanogaster* 3 dpi. (B) Viral titre over a time-course. In both panels each point is from ~15 female flies infected by intrathoracic inoculation. Titre was estimated using qPCR relative to mRNA from the *Drosophila* gene *EF1α100E*. Points below the detection limit are jittered vertically. The box shows the median and interquartile range. (C) PCR genotyping for the presence of the *Doc* insertion.

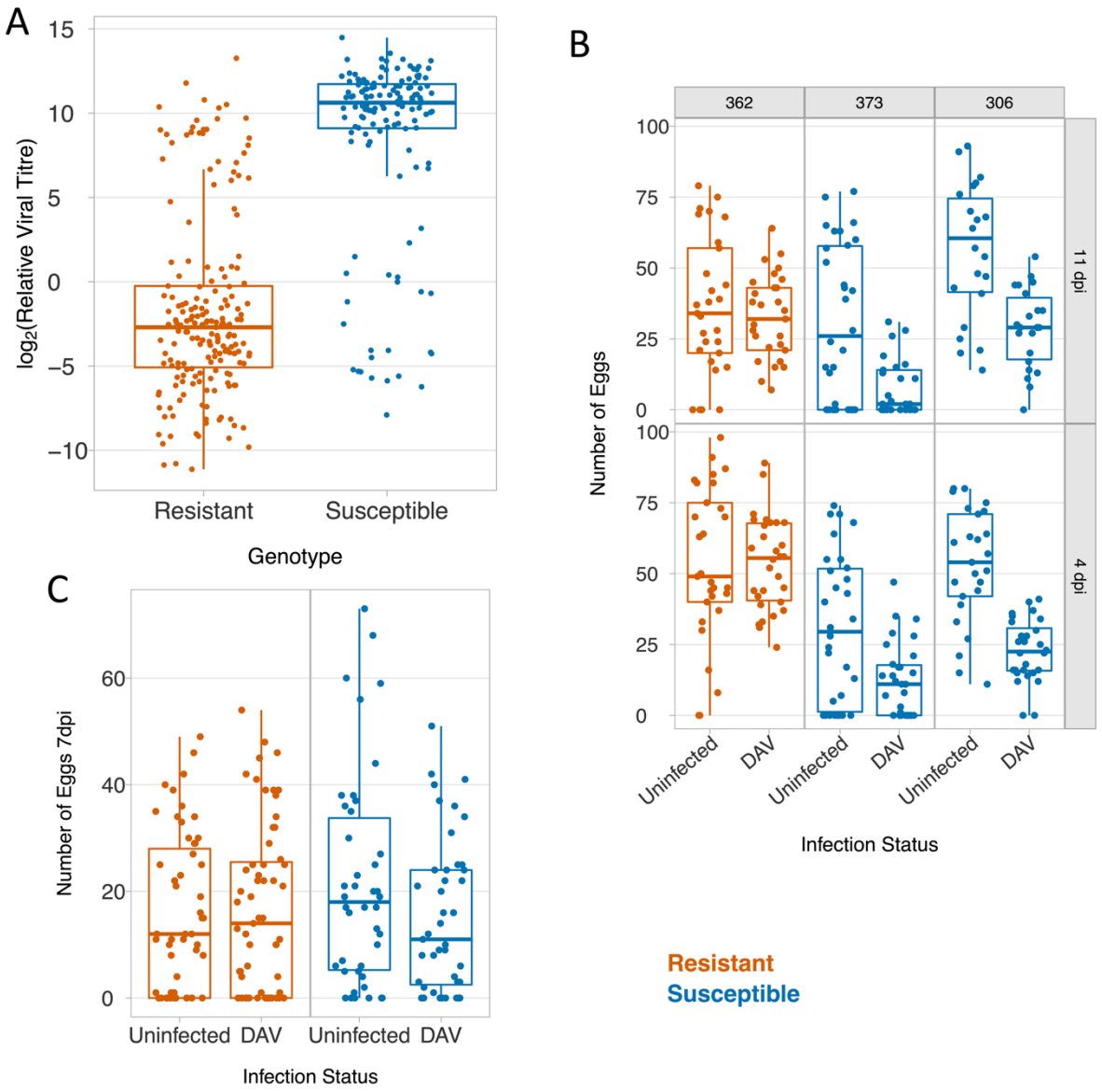

Figure 2. The effect of resistance on host fitness after infection. (A) Viral titre relative to *EF1α100E* mRNA 7 days after oral infection. (B/C) The effect of DAV infection on the number of eggs laid by single females over 20 hours following infection by intrathoracic inoculation (B) or oral infection (C). In A and C, the resistant and susceptible genotypes are recombinants between resistant and susceptible DGRP lines (see 'high resolution genetic mapping' for details). In panel B, the three *Drosophila* lines are the F1 progeny of a cross between a line from the DGRP panel and an isogenic w^{1118} line.

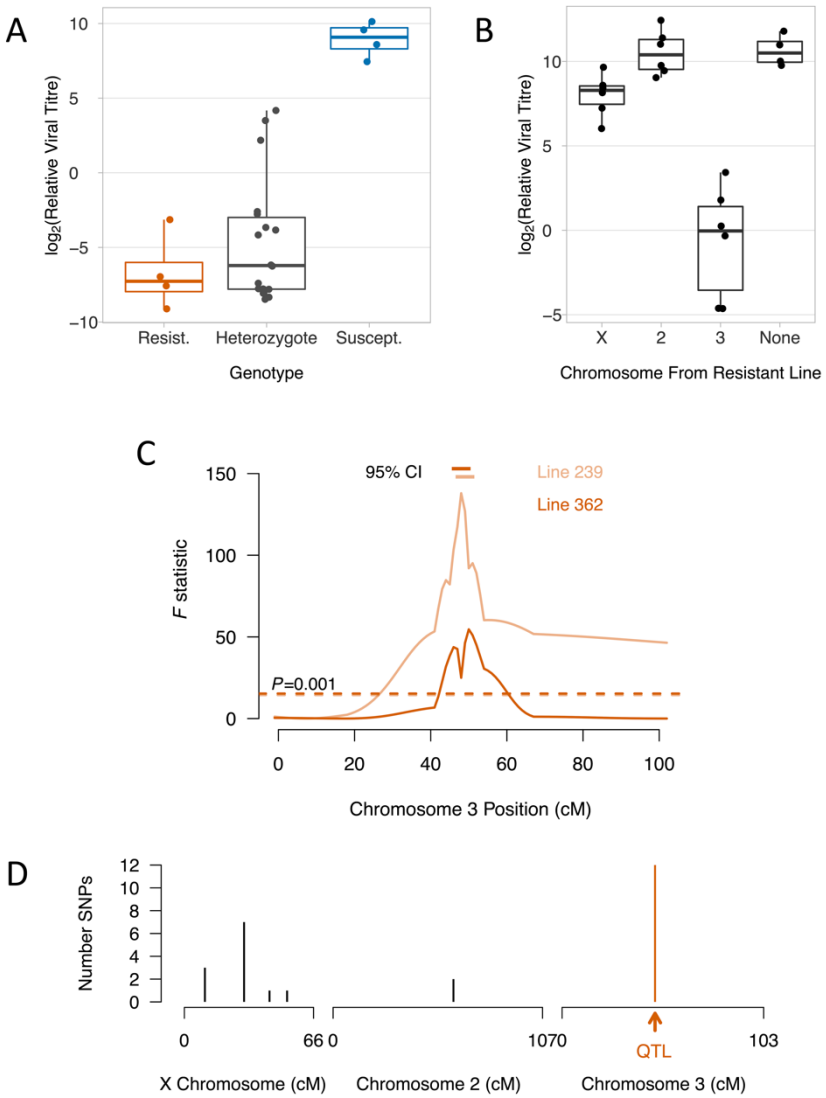

Figure 3. Genetic mapping of DAV resistance. (A) Dominance in resistant flies (DGRP 239 and 362), susceptible flies (DGRP 48 and 91) and F1 progeny. Pairs of lines were combined as they did not differ significantly. Each point is 20 females. (B) Resistance was mapped to chromosome by substituting chromosomes between lines. (C) The resistance QTL was mapped by crossing DGRP 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 markers inferred (31). The dashed line is a significance threshold obtained by permutation. The horizontal bars are 95% bootstrap intervals on the QTL location. In A-C flies were infected by inoculation. Viral titre was estimated 3 dpi relative to *EF1α100E* mRNA. (D) The genomic location of SNPs for which the two resistant lines one allele and the 180 susceptible lines had a different allele.

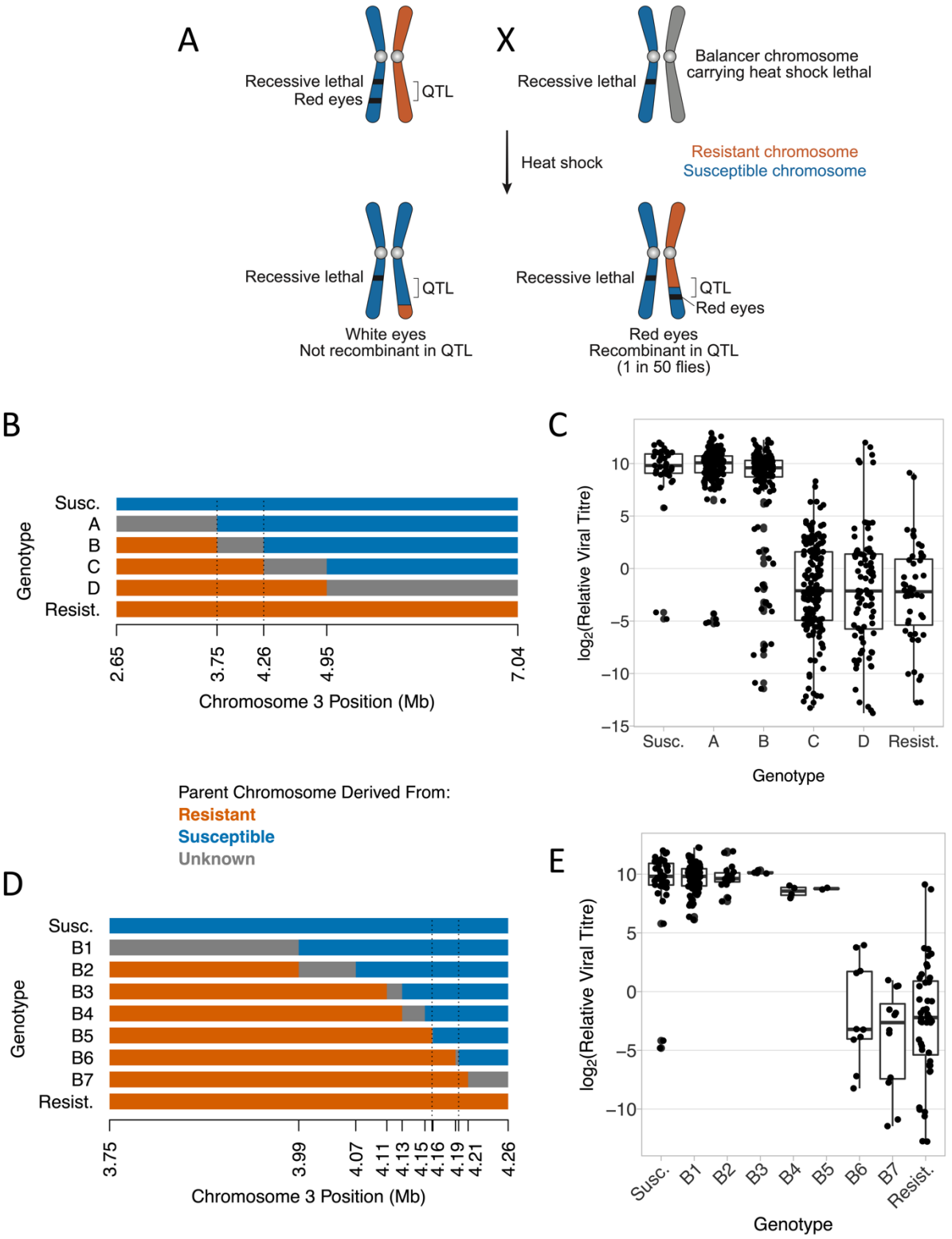

Figure 4. High resolution genetic mapping of DAV resistance. (A) The genetic cross used to select recombinants within the QTL controlling susceptibility to DAV. (B) The recombinant genotypes generated in the first phase of mapping with the location of five molecular markers used to genotype the recombinants. The dashed lines mark the region controlling susceptibility inferred from panel C. (C) The viral titre of the genotypes in (B). Each point is an independent recombinant line (mean, 16.6 flies/line). (D) Recombinants between the dotted lines in (B) with the location of 12 markers. Dashed lines mark the region controlling

susceptibility inferred panel E. (E) The viral titre of the genotypes shown in D. The points are replicates of the infection assay (5-20 flies), and some lines are represented by multiple replicates. Flies were infected by inoculation, titre was estimated 3 dpi relative to *EF1α100E*. Titre measurements in (E) are a subset of (C).

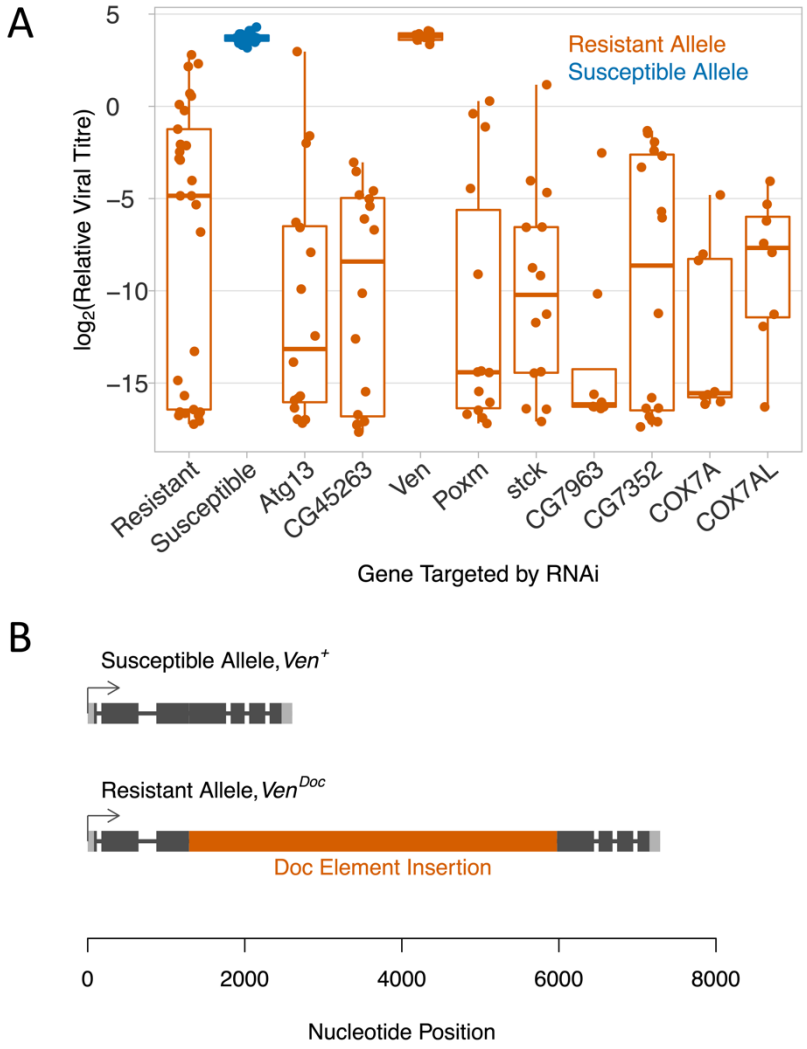

Figure 5. *Veneno* is associated with susceptibility to DAV. (A) The effect on DAV titre of knocking down the expression of nine genes in the region associated with DAV resistance. Viral titre was estimated 3 dpi using qPCR relative to *EF1α100E*. (B) The structure of *Veneno* in resistant and susceptible lines.

A

Allele	gRNA Target	% Indel	% Frameshift
Ven^{+}	Downstream Doc	99%	53.38%
Ven^{Doc}	Downstream Doc	98.2%	59.72%
Ven^{+}	Upstream Doc	82.5%	56.38%
Ven^{Doc}	Upstream Doc	99.7%	67.24%

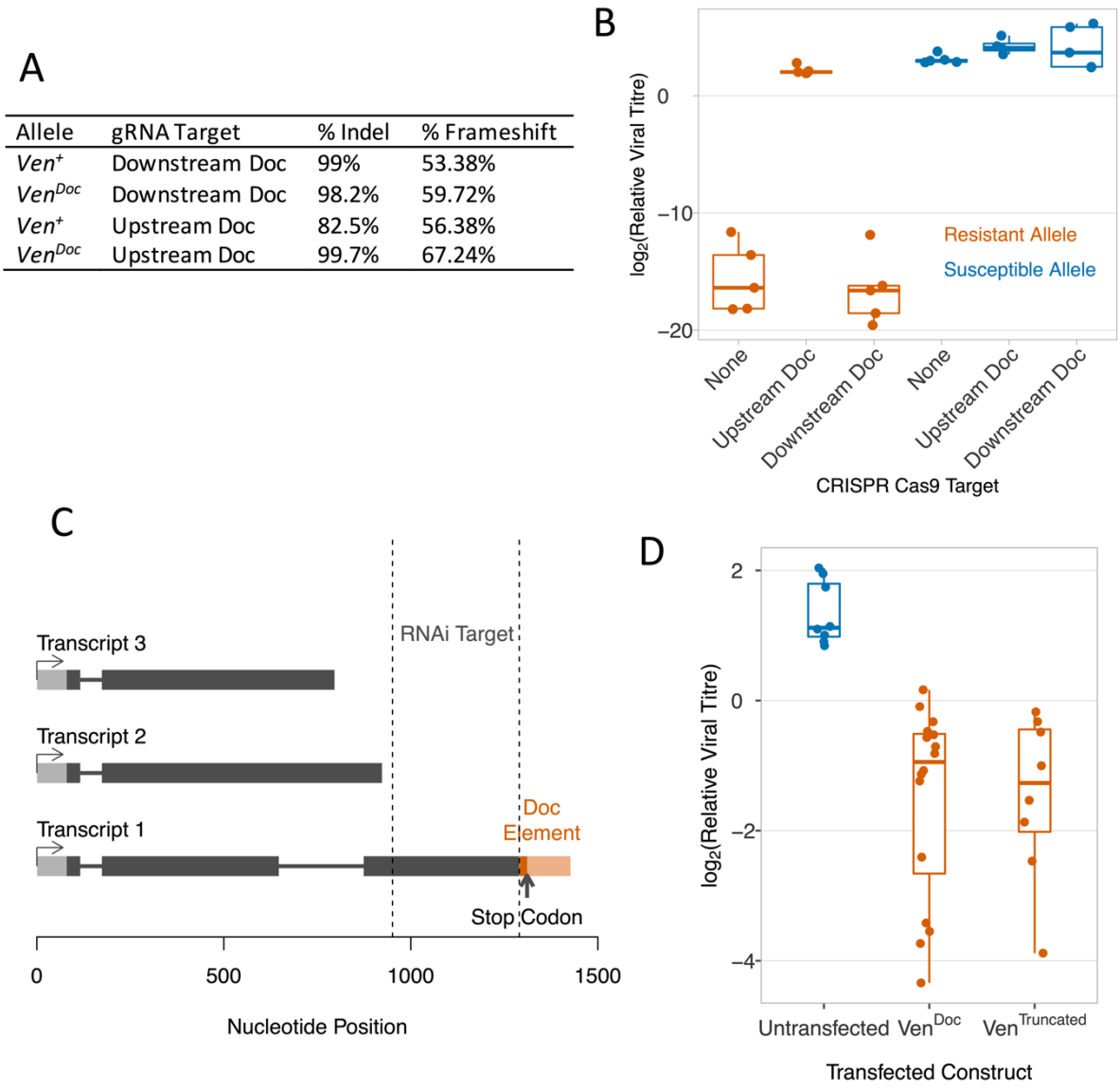

Figure 6. The *Doc* element insertion in *Veneno* creates a novel transcription factor. (A and B) Exons of *Veneno* were mutated by expressing gRNAs and Cas9 in somatic cells, targeting sequences upstream and downstream of the *Doc* insertion. (A) The gRNA targets were Illumina sequenced to estimate the proportion mutated chromosomes. (B) DAV titre estimated 3 dpi relative to *RpL32*. (C) Oxford Nanopore Technologies transcriptome sequencing identified three polyadenylated transcripts of *Ven*^{Doc}. (D) Clonal stably transfected cell lines expressing V5-tagged Transcript 1 of *Ven*^{Doc}, or the equivalent construct where sequence from the *Doc* element has been deleted (*Ven*^{Truncated}). Viral titres relative to *RpL32* were estimated 3 dpi.

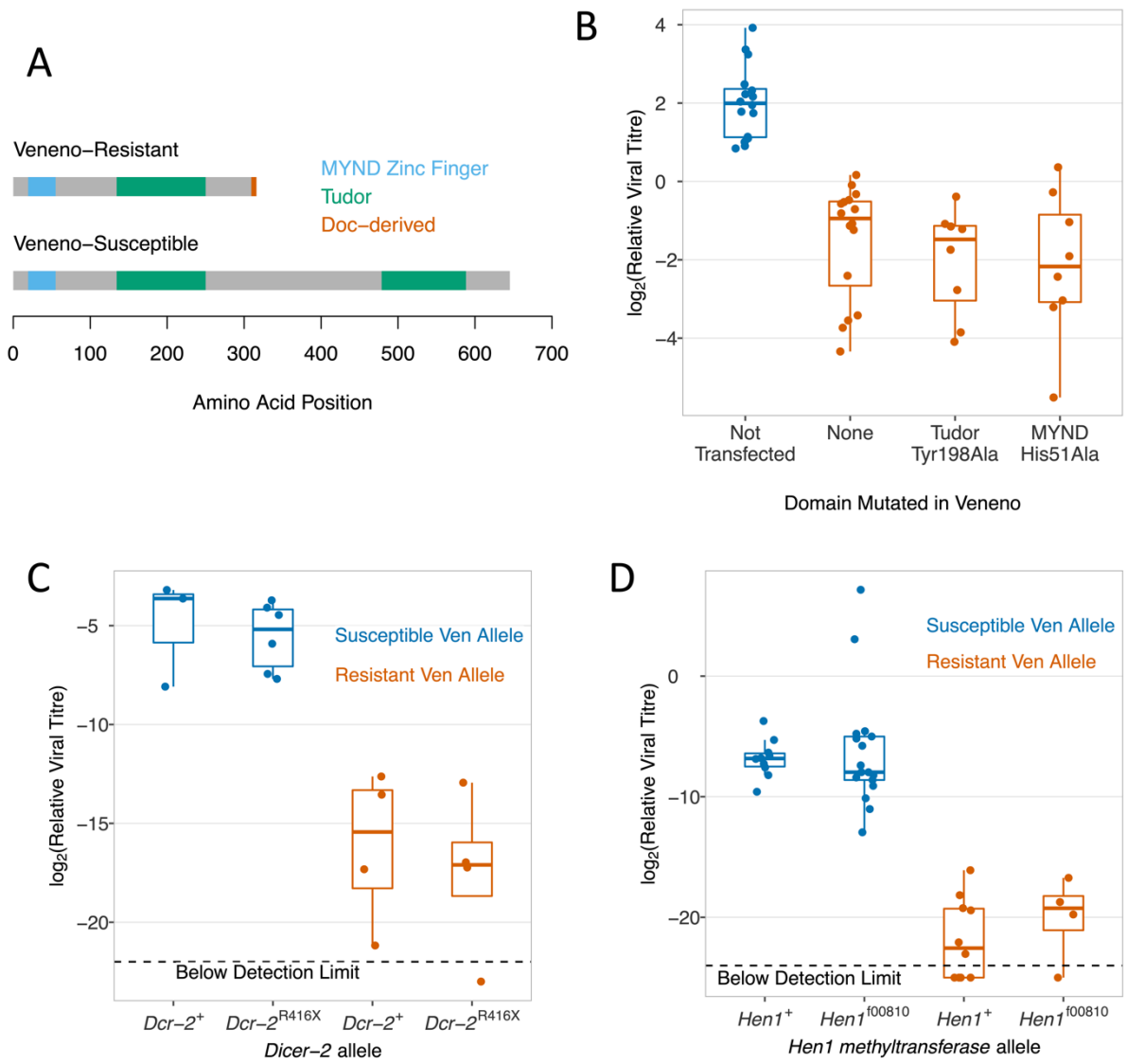

Figure 7. DAV resistance is independent of the RNAi pathway, and Tudor and MYND domains in Veneno . (A) Protein domains encoded by *Ven*$^{+}$ and *Ven*^{doc} Transcript 1. (B) Clonal stably-transfected cell lines expressing V5-tagged Transcript 1 of *Ven*^{Doc} with the mutations His51Ala and Tyr198Ala that disrupt the MYND and Tudor domains respectively. Two susceptible controls (blue) are combined (untransfected cells and cells transfected with GFP; same data as in Figure 6D). (C-D) In flies, *Dcr-2* and *Hen1* mutants were combined with *Ven*$^{+}$ and *Ven*^{doc}. Viral titre was estimated 3 dpi using qPCR relative *RpL32*.

PNAS

Supplementary Information for

A novel transposable element-mediated mechanism causes antiviral resistance in *Drosophila* through truncating the Veneno protein

Osama Brosh, Daniel K. Fabian, Rodrigo Cogni, Ignacio Tolosana, Jonathan P Day, Francesca Olivieri, Manon Merckx, Nazli Akilli, Piotr Szkuta and Francis M. Jiggins

Francis M. Jiggins
Email: fmj1001@cam.ac.uk

This PDF file includes:

Figures S1 to S5
Table S1
Supplementary Methods
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

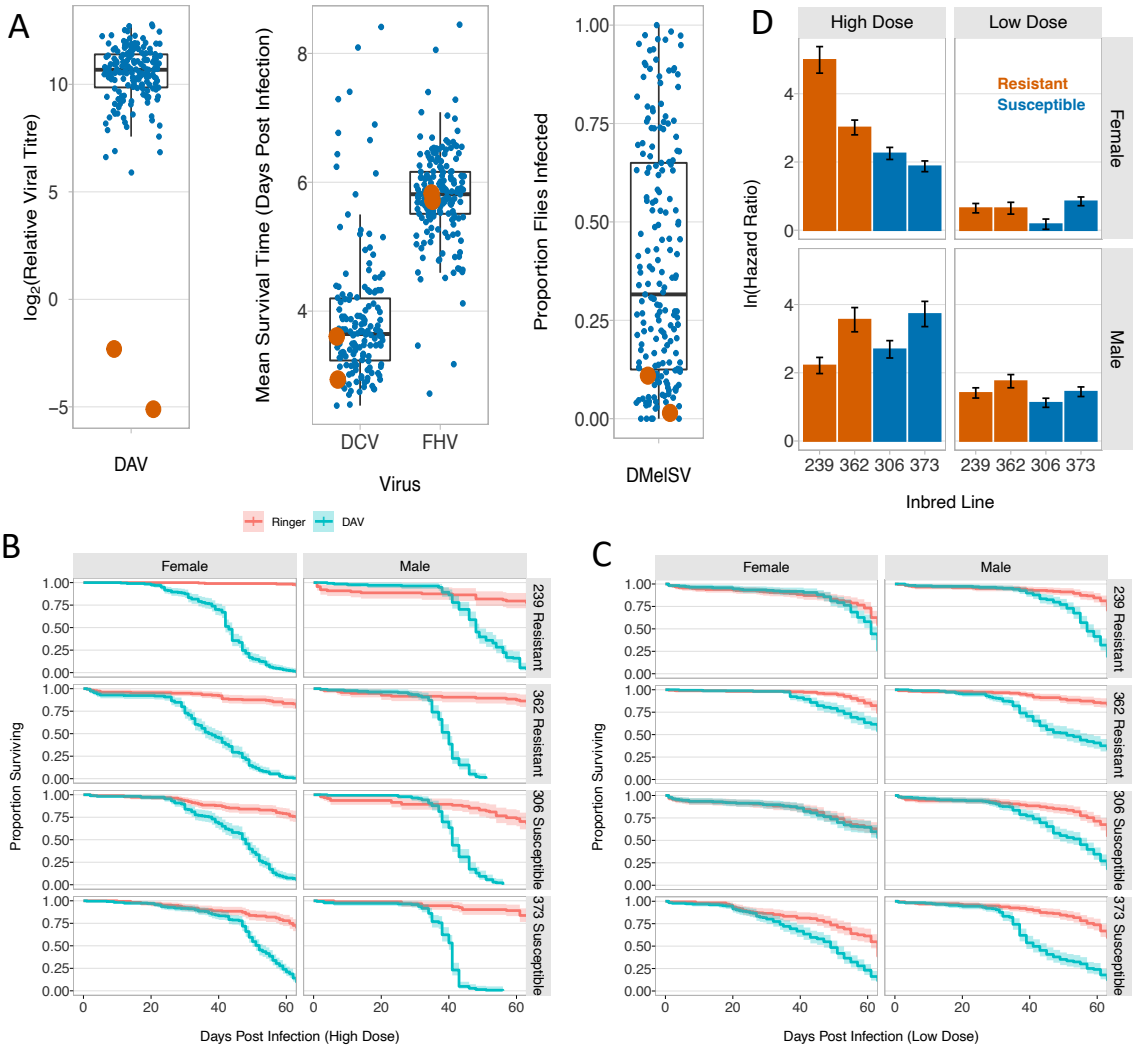

Figure S1. The fitness of resistant and susceptible flies after DAV infection. (A) Susceptibility of inbred DGRP lines to different viruses. Each point is the mean phenotype of a line, and DAV-resistant lines 239 and 362 are shown in orange. (B-C) The survival of lines after inoculation with either a (B) high or (C) low dose of DAV. Control flies were inoculated with Ringer's solution. The shaded area is the 95% confidence interval. (D) The hazard ratio comparing mortality rates in control versus infected flies. Error bars are standard errors of the estimates. The four *Drosophila* lines are the F1 progeny of a cross between a line from the DGRP panel and an isogenic w^{1118} line. This cross was performed to avoid any negative effects of inbreeding.

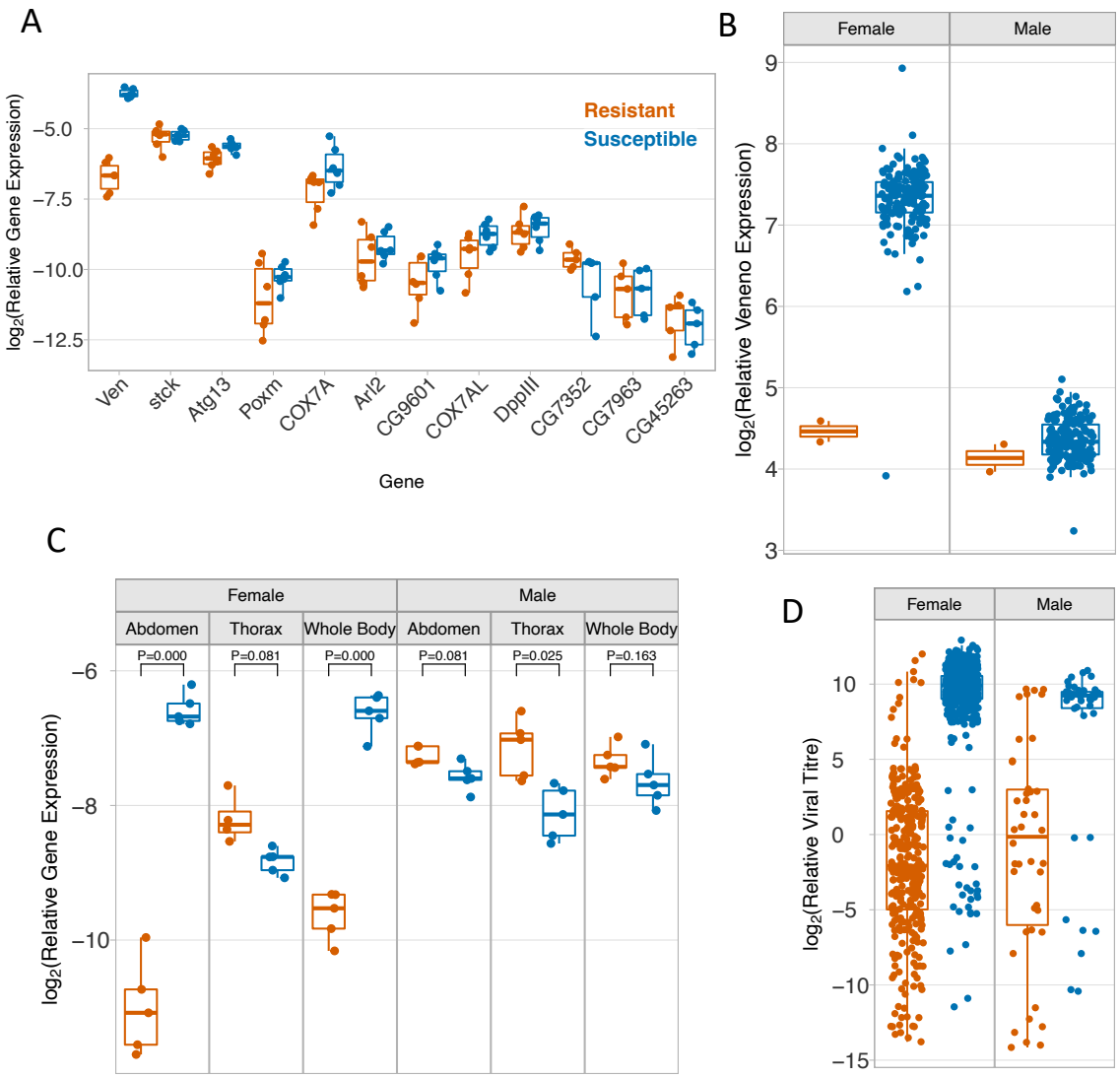

Figure S2. The expression of *Veneno* and other genes in the region associated with susceptibility to DAV. (A) The expression of the 12 genes in the region containing the resistance gene. Female flies were infected by intrathoracic inoculation, and each point is a single replicate 3 dpi at 25C of an independent homozygous recombinant line from the high resolution genetic mapping cross. Gene expression for each line was measured from a single replicate from 5-20 flies, 3 days post infection with DAV by quantitative PCR relative to *RpL32*. (B) the expression of Veneno in the DGRP lines from published microarray data on whole adult flies. (C) The expression of *Veneno* in the head, thorax and whole body of male and female flies 3 dpi. The *Veneno* primers amplify a region of the gene upstream of the *Doc* element insertion. The *P* values are from Welch's *t* test corrected for multiple comparisons by the Holm-Bonferroni method. (D) DAV titres in male and female flies. The data for females is replotted from Figure 3C in the main text. Data for males was generated at the same time from a subset of the lines. Details are given in the legend of Figure 3.

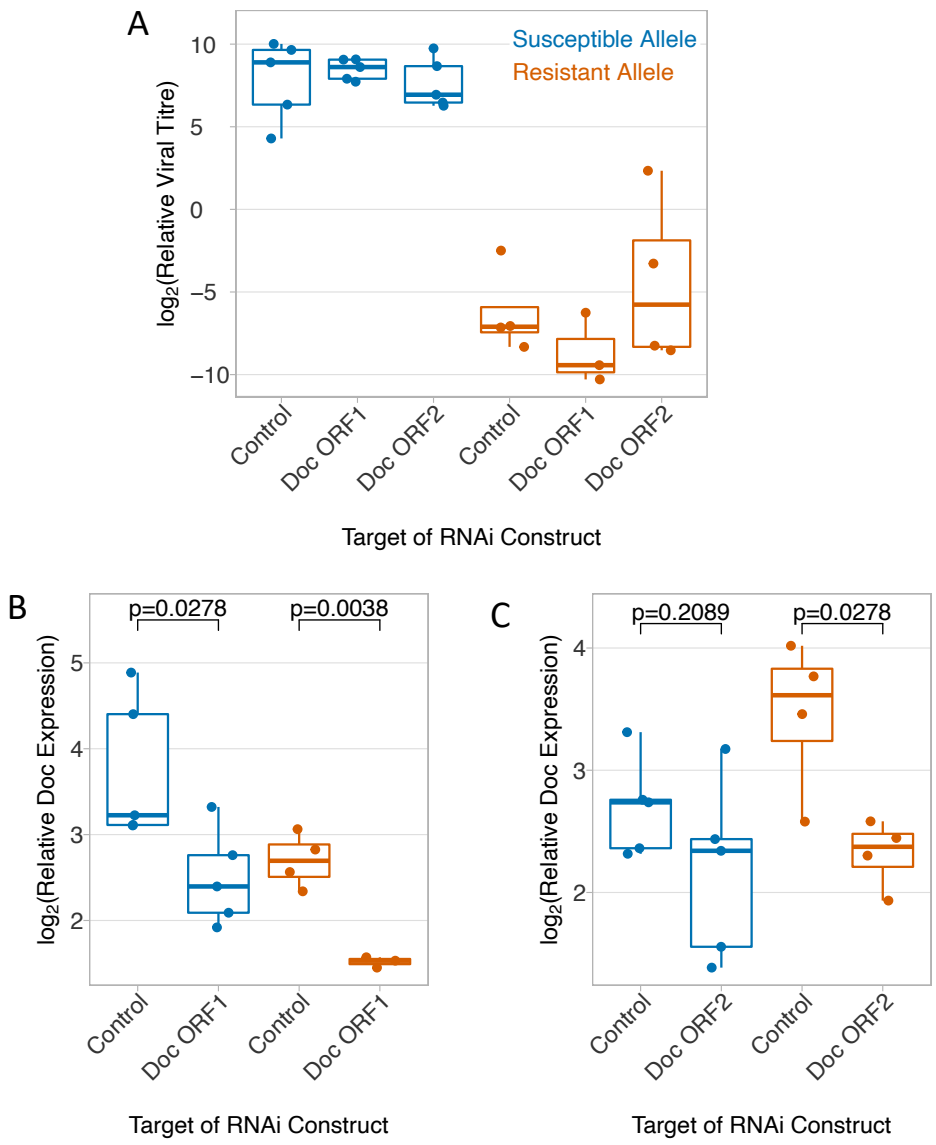

Figure S3. The effect of knocking down *Doc* element expression on susceptibility to DAV. (A) Viral titre and (B-C) *Doc* element expression. Three days post infection, viral titre and *Doc* expression was estimated using quantitative PCR relative to *RpL32.* The flies expressed RNAi constructs targeting the 5' or 3' open reading frames of the Doc element insertion (ORF1 and ORF2 respectively). The UAS-RNAi constructs were expressed under the control of Actin5C-Gal4 at 29C. Each dot represents 6-15 flies. The orange points had Ven^{doc}/Ven^{+} on chromosome 3, while the blue points are Ven^{+}. The control lines have an empty attB docking site where the UAS-RNAi constructs were integrated. 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 *RpL32*. Boxes show median and interquartile range. P values are from Welch's *t* test.

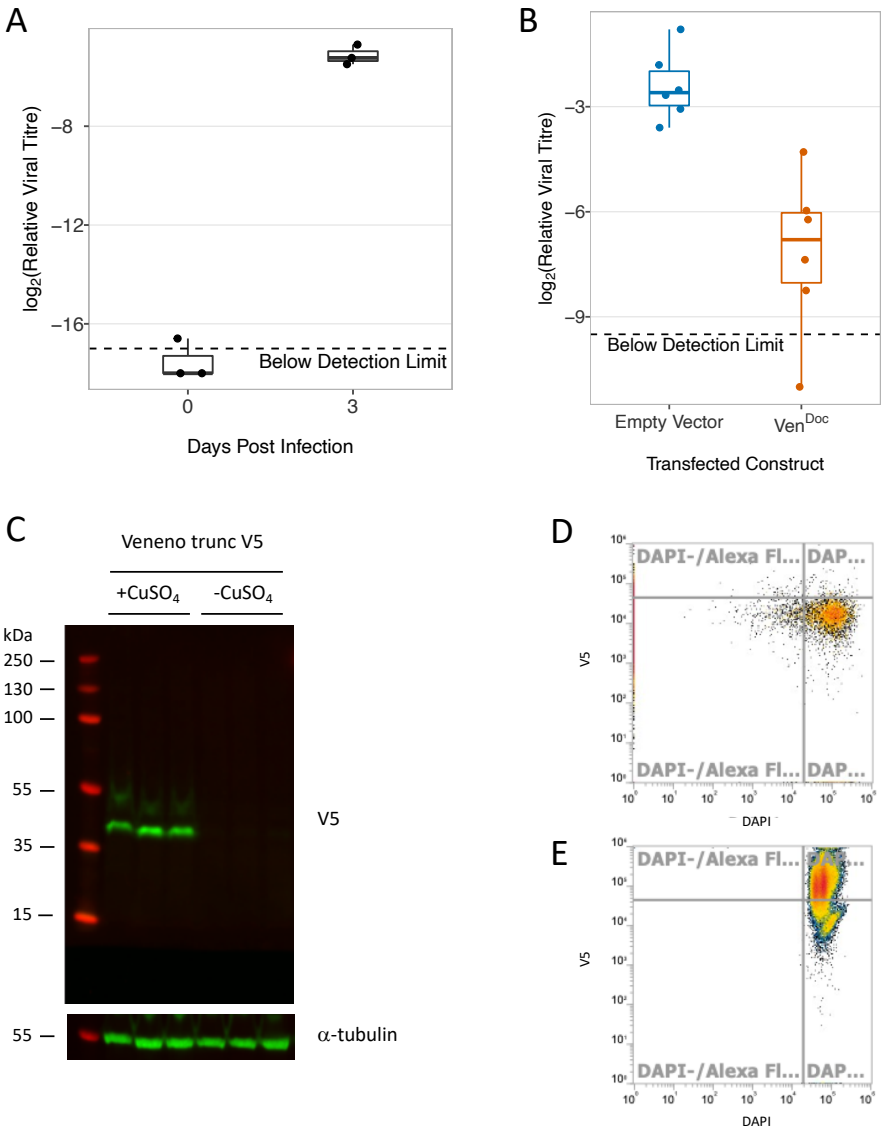

Figure S4. Expressing *Ven^{Doc}* makes *Drosophila* cells resistant to DAV. (A) DAV titres in DL2 cells at zero and three days post infection. (B) DL2 cells were transfected with a plasmid expressing Transcript 1 of *Ven^{doc}*. Puromycin was used to select stably transfected cells. Note these are not clonal cell lines and *Ven^{doc}* was not V5 tagged. The cells were infected with DAV and viral titres relative to *Rpl32* mRNA was estimated using quantitative PCR 3 dpi. The control cells are untransfected DL2 cells. Boxes show median and interquartile range. (C) Transcript 1 of *Ven^{doc}* was cloned with a V5 and His tag at the C-terminus and stably transfected into cells. In 3 wells protein expression was induced by $CuSO_4$ while 3 control wells were left untreated. A Western blot was probed anti-V5 primary antibody. To determine loading quantities of protein samples, membranes were stripped and re-probed with anti-a-tubulin primary antibody. (D, E) Flow cytometric analysis of (D) untransfected cells and (E) cells transfected with V5-tagged *Ven^{doc}* . O.2% of the cells in (D) and 84% of cells in (E) are in the upper-right quadrant and therefore express a V5-tagged protein.

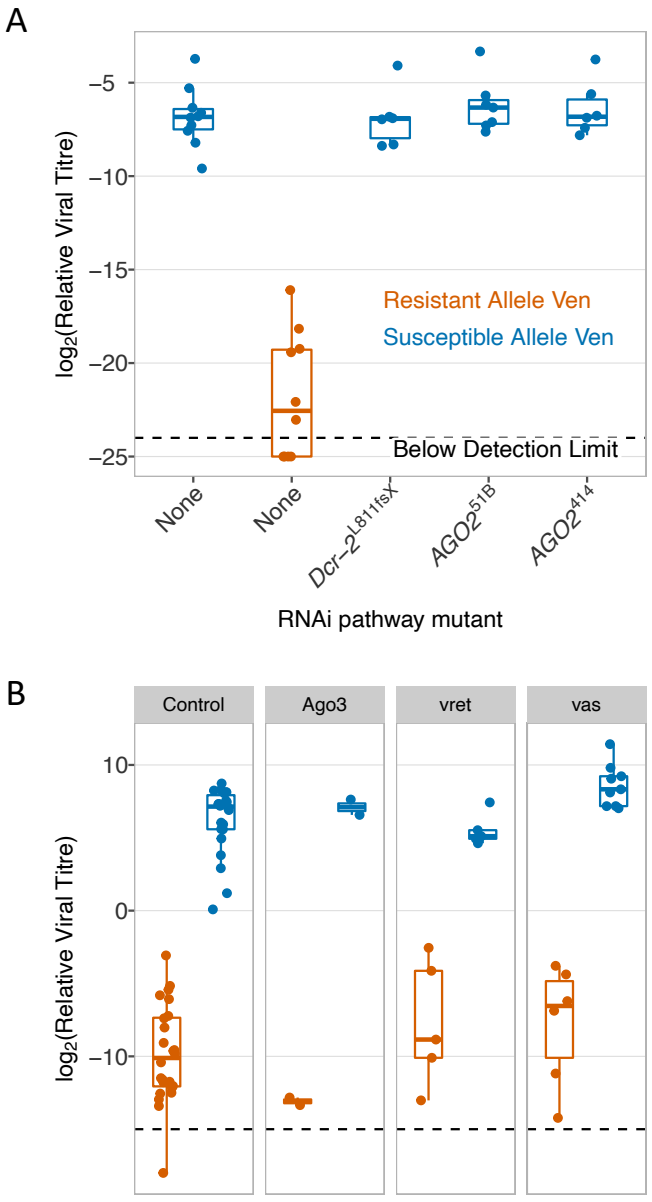

Figure S5. Effect of the siRNA and piRNA pathways on susceptibility to DAV. (A) Viral titres in siRNA pathway mutants. The lines carrying no mutations are DGRP lines 362 and 850 (resistant *Ven*^{Doc} and susceptible *Ven*$^{+}$ respectively). (B) Viral titres in lines where three genes involved in piRNA biogenesis were knocked down by RNAi. The RNAi constructs were expressed throughout somatic and germline cells under the control of *Actin5C-Gal4.* The resistant and susceptible lines carry third chromosomes from DGRP lines 362 and 850 respectively. In both experiments, viral titre was estimated 3 dpi using quantitative PCR to measure the concentration of viral RNA relative to mRNA from the *Drosophila* gene *RpL32*. Flies were infected by intrathoracic inoculation. Boxes show median and interquartile range.

Trait	*P*	Pearson rho	*N*	Source
Virgin Lifespan	0.43	0.06	173	Ivanov et al 2015
Lifespan	0.28	0.09	153	Durham et al 2014
Fecundity Week 1	0.64	-0.04	153	Durham et al 2014
Fecundity Week 3	0.86	0.01	153	Durham et al 2014
Fecundity Week 5	0.76	0.03	146	Durham et al 2014
Fecundity Week 7	0.85	0.02	111	Durham et al 2014
Total Fecundity	0.94	-0.01	153	Durham et al 2014
Unfertilized Rate	0.15	-0.16	77	Horvath et al 2018
Ovariole Number	0.95	0.00	177	Lobell et al 2017

Table S1. Correlation between DAV titers and published measurements of fitness-related.

Supplementary Methods

Virus production

Drosophila A Virus was kindly provided by Dr Karyn Johnson (The University of Queensland, Australia) (1) and replicated in a virus-free isogenic line of *Drosophila melanogaster* (w^{1118}). Ten days after infection, flies were immersed in liquid nitrogen and frozen at -80°C to rupture cells, and then homogenized in Ringer's solution (4 µl/fly) (as in Ambrose *et al.* 2009 (1)). The homogenate was centrifuged twice at 4°C, each time retaining the supernatant. This was passed through Millex PVDF 0.45µM and 0.22µM syringe filters (Millipore, Billerica, MA, USA) to remove any remaining host cells or bacteria. Aliquots were stored at -80°C. Before infections, a DAV aliquot was serially diluted to either 10^{-5} (mapping & life history traits, ~3000 TCID50/100µl) or 1:2 (life history traits, 1.5×10^8 TCID50/100µl) in 0.2 µm filtered Ringer's solution.

To create a DAV stock for feeding assays, we pricked 3 – 6-day old females of two non-resistant DGRP lines (306 and 373) with DAV as described below. We further created a virus-free control stock by pricking flies with Ringer's solution. Flies were then kept at 25°C and 70% humidity with a 12 hours light/dark cycle. After 5 days, a pool of 100 DAV-infected flies (both fly lines mixed) was homogenized with beads in 250µl of Ringer's solution. Another 250µl of Ringer's solution was then added and centrifuged at 13,000g at 4°C for 30 seconds previous to transferring the supernatant to a fresh tube. After repeating the centrifugation step, the DAV and control solution were aliquoted and stored at -80°C. We confirmed the infectivity of the DAV solution and absence of DAV in the control solution by pricking several resistant and susceptible lines and measuring their viral load (see below).

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 -80C until RNA extraction. After RNA extraction, reverse transcription, and measurement of viral titre, we calculated TCID50 using the Reed and Muench method (Reed, 1938).

DAV infections

Experimental flies were obtained by putting up to 3 males and 3 females into a vial with standard cornmeal-agar food with dry yeast sprinkled on top and allowing them to lay eggs for 1-3 days. The emerging generation was then infected with DAV through inoculation ('pricking') or feeding (see below). All flies were kept at 25˚C during the experiments (except from some RNAi experiments, see below). On the day of collection, infected flies were immersed in liquid nitrogen and kept frozen at -80ºC.

For inoculation (pricking), flies were pricked into the left pleural suture on the thorax with a 0.15 mm diameter anodized steel needle (Austerlitz Insect Pins) bent ~0.25 mm from the end (half of the dorsal width of the thorax), dipped into viral solution as in (2) under CO_2 anesthesia. Unless otherwise stated, infected flies were kept in standard cornmeal-agar food for 72 ± 3 hours and then immersed in liquid nitrogen and stored at -80˚C.

To orally infect flies, 10µl of DAV (or the control Ringer's solution) was added to a yeast solution (0.15g dried yeast / ml of deionized H_2O). 100µl of this solution were pipetted into vials containing cornmeal-agar food and allowed to dry for 24h. Flies were put into infected and control food vials and left for 3 days. After that, flies were transferred into fresh cornmeal vials without yeast and without virus/control solution every 2 days. On the 7^{th} day post infection, flies were collected for analysis of viral load.

RNA extraction

For screening the DGRP panel and QTL mapping, 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. For all other RNA extractions, we added 250µl Trizol to the pool of flies and homogenized in a TissueLyser at 30Hz for 2 minutes with ~12 x 1mm zirconia beads, then stored the samples at -80C. RNA was then extracted according to the manufacturer's protocol.

For transcript sequencing, we quantified RNA using a Qubit™ 2.0 Fluorometer and the Qubit RNA HS Assay Kit (ThermoFisher) following manufacturer's protocols. We made a 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.

When required, to reverse-transcribe RNA into complementary DNA (cDNA), we used GoScript™ Reverse Transcriptase (Promega) and random hexamer primers (RanHex). In the first stage, we incubated 0.6µL of RanHex, 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.

Quantitative PCR

The copy-number of viral RNA or mRNA was estimated in two ways. The first approach was to measure 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). All primers are given in Table S2.

Oligo Name	Sequence
DAV-primer-F	AAGGCATACTTGATGGTA
DAV-primer-R	GGGTTCCTTCCTTTATATG
DAV-probe	[6FAM]TAGCACAACAGGAAGTCCAGTATCAAT[BHQ1]
EF1-primer-F	ACGTCTACAAGATCGGAG
EF1-primer-R	CAGACTTTACTTCGGTGAC
EF1-probe	[HEX]CATCGGAACCGTACCAGTAGGT[BHQ1]
Act5C-F	GACGAAGAAGTTGCTGCTCTGGTTG
Act5C-R	TGAGGATACCACGCTTGCTCTGC

Table S2 Primers for quantitative PCR.

For the second approach to measure viral titre, we reverse transcribed RNA from Trizol extractions into cDNA. We diluted the 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 SensiFAST reagents, 2µL nfH_2O, (2.5mM), 2µL of 1:10 diluted cDNA, and 1µL of primers for either DAV in the virus reaction, and either *Rpl32* or *Act5C* in the housekeeping gene reaction. 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. All primers are given in https://doi.org/10.17863/CAM.84829.

We then calculated Ct values (that is the number of cycles at which the reaction fluorescence crosses the fluorescence threshold) using a threshold ΔRn value of 1 (ΔRn is the difference between normalized experimental reporter value and normalized baseline reporter value (ROX), where the reporter value is the fluorescent signal generated by SYBR Green). We then found ΔCt by subtracting the Ct value of our housekeeping gene from the Ct value of our virus. Viral titre is proportional to $2^{-\Delta Ct}$ so we used $-\Delta$Ct as an estimate of $\log_2$(relative viral titre). For each sample, two RT-qPCR reactions were carried out and the mean of these two technical replicates was used as the relative viral titer in the statistical analysis.

We measured transcript abundance of 12 candidate genes in a total of 13 resistant and susceptible female recombinant lines generated in the 'fine mapping' experiment and both parental lines. For this, extracted RNA from the recombinant screen was re-used. The recombinant lines used here were measured in a single replicate, so that our unit of replication was "resistant" and "susceptible" genotypes. Gene expression primers were designed by obtaining FASTA files of the coding sequences (CDS) of all known transcripts of each gene from GenBank (NCBI). Sequences from multiple transcripts of the same gene were aligned using T-coffee (3) and primers designed in regions common to all known transcripts using Primer-BLAST.

Primer-efficiency was determined by serially diluting a cDNA mix created by mixing 5µl of 6 resistant and 6 susceptible samples 8 times starting at a 1:5 dilution (i.e. 1:5 to 1:390625). For each primer, we quantified the target gene in two independent serial dilutions, and obtained the efficiency through regressing the CTs along the dilutions. All primers used had efficiencies close to 100%.

Characterising genetic variation in resistance to DAV

To assay the viral load in the panel of 182 lines from the Drosophila Genetics Reference Panel (4), we infected all the lines with DAV. Prior to this, lines infected with *Wolbachia* were treated with tetracycline as in (2). For each line we infected up to five replicates of 15 mated, 3-6 days old, females by pricking. Infected flies were collected three days post infection. For the timeseries experiment, we tested up to 6 replicates/timepoint/line. This experiment used the two resistant (239 and 362) and two susceptible DGRP lines (306 and 373). Flies from separate vials were collected at 1, 3, 7, 11 and 28 days post infection. Viral load was measured as explained above.

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. Flies were infected with DAV and virus titer was measured as described above.

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.

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 determine that it has the susceptible phenotype. 84 recombinant lines for DGRP-239 and 72 lines for DGRP-362 were produced. DAV resistance was assayed in a single female for each line as described above. We genotyped 10 SNPs distributed along the third chromosome by High Resolution Melting Curve analysis (HRM, see “Genotyping” below).

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. 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 in such a small region are 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 QTL region (Figure 4B). Generating this number of recombinants without a phenotypic marker of recombination would have required us to screen over ~30,000 fly lines.

These recombinant lines were then genotyped and adult females assayed for DAV load upon infection through pricking, including the parental lines as controls. 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. All SNP position and primers used for mapping are shown in https://doi.org/10.17863/CAM.84829. Two hundred and eighteen of these lines were also infected orally. This feeding experiment was done using pools of 1-15 flies (average = 14.1).

DNA Extractions and Genotyping

For the QTL screen, DNA was extracted with ZR-96 Quick-gDNA kit (Zymo Research). For Sanger 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 all other experiments we have extracted DNA from ~1-10 flies in 150µL of 5% w/v Chelex suspension prepared with Chelex 100 (Sigma) and 1mm Zirconia beads in a PCR tube, then homogenized using a TissueLyser for 4 minutes at 30Hz. We added 1.5µL proteinase K and incubated overnight at 56C then centrifuged, removed 100µL of supernatant, and inactivated proteinase K by incubating at 94C for 15 minutes.

To define the QTL and fine-scale mapping experiments, we selected polymorphisms using the VCF file from DGRP freeze 2 (http://dgrp2.gnets.ncsu.edu/data.html). For the QTL analysis, we selected SNPs 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. As we did not know the genomic sequence of the susceptible parent in the fine-scale mapping analysis, 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.

We used three different genotyping methods throughout this study. First, high-resolution melt analysis (HRM) was performed with MeltDoctor HRM Master Mix (Applied Biosystems) according to manufacturer protocol. We avoided G/C and A/T SNPs since they are hard to separate using HRM. We further designed HRM primers targeting a single variant to avoid possible interference from other SNPs (SNPs that have no other SNPs 100bp on either side). For the QTL analysis, we picked 10 SNPs distributed along the third chromosome. All HRM reactions were run with multiple homozygous and heterozygous controls.

Second, for the fine-scale recombination mapping, we identified restriction site polymorphisms using RestrictionMapper (http://www.restrictionmapper.org). 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). We then performed

PCR in 15µl reaction volumes using Taq Polymerase (NEB, used throughout) 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 (primers: https://doi.org/10.17863/CAM.84829). 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.

Third, we also used Sanger sequencing for genotyping in the fine-scale recombination mapping and also for screening CRISPR-Cas9 mutants (see below). As double-recombinants between two nearby markers are extremely rare, we only Sanger-sequenced lines where the flanking markers came from different parents.

PCR

All PCR primers were designed using Primer-BLAST (5) and are shown in https://doi.org/10.17863/CAM.84829. PCR products (including restriction enzyme treated products) were visualized using a 1-2% w/v TAE agarose gel electrophoresis with EtBr after running for ~30min at 120V. The general method we used for amplifying DNA was TouchDown PCR. For a 20µL reaction we used 15.4µL nuclease free water (nfH2O), 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: 95C for 2 min, 10 cycles of 95C for 30 sec, 62C minus 1C/cycle for 15 sec, and 72C for 1min/kB, then 25 cycles of 95C for 30 sec, 52C for 15 sec and 72C for 1min/kB, and finally 68C for 4 min. Annealing temperatures were adjusted when necessary.

Sanger Sequencing

We PCR-amplified the target region and 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 37C for one hour, then at 72C for 15 minutes. 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 nfH2O, 1 µL of 3.2µM primer, 1µL of BigDye Ready Reaction Mix, and 2µL of cleaned PCR product or plasmid DNA. We ran the reaction on a thermal cycler for 96C for 1 min, then 25 cycles of 96C for 10 sec, 50C for 5 sec and 60C for 4 min. These were sequenced by Source Bioscience (Cambridge) on an AppliedBiosystems Sequencer. Chromatograms of sequences were analyzed using Geneious (version 4.8) and Benchling (www.benchling.com).

Locating, sequencing and genotyping the Doc Element Insertion

We used TouchDown PCR (see above) to amplify the DNA of DGRP lines 306, 373, 362, and 239 using primer pairs expected to amplify *Veneno* (3JK_1a, 3JK_1b, 3JK_2a, and 3JK_2b, https://doi.org/10.17863/CAM.84829) and ran the products through an agarose gel. We cleaned up the products of the PCR reactions and Sanger sequenced them using the PCR primers. Sequences that contained fragments of *Doc* element were used to design further primer pairs (CGthenDoc 1 to 7 and DocthenCG 1 to 7; https://doi.org/10.17863/CAM.84829) 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.

To amplify the *Doc* element itself, we used Platinum Pfx DNA Polymerase (ThermoFisher) and primers spanning the entire Doc insertion (CGspanDoc 1 to 7; https://doi.org/10.17863/CAM.84829). 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, 1µL of template DNA (DGRP line 362), 0.2µL of Platinum Pfx DNA Polymerase, and 15.2 µL of nfH_2O. We used a thermal cycle of 94C for 2 min, then 35 cycles of 94C for 15 sec, 55C for 30 sec and 68C for 6 min.

We ran the products in an agarose gel to find the primer pairs which worked best, and we cleaned the product of primer pairs CGspanDoc4 and CGspanDoc5; https://doi.org/10.17863/CAM.84829) using ExoSap Cleanup (see above), and sequenced each product. We assembled the sequences to the reference sequence of the *Doc* element (6) as well as to the reference sequence of *Veneno* which we modified to include the sequence of the *Doc* element at the insertion site.

We genotyped DGRP lines for presence/absence of the *Doc* element insertion using TouchDown Taq PCR amplification with 3 primers: DocthenCGR2 (CCGTGGTACCTTCGAAACGA), DocThenCGF2 (TGTCCGGAGTTCTGTTGCTT), and CGThenDocF5 (GCAGGAATTGCAATGGGTTGA; Figure S6). In the presence of the *Doc* insertion, DocthenCGR2 and DocThenCGF2 amplify a 730bp region which includes sequences from the *Doc* element and the upstream *Veneno* sequence. In the absence of the *Doc* insertion, the only

product amplified is the one formed by DocthenCGR2 and CGThenDocF5 which was 390bp. We ran the amplified product on an agarose gel to determine the length and therefore the genotype.

Figure S6. Location of three PCR primers used to test for the presence or absence of the *Doc* insertion in *Veneno.*

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-down of candidate genes in a resistant genetic background. To do this we used lines that express RNAi constructs from the VDRC panel (7). 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.

G 1:	**Males**		**Females**
	w*; +; R(DGRP-362)		
		X	w*; if/CyO; MKRS/TM6B
	w*; UAS(w+); +		
G 2:	**Males**		**Females**
	w*; +/CYO; MKRS/R	X	w*; UAS/CyO; +/TM6B
G 3:	**Males**		**Females**
	w*; UAS/CyO; R/TM6B	X	w*; UAS/CyO; R/TM6B
G4	w*; UAS; R(DGRP-362)		

This experiment revealed high viral titres of resistant flies knocked-down for *Veneno* (CG9684). As we detected differences in the *EF1* housekeeping gene expression in this assay, we confirmed this result using a SYBR-green based qPCR assay (see below at gene expression) with *Act5C* as housekeeping gene. To rule out any potential technical issues, we screened several flies of the UAS-CG9684_RNAi; 362 stock for the presence of the restriction marker G at position 3R:4260142 (see above), which is diagnostic for the resistant chromosome. We also infected this line and measured viral load without crossing it to a driver line to verify that this line is indeed resistant.

We performed several representative assays to confirm that the RNAi-knockdown was successful. First, we observed 100% lethality in the larval stage of offspring between UAS-*stck* and da-Gal4 flies. Second, all the offspring from the cross UAS-*stck* with tub-Gal80^{ts}; da-Gal4 died in the pupal stage when raised at 29ºC, while – as expected – most survived at 18ºC. This indicates that both of our driver lines successfully express Gal4, and that Gal80^{ts} expectedly inhibits Gal4 at temperatures lower than 29ºC. Third, the expression of a subset of the genes was measured by qPCR to confirm knockdown.

The final assay used *daughterless-Gal4* (*Poxm, CG7963, COX7A, ATG13, CG45263, CG7352, Ven*) and *daughterless-Gal4; tubulin-Gal80^{ts}* (*Cox7AL, CG7352, Atg13, CG45263*). In Figure 5, three resistant controls are combined as they yield similar results (*w*;+;*DGRP-362, *w**;*daughterless-Gal4;*DGRP-362*,* and w*;*daughterless-Gal4/+; tubulin-Gal80^{ts}* /DGRP-362). Two

susceptible controls are combined as they yield similar results (*daughterless-Gal4* and *daughterless-Gal4; tubulin-Gal80^{ts}*).

Knocking down *Doc* expression 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. (https://doi.org/10.17863/CAM.84829).

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, with chromosome 2 landing site (attp40) to obtain *w**;UAS-gene/+;+.

We then followed the following crossing scheme (Figure S7) using either this *w**;UAS-RNAi/+;+ line or an untransformed control line *w**;attP/+;+ as well as a double balancer line, DGRP lines 362 and 850, and a line carrying Actiß n5C::Gal4 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 titres using qPCR.

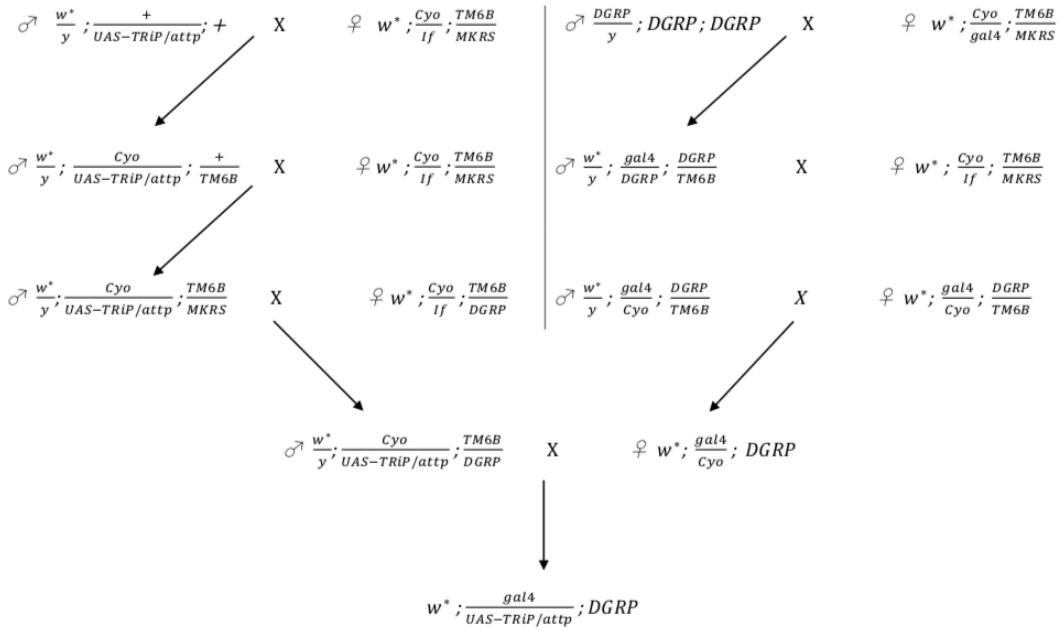

Figure S7. Crossing scheme to produce *w**;UAS-gene/Gal4;DGRP_362 and *w**;UAS-gene/Gal4;DGRP_850.

CRISPR/Cas9 mutagenesis of *Veneno*

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, which expresses a single gRNA using U6:3 which is the strongest U6 promoter in Drosophila [http://www.crisprflydesign.org/plasmids/; (8)] (this was kindly provided by Simon Bullock). gRNA expression oligos were designed and cloned with pCFD3 following the instructions on this page http://www.crisprflydesign.org/wp-content/uploads/2014/05/Cloning-with-pCFD3.pdf. Maxipreps of the plasmids were sent for transformation onto the second chromosome attp40 landing site using phiC31 expressing stocks at The University of Cambridge Department of Genetics Fly Facility, producing flies with U6:3-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, where DGRP is the third chromosome from either a resistant or susceptible DGRP line (lines 362 and 59 respectively; Figure S8).

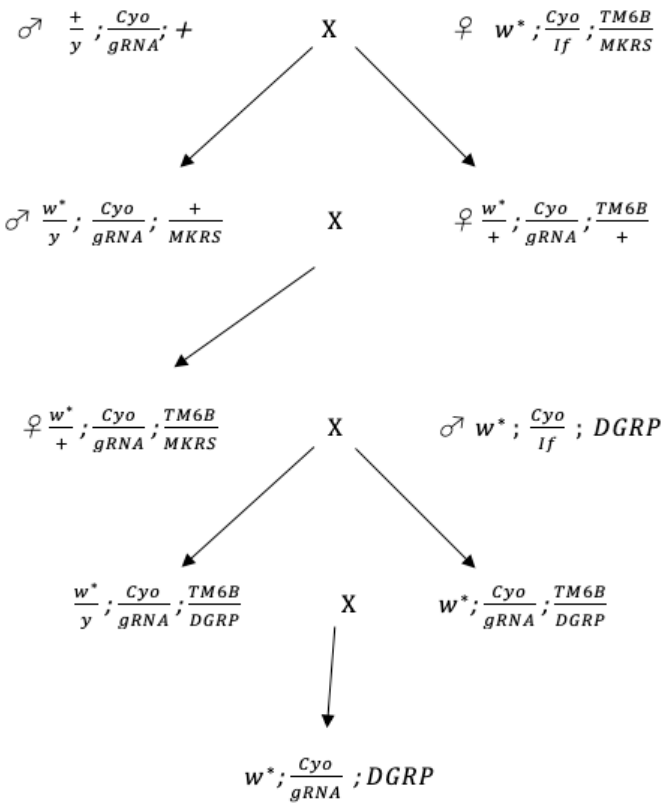

Figure S8. Crossing scheme to produce lines expressing guide RNAs to for Cas9 mutagenesis of *Veneno.* DGRP represents the third chromosome from DGRP lines 59 or 362.

We used CFD1 (Bloomington: 54590), which ubiquitously expresses *Cas9* under the control of an *actin 5C* promoter, (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 S9).

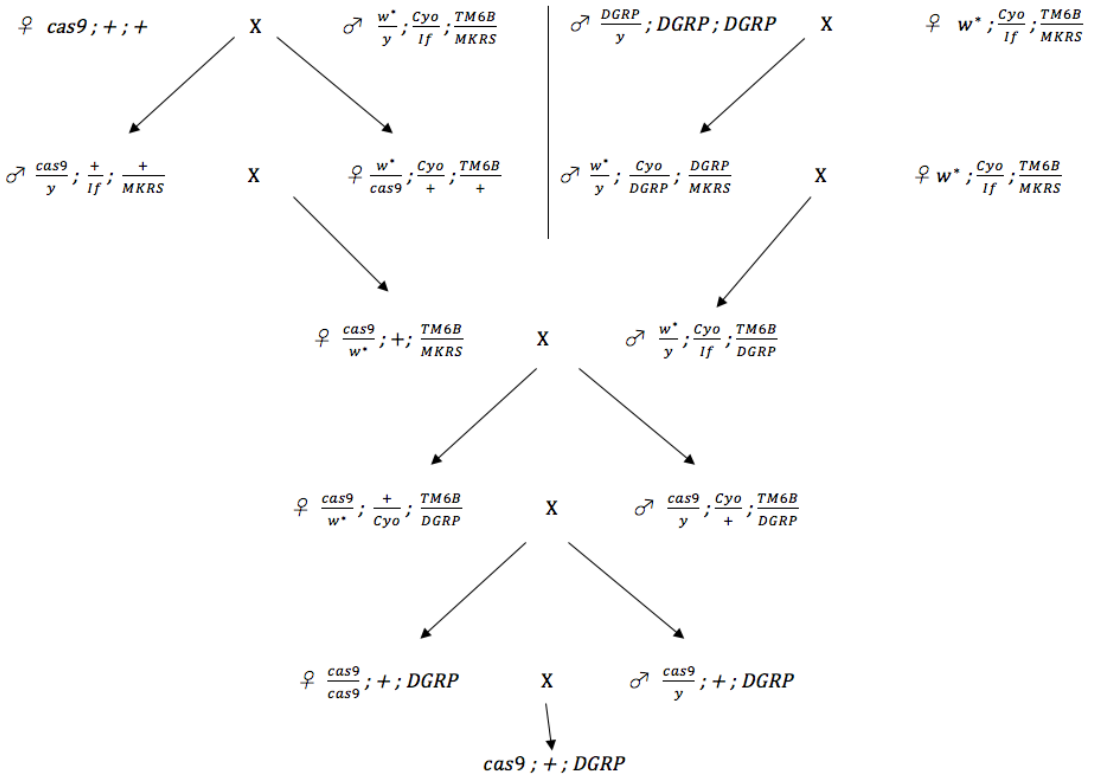

Figure S9. Crossing scheme to produce resistant and susceptible lines expressing *Cas9*. DGRP represents the third chromosome from DGRP lines 59 or 362.

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 resistant or susceptible third chromosome. We infected lines cas9/*w**;gRNA/+;DGRP, and cas9/*w**;attp/+;DGRP with DAV then extracted their RNA with TRIzol, reverse transcribed it, and measured DAV titres using qPCR.

We used Illumina sequencing to check whether the sites targeted for mutation using CRISPR/Cas9 were successfully mutated. We designed primers to amplify the guide RNA target sites, then added overhangs to allow amplification by Nextera XT index primers (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG was added to the 5' end of the forward primer and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG to the 5' end of the reverse primer; https://doi.org/10.17863/CAM.84829)

We amplified target sites by PCR using Phusion® High-Fidelity DNA Polymerase with settings in Table S3. The reaction contained 6µL Phusion HF buffer, 1.5µL dNTP (0.5mM), 20.4µL nfH2O, 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 S3. Cycling Conditions to amplify guide RNA 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) (Illumina), using the cycle settings in Table S4.

	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 S4. 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. An equimolar pool of the libraries was created and sequenced using Illumina MiSeq with 2x250bp paired end reads using the Nano v2 kit.

siRNA pathway mutants

We combined siRNA pathway mutants with resistant and susceptible third chromosomes. 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* mutation on the second chromosome, and line ;*Hen1*^{f00810}; which has a *Hen1* mutation on its second chromosome. Using the mutant lines, DGRP lines 362, 306, and 373, 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 crossing scheme in Figure S10, and then we infected those lines with DAV, extracted their RNA, reverse transcribed it, and tested it for DAV infection using qPCR.

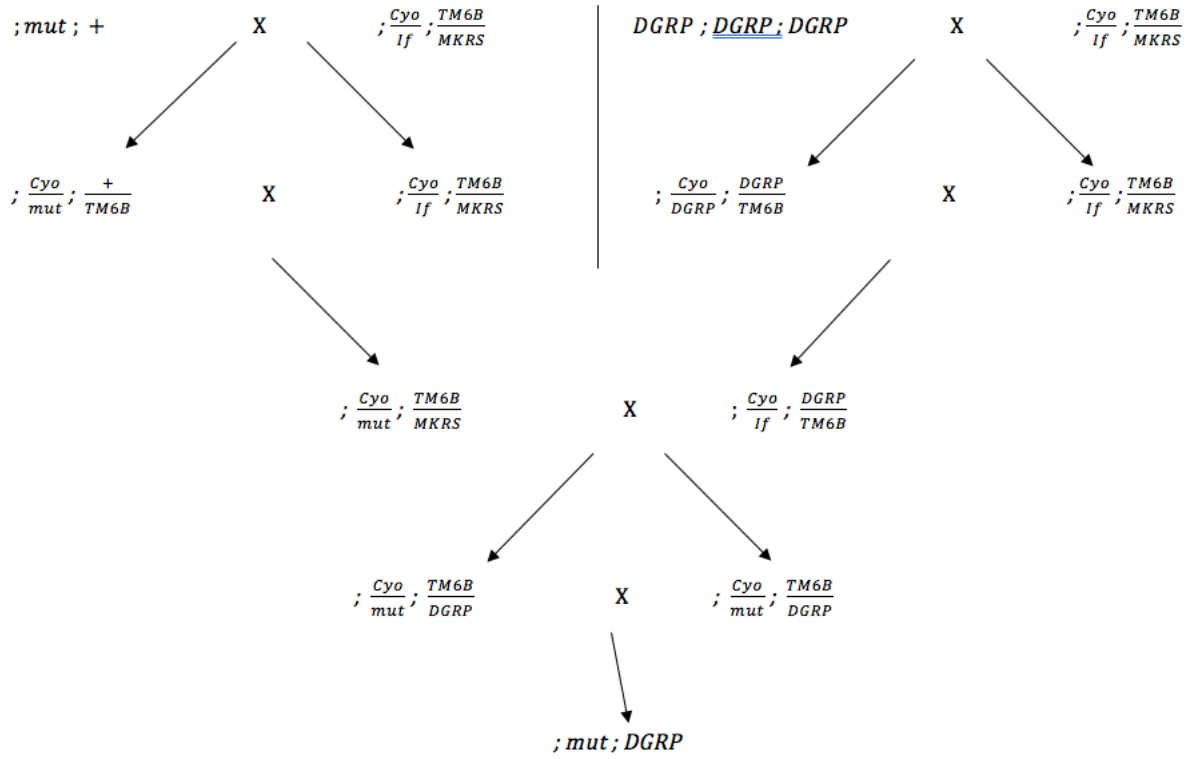

Figure S10. Crossing scheme to combine RNAi pathway mutations ('mut') with resistant and susceptible third chromosomes ('DGRP')

Knocking-down genes in the piRNA pathway

To knock down genes involved in piRNA biogenesis, we used fly stocks which carried TRiP RNAi (9) constructs targeting genes of interest. Using a fly line which expresses Gal4 under the control of an *actin5C* promoter (provided by Alex Whitworth), *w**;CyO/If;DGRP lines (DGRP is the third

chromosome from DGRP line 362 or 850) and double balancers, we created lines w*;Act5C-Gal4/UAS-TRiP;DGRP/+ or w*;Act5C-Gal4/+;DGRP/UAS-TRiP (depending on whether the TRiP-RNAi construct was on the second or third chromosome) according to the crosses in Figure S11. We infected the resulting lines with DAV then extracted their RNA with TRIzol, reverse transcribed it, and tested it for DAV infection using qPCR.

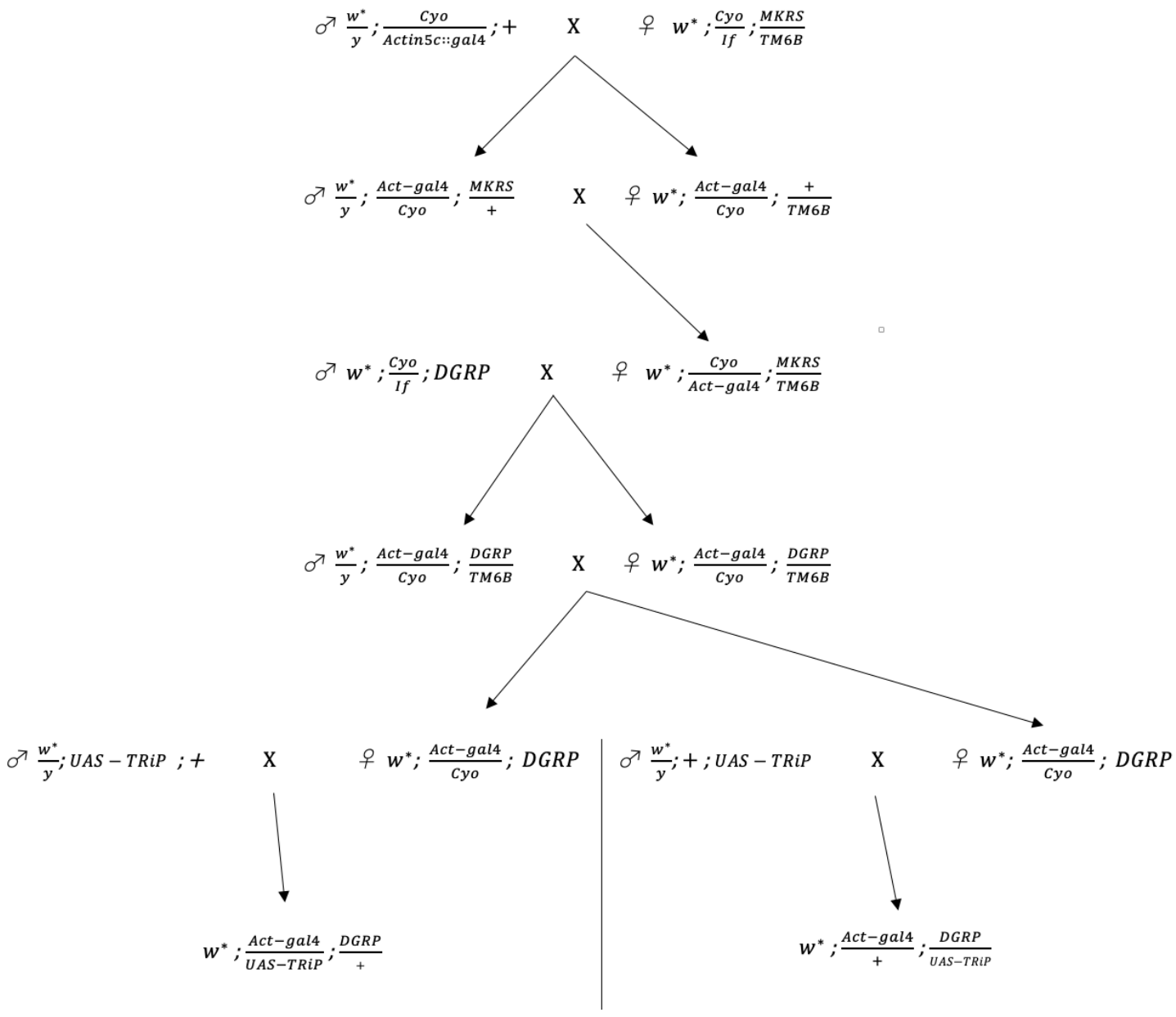

Figure S11. Crossing scheme to knock down genes involved in piRNA biogenesis. DGRP represents the third chromosome from DGRP lines 850 or 362.

Oxford Nanopore Technologies Transcriptome Sequencing

We used Oxford Nanopore Technologies PCR cDNA sequencing to sequence full length transcripts of 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). We sequenced on the MinIon (Oxford Nanopore) connected to a MinIT device (Oxford Nanopore).

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 (10) (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 utilized samtools (11) (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 into 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 with DOI: 10.5281/zenodo.5634071.

Fecundity and Lifespan after DAV infection

To assess whether the resistance allele confers a fitness benefit upon infection with DAV, we measured survival and fecundity in resistant and susceptible lines. To avoid potential effects of homozygous deleterious alleles, we first crossed males of two resistant (239 and 362) and two susceptible lines (306 and 373) to virgins of w^{1118}. We confirmed that w^{1118} is susceptible before the experiment, and we knew that the resistant allele is dominant. Crosses were setup in bottles with standard cornmeal food and dry yeast on top with 6 males and 6 virgin females, which were

transferred to fresh food after two days to keep larval competition low. All crosses with 239 resulted in infertile offspring, which therefore could only be used in the lifespan assay. Male and female offspring were collected within a 24-hour period to control for the age of the flies. At the age of 3 days post eclosion, male and female flies were pricked with DAV using a 1:2 (fecundity & lifespan) or a $1:10^{-5}$ (lifespan) dilution or a virus-free Ringer's solution.

Lifespan was then measured in each sex separately with 6 – 24 flies per vial (mean 19.1) in up to 7 replicates. Flies that died within 12 hours post infection were excluded from the analysis. The number of dead flies was recorded daily until day 31, and then every 2-3 days until day 65 post infection. Flies were transferred to vials with fresh food every 3 days. In total, we measured 2423 flies for the 1:2 and 2869 for the $1:10^{-5}$ dilution experiments. Flies that were alive at the end of the experiment were censored.

In the fecundity assay, we controlled the amount of yeast in vials by adding 100µl of a yeast solution (0.15g yeast per 1 ml of deionized H_2O) on top of cornmeal food, which was then left drying for 24 h. We then added 1 female and 2 males in each vial with up to 30 replicates. Flies were transferred to fresh medium every other day, or 20 hours before fecundity, as defined by the number of eggs, was scored. The number of eggs in each vial were counted under a standard light microscope at 4 and 11 dpi as described in (12). Flies that did not lay any eggs throughout the experiment were excluded from the analysis. Males that died during the experiment were replaced with mock- or DAV-infected males created at the beginning of the experiment. In total, we measured fecundity of 172 females.

We further measured fecundity by feeding DAV to 21 recombinant lines from the fine-mapping experiment and the two parental DGRP lines. The assay was performed similarly as above, however with the following changes. To obtain experimental flies, we put 10 males and 10 females together in a small Petri dish with cornmeal food and yeast and let them lay eggs for 24 h. We then picked 10-15 larvae using a spatula and transferred them into vials with the same food to control larval density. Next, adult flies that eclosed within a 24h period were aged to 1 day, and 1 female and 2 males transferred into vials with DAV or control-solution (see above). After 48 h, flies were transferred to vials with added yeast solution (see fecundity assay above) and this was repeated every other day thereafter. We counted the numbers of eggs at 7 and 13 dpi, and measured a total of 168 females in this assay.

To measure larval to adult survival, 20 first-instar larvae were picked and placed in a vial with no live yeast added to the surface of the food. We then pipetted virus or control solution on top of the larvae as described above. Fourteen days later we counted the number of adults.

Cell Culture

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. Cells 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 later, 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 or, in lines expressing gene with a V5 affinity tag using flow cytometry.

Creating plasmids

To make plasmid constructs with inserts for transfection into cells or injecting into flies we followed the following procedure. 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 where desired such as to place the sequence in frame or to generate a stop (https://doi.org/10.17863/CAM.84829). One or more inserts were amplified using KAPA HiFi Hot Start Ready Mix, following manufacturer protocol, and using the PCR settings in Table S5.

	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 S5 Cycling Conditions for insert amplification with KAPA HiFi Hot Start Ready Mix

We then ran 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.

The vector was digested with either NotI-HF (NEB) or EcoRI-HF (NEB) 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 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 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.

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 of correct size by running a PCR reaction targeted around the insertion site (https://doi.org/10.17863/CAM.84829) 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 (https://doi.org/10.17863/CAM.84829). Samples with the correct insert were then grown further for freezing and for plasmid extraction.

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.

Expressing *Ven^{Doc}* in cells

We created plasmids expressing *Ven^{Doc}* Transcript 1 coding sequence using NEBuilder. We inserted the sequence of *Ven^{Doc}* with both UTRs removed into pMT-puro which has been digested with NotI-HF. The *Ven^{Doc}* sequence was amplified in three parts from DGRP_362 DNA (https://doi.org/10.17863/CAM.84829) and was inserted either without an added stop codon to produce pMT-puro-*Ven^{Doc}* –V5 (Figure S12) or with a stop codon to produce pMT-puro-*Ven^{Doc}* (Figure S13). Plasmid sequences can be found at: https://benchling.com/obrosh/f_/Pack5uIT-ven_doc-plasmids/ .

Figure S12 Plasmid to express *Ven^{Doc}* Transcript 1 in cell culture *(pMT-puro-Ven^{Doc} -V5).* Expresses PuroR gene causing puromycin resistance, and *Ven^{Doc}* transcript 1 attached to the V5 tag under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder to insert Ven^{Doc} transcript 1 in two fragments which we amplified from DGRP_362 DNA using primers in https://doi.org/10.17863/CAM.84829.

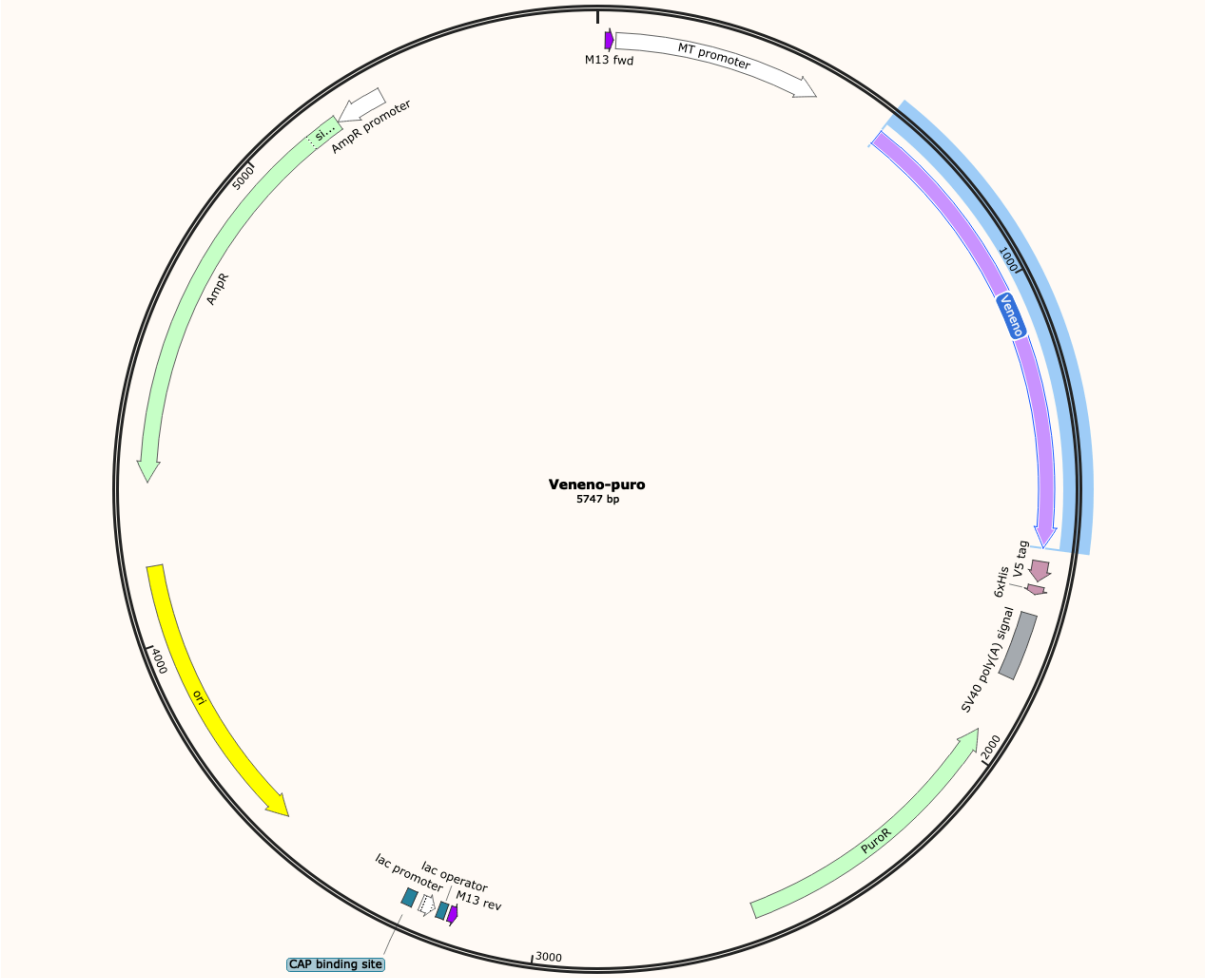

Figure S13 Plasmid to express *Ven^{Doc}* Transcript 1 in cell culture (pMT-puro-Ven^{Doc}). Expresses PuroR gene causing puromycin resistance, and *Ven^{Doc}* transcript 1 under the control of the MT promoter. We digested pMT-puro with NotI-HF, then used NEBuilder to insert Ven^{Doc} transcript 1 with an alteration resulting in a stop codon at the end of the Ven^{Doc} transcript 1 sequence before the V5 tag, in two fragments which we amplified from DGRP_362 DNA using primers in https://doi.org/10.17863/CAM.84829.

We infected with DAV DL2 cells transfected with pMT-puro-*Ven^{Doc}* and DL2 cells transfected with pMT-puro (as control) with cells at a 1.5x10^5 cells/mL density and ~2.1 viruses per well. 3 days later we extracted the RNA using TRIzol extraction, reverse transcribed it, and used qPCR to measure viral titre.

Drosophila DL2 cells were transfected with a truncated form of the veneno transcript expressed in DGRP fly lines resistant to DAV (veneno trunc). The veneno transcript was cloned to encode a V5 and His tag at the C-terminus. The expected molecular weight of the tagged, truncated veneno protein is 39kDa. Stable cell lines were selected by treating transfected cells with 10ug/ml puromycin and screening for a line that had the highest level of expression after protein induction

using flow cytometry. This line was seeded into wells of a 6-well plate and incubated for 48 hours. In 3 wells protein expression was induced for 24 hours by adding 500uM $CuSO_4$, 3 control wells were left untreated. Cells were harvested, washed in PBS, lysed in 50ul IGEPAL lysis buffer supplemented with protease inhibitor cocktail (Roche). After pelleting debris by centrifugation at 12,000g at 4°C for 5 minutes, 45 of lysate was removed to a fresh tube and 15ul of 4x SDS loading buffer was added. Samples were denatured by heating to 95°C for 5 minutes and then equal quantities were separated by SDS-PAGE using a NuPAGE 4-12% Bis-Tris Gel (Thermo Fisher NP0322) including a lane containing PageRuler plus pre-stained protein ladder (Thermofisher 26619). Protein was transferred to Amersham Protran 0.45 um Nitrocellulose blotting membrane (GE Healthcare 10600008) and blocked in TBST with 5% milk for 30 minutes. Anti-V5 monoclonal antibody (Thermofisher R960-25) was diluted 1:5000 in 5% milk in TBST and the membrane was incubated with this over-night at 4°C. After washing, the membrane was incubated with IRDye 800CW Goat anti-mouse secondary (LiCor 326 32210) (1:5000 dilution in TBST with 5% milk). To determine loading quantities of protein samples membranes were stripped and re-probed with anti-a-tubulin primary antibody (Sigma T6199) diluted 1:10000 in 5% milk, washed and incubated with IRDye 800CW Goat anti-mouse secondary. Membranes were detected using Li-Cor Odyssey XF Imager for 10 minutes detection time for the 800nm wavelength and 2 minutes for the 700nm wavelength.

Western Blots

Drosophila DL2 cells were transfected with a truncated form of the veneno transcript expressed in DGRP fly lines resistant to DAV. The veneno transcript was cloned to encode a V5 and His tag at the C-terminus. The expected molecular weight of the tagged, truncated veneno protein is 39kDa. A stably transfected cell line was seeded into wells of a 6-well plate and incubated for 48 hours. In 3 wells protein expression was induced for 24 hours by adding 500uM $CuSO_4$, 3 control wells were left untreated. Cells were harvested, washed in PBS, lysed in 50ul IGEPAL lysis buffer supplemented with protease inhibitor cocktail (Roche). After pelleting debris by centrifugation at 12,000g at 4°C for 5 minutes, 45 of lysate was removed to a fresh tube and 15ul of 4x SDS loading buffer was added. Samples were denatured by heating to 95°C for 5 minutes and then equal quantities were separated by SDS-PAGE using a NuPAGE 4-12% Bis-Tris Gel (Thermo Fisher NP0322) including a lane containing PageRuler plus pre-stained protein ladder (Thermofisher 26619). Protein was transferred to Amersham Protran 0.45 um Nitrocellulose blotting membrane (GE Healthcare 10600008) and blocked in TBST with 5% milk for 30 minutes. Anti-V5 monoclonal antibody (Thermofisher R960-25) was diluted 1:5000 in 5% milk in TBST and the membrane was incubated with this over-night at 4°C. After washing, the membrane was incubated with IRDye 800CW Goat anti-mouse secondary (LiCor 326 32210) (1:5000 dilution in TBST with 5% milk). To determine loading quantities of protein samples membranes were

stripped and re-probed with anti-a-tubulin primary antibody (Sigma T6199) diluted 1:10000 in 5% milk, washed and incubated with IRDye 800CW Goat anti-mouse secondary. Membranes were detected using Li-Cor Odyssey XF Imager for 10 minutes detection time for the 800nm wavelength and 2 minutes for the 700nm wavelength.

Site directed mutagenesis

We created plasmids expressing versions of *Ven^{Doc}* with the mutations in Table S6 by mutating the pMT-puro-*Ven^{Doc}*-V5 plasmid using the Q5 site directed mutagenesis kit (NEB) and the primers in https://doi.org/10.17863/CAM.84829, plasmid sequences can be found at: https://benchling.com/obrosh/f_/Pack5uIT-ven_doc-plasmids/ . 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 S7).

Mutation name	Mutation Description
NoDoc	deletion of the Doc element sequence (to produce Ven^{truncated}
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} (control for PremStop).

Table S6 Mutations made to *Ven^{Doc}* transcript 1 using Q5 site directed mutagenesis. A mutation similar to H51A was found to inactivate the MYND Zn finger domain (13). Mutations similar to Y198A were found to inactivate the Tudor domain (14).

	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 S7 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 (https://doi.org/10.17863/CAM.84829) 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 https://doi.org/10.17863/CAM.84829). Samples with the correct insert were then grown further for freezing and for plasmid extraction. We then prepared the plasmids and transfected the cells as described above.

We cloned the mutant cell lines (and the unmutated pMT-puro-*Ven^{Doc}* –V5 line) by following methods similar to Zitzmann et al (15): 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 confirmed the expression of the V5 tag in the selected cells via flow cytometry, and the expression of the transgene via qPCR (https://doi.org/10.17863/CAM.84829).

Infecting cells with DAV

We added $CuSO_4$ 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. We placed 90µL of our cells in a 96 well plate (Cellstar). 24-48 hours later, we diluted the cells to ~$1.5x10^5$ cells/mL or ~$1.5x10^6$ cells/mL in medium containing 5mM $CuSO_4$, and counted the cells under the microscope using Fast-Read 102 cell counting chambers (Biosigma). 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 TCID50 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 TCID50 or 4.2 viral particles/well). 72 ± 3 hours later, cells were spun down, medium was removed, and cells were dissolved with 30µL nfH2O, then frozen in 200µL TRIzol Reagent (ThermoFisher) at -80C until RNA extraction, or dissolved in 50µL nfH2O then frozen in 150µL TRIzol LS Reagent (ThermoFisher) at -80C 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 followed the following protocol. Transcription of our insert was induced by incubation with 5mM of $CuSO_4$ 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 (voltage = 280V), and the violet laser 1 channel (voltage = 260V) for DAPI as well as forward scatter channel (FSC, voltage = 120V) and the side scatter channel (SSC, voltage =

380V). 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.

Statistical Analysis

To estimate the repeatability of our resistance assay, we fitted in a linear mixed-effect model. Let $y_{i,j,k}$ be the $\log_2$ virus titer (delta Ct) in vial *k* from line *j* measured on qPCR plate *i*:

$$y_{i,j,k} = \beta + plate_i + line_j + \varepsilon_{i,j,k} \quad (1)$$

where β is the overall mean virus titer, $plate_i$ is a random variable representing the deviation from the overall mean of virus tier measured on qPCR plate *i* and $line_j$ is a random variable representing the deviation of line *j*. $\varepsilon_{i,j,k}$ is the residual error. The model parameters were estimated by REML using the *lmer* function in R. Using the parameters estimated in this model we calculated the repeatability, *R,* of our assay:

$$R = \frac{\sigma^2_{line}}{\sigma^2_{line}+\sigma^2_{plate}+\sigma^2_{\varepsilon}} \quad (2)$$

Where σ^2_{line} is the between-line variance, σ^2_{plate} is the between qPCR plate variance and σ^2_{ε} is the residual variance.

Data availability

The raw data and scripts that reproduce all the figures and statistics are available in the Cambridge Data Repository https://doi.org/10.17863/CAM.84829.

SI References

1. R. L. Ambrose, *et al.*, Drosophila A virus is an unusual RNA virus with a T=3 icosahedral core and permuted RNA-dependent RNA polymerase. *J. Gen. Virol.* **90** (2009).
2. J. Martinez, *et al.*, Addicted? Reduced host resistance in populations with defensive symbionts. *Proc. R. Soc. B Biol. Sci.* **283** (2016).
3. C. Notredame, D. G. Higgins, J. Heringa, T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302** (2000).
4. T. F. C. Mackay, *et al.*, The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
5. J. Ye, *et al.*, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13** (2012).
6. K. O'Hare, M. R. K. Alley, T. E. Cullingford, A. Driver, M. J. Sanderson, DNA sequence of the Doe retroposon in the white-one mutant of Drosophila melanogaster and of secondary insertions in the phenotypically altered derivatives white honey and white-eosin. *MGG Mol. Gen. Genet.* **225** (1991).

7. G. Dietzl, *et al.*, A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature* **448** (2007).
8. F. Port, H. M. Chen, T. Lee, S. L. Bullock, Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **111** (2014).
9. J. Zirin, *et al.*, Large-scale transgenic Drosophila resource collections for loss- And gain-of-function studies. *Genetics* **214** (2020).
10. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34** (2018).
11. H. Li, *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
12. D. K. Fabian, *et al.*, Spatially varying selection shapes life history clines among populations of Drosophila melanogaster from sub-Saharan Africa. *J. Evol. Biol.* **28** (2015).
13. Y. Liu, *et al.*, Structural Basis for Recognition of SMRT/N-CoR by the MYND Domain and Its Contribution to AML1/ETO's Activity. *Cancer Cell* **11** (2007).
14. H. Liu, *et al.*, Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev.* **24** (2010).
15. J. Zitzmann, *et al.*, Single-cell cloning enables the selection of more productive Drosophila melanogaster S2 cells for recombinant protein expression. *Biotechnol. Reports* **19** (2018).