

# Host-pathogen coevolution increases genetic variation in susceptibility to infection

Short title: Selection by pathogens increases genetic variation in susceptibility

Elizabeth ML Duxbury<sup>1\*</sup>, Jonathan P Day<sup>1\*</sup>, Davide Maria Vespasiani<sup>1</sup>, Yannik Thüringer<sup>1</sup>,  
Ignacio Tolosana<sup>1</sup>, Sophia CL Smith<sup>1</sup>, Lucia Tagliaferri<sup>1</sup>, Altug Kamacioglu<sup>1</sup>, Imogen Lindsley<sup>1</sup>,  
Luca Love<sup>1</sup>, Robert L Unckless<sup>2</sup>, Francis M Jiggins<sup>1+</sup>, Ben Longdon<sup>1,3+</sup>

\* joint first authors

+ equal contributions, joint senior and corresponding authors, [fmj1001@cam.ac.uk](mailto:fmj1001@cam.ac.uk) and  
[b.longdon2@exeter.ac.uk](mailto:b.longdon2@exeter.ac.uk)

<sup>1</sup> Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK

<sup>2</sup> Department of Molecular Biosciences, University of Kansas

<sup>3</sup> Centre for Ecology and Conservation, Biosciences, University of Exeter (Penryn Campus),  
Cornwall, TR10 9FE, UK

## Abstract

It is common to find considerable genetic variation in susceptibility to infection in natural populations. We have investigated whether natural selection increases this variation by testing whether host populations show more genetic variation in susceptibility to pathogens that they naturally encounter than novel pathogens. In a large cross-infection experiment involving four species of *Drosophila* and four host-specific viruses, we always found greater genetic variation in susceptibility to viruses that had coevolved with their host. We went on to examine the genetic architecture of resistance in one host species, finding that there are more major-effect genetic variants in coevolved host-parasite interactions. We conclude that selection by pathogens increases genetic variation in host susceptibility, and much of this effect is caused by the occurrence of major-effect resistance polymorphisms within populations.

## Introduction

From bacteria to plants and insects to humans, it is common to find considerable genetic variation in susceptibility to infection in natural populations [1-4]. This variation in susceptibility can determine the impact of disease on health and economic output [5-8]. In nature and breeding programs, it determines the ability of populations to evolve resistance to infection. Insect populations, like those of other organisms, typically contain considerable genetic variation in susceptibility to infection [2, 4, 9, 10], and provide a convenient laboratory model in which to investigate basic questions about how this variation is maintained [11]. Within vector species like mosquitoes, resistant genotypes are less likely to transmit parasites, and this has the potential to reduce disease in vertebrate populations [12]. Where pathogens are contributing the decline of beneficial species like pollinators, high levels of genetic variation may allow populations to recover [13]. Understanding the origins of genetic variation in susceptibility is therefore a fundamental question in infectious disease biology.

As pathogens are harmful, natural selection is expected to favour resistant host genotypes. Directional selection on standing genetic variation will drive alleles to fixation, removing variants from the population [14-16]. However, as directional selection also increases the frequency of mutations that change the trait in the direction of selection, at equilibrium it is expected to have no effect on levels of standing genetic variation (relative to mutation-drift balance; [17]). However, selection mediated by pathogens may be different. Coevolution with pathogens can result in the maintenance of both resistant and susceptible alleles by negative frequency dependent selection [18, 19]. Similarly, when infection prevalence exhibits geographical or temporal variation, selection can maintain genetic variation, especially if pleiotropic costs to resistance provide an advantage to susceptible individuals when infection is rare [20-22]. Even when there is simple directional selection on alleles that increase resistance, the direction of selection by pathogens may frequently change so populations may not be at equilibrium. If selection favours rare alleles – such as new mutations – directional selection can transiently increase genetic variation during their spread through the population [23-25].

To examine how selection by pathogens affects levels of genetic variation we used a natural host-parasite system; *Drosophila* and sigma viruses [26, 27]. Sigma viruses are a clade of insect RNA viruses with negative-sense genomes in the family Rhabdoviridae [28-31]. They are vertically transmitted, and each sigma virus infects a single host species, simplifying studies of coevolution [28, 31]. In *Drosophila melanogaster* infection reduces host fitness by approximately 25% and prevalence in wild populations is typically around 10% [32, 33]. As such, there is the necessary selective pressure for resistance to evolve [25, 34]. In *D. melanogaster*, three major-effect resistance alleles have been identified [2, 11, 24, 25, 34-36]. There has been a recent sweep of genotypes of *D. melanogaster* sigma virus (DMelSV) that are able to overcome one of these host resistance genes [37-39]. Given the power of *Drosophila* genetics, this system is an excellent model of a coevolutionary arms race between hosts and parasites.

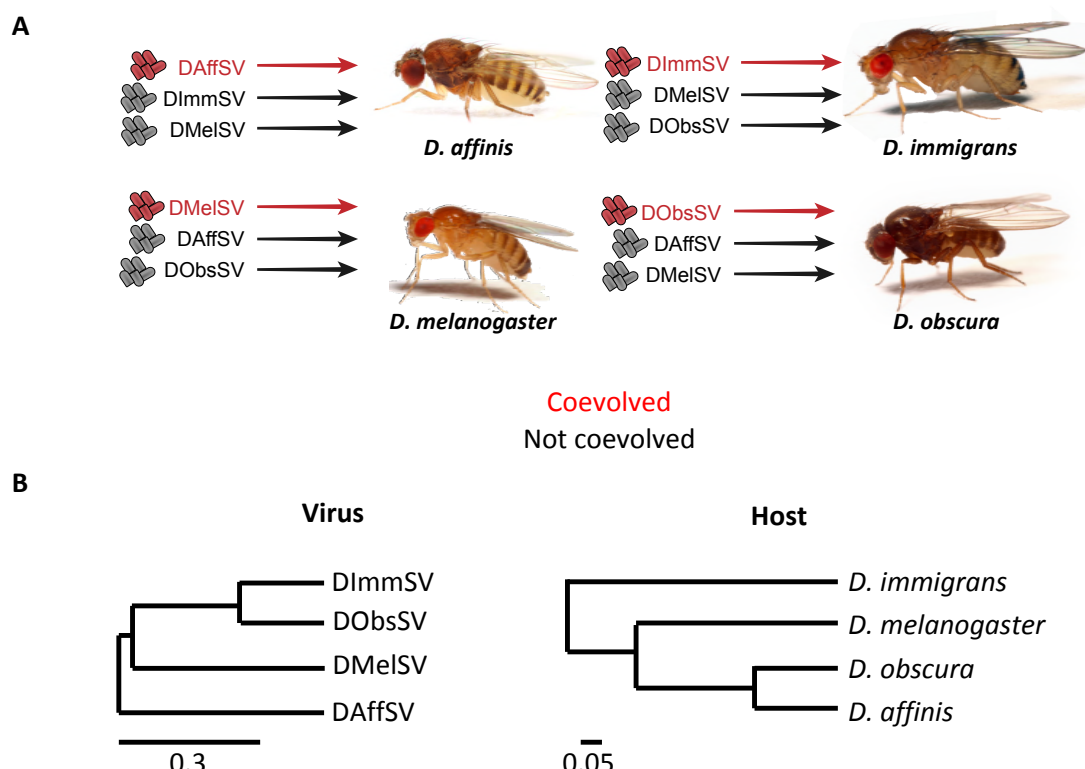
Sigma viruses offer a novel way to test how coevolution with a pathogen alters the amount of genetic variation in host susceptibility. As sigma viruses are vertically transmitted, we can be certain of which hosts and viruses are natural coevolving combinations, and which do not naturally coevolve. In this study we have used four species of *Drosophila* (*D. affinis*, *D. immigrans*, *D. melanogaster* and *D. obscura*) that shared a common ancestor approximately 40 million years ago [40, 41], and their natural sigma viruses (DAffSV, DImmSV, DMelSV and DObsSV respectively) [29, 30, 42]. We have tested whether selection by parasites increases the amount of genetic variation in host susceptibility by comparing viruses that naturally infect and coevolve with each species to non-coevolving viruses. We then examined how selection by pathogens alters the genetic architecture of resistance by mapping loci that confer resistance to DAffSV, DMelSV and DObsSV in *D. melanogaster*.

## Results

### Genetic variation in susceptibility to infection is greatest in coevolved host-virus associations

To test whether selection by parasites increases genetic variation in susceptibility to infection, we compared coevolved host-virus associations with novel associations that have no history of coevolution. We used four different species of *Drosophila*, each of which is naturally host to a different sigma virus [26-28, 30, 31]. We collected four species from the wild and created genetically diverse populations in the laboratory. Using flies from these populations we crossed single males to single females to create full-sib families. The progeny of these crosses were then injected with either the virus isolated from that species, or a virus isolated from one of the other species (Figure 1). Studies on DMelSV have shown that those loci providing resistance against naturally acquired infections in the lab and wild also provide resistance when the virus is injected, and transmission rates are correlated to viral load [2, 27, 43-46]. Fifteen days post infection we extracted RNA from the flies and measured viral load by quantitative RT-PCR. This allowed us to estimate the genetic variance ( $V_G$ ) in viral load from the across-family variance. In total we infected 52,592 flies and

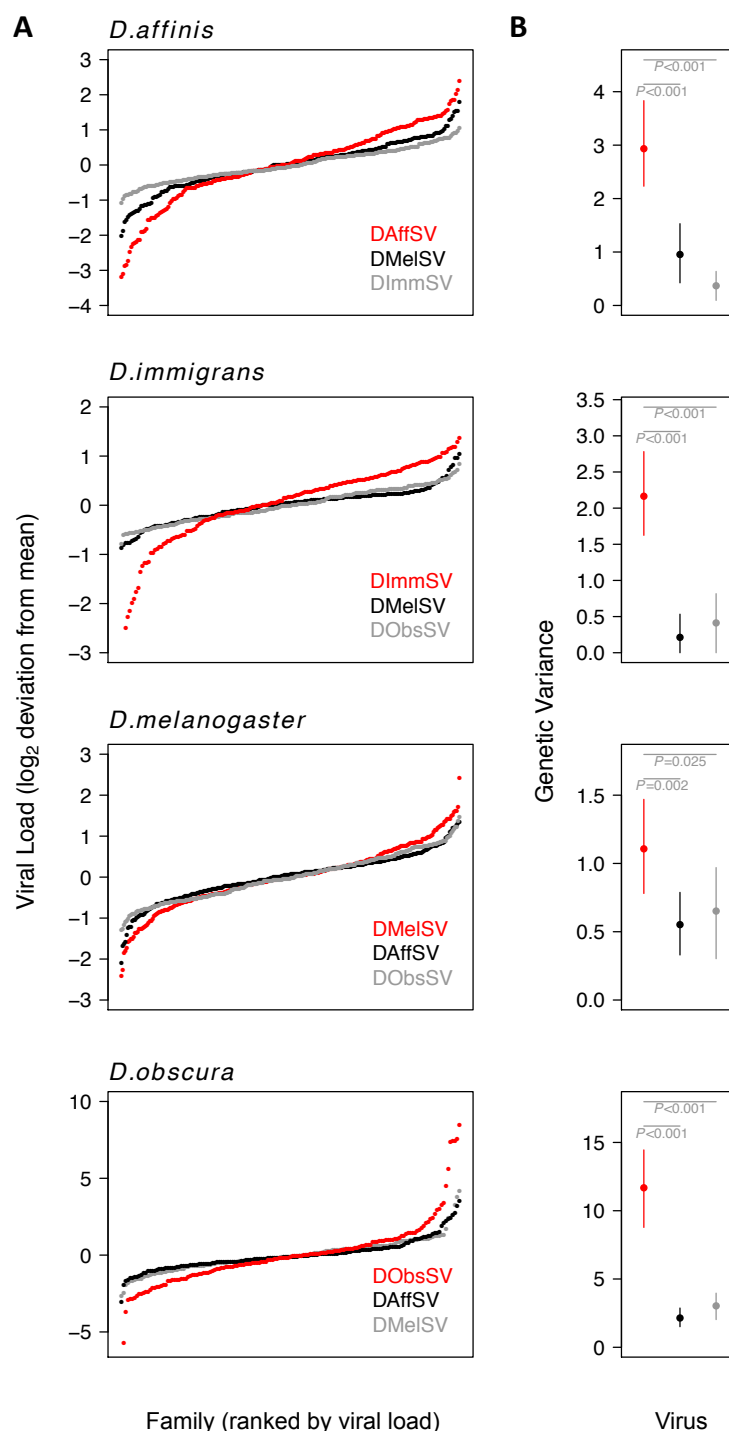
measured the viral load in 4295 biological replicates across 1436 full-sib families (details of sample sizes in Supplementary Methods).



**Figure 1. Experimental design and phylogenies of host and parasite.** A) Four species of *Drosophila* were independently infected both with a sigma virus with which they have coevolved in nature (red) and two viruses that naturally infect another species (black). B) Phylogenies of the sigma viruses and their *Drosophila* hosts, redrawn from [29, 47]. Scale bars represent substitutions per site under a relaxed clock model. Sigma viruses are highly divergent from one another with amino acid identities of <55% in the most conserved gene. The four species of *Drosophila* are estimated to have shared a common ancestor approximately 40 million years ago [40, 41].

Within populations of all four species, we found significantly greater genetic variance in susceptibility to the sigma virus that naturally infects that species compared to viruses from other species (Figure 2 and Supplementary Tables S1 and S2). These different variances reflect considerable differences in the mean viral load between families (Figure 2A). For example, when families of *D. obscura* in the 2<sup>nd</sup> and 98<sup>th</sup> percentile were compared, there was a 1294 fold difference between the viral loads of the coevolved virus in the families (Figure 2A). In contrast, for the non-coevolved viruses there was a 27 fold difference in DMelSV loads and a 19 fold difference in DAffSV loads (see Figure 2 for statistics). In *D. affinis* the equivalent statistics were a 26 fold difference in the load of the coevolved virus, compared to a 7-fold and 3 fold difference in DMelSV loads and DImmSV loads. In *D. immigrans* there was an 11-fold difference for the coevolved virus compared to a 2-fold difference for DObsSV and 3-fold for DMelSV. In *D. melanogaster* there was a 9-fold difference for the coevolved virus loads compared to a 4-fold difference for DObsSV and 5-

151 fold for DAffSV. Across the four species, there was no consistent difference between the  
152 mean viral load in coevolved versus non-coevolved associations, and the genetic variance in  
153 viral load was not correlated with the mean viral load (Supplementary Figure S1, Spearman's  
154 correlation:  $\rho = -0.38$ ,  $S = 296$ ,  $P = 0.22$ ).  
155



156  
157  
158 **Figure 2. Genetic variation in susceptibility to coevolved and non-coevolved viruses.** The  
159 viral load was measured 15 days post infection by quantitative RT-PCR relative to a  
160 *Drosophila* reference gene (*RpL32*). **A)** The points show model prediction family means from

our GLM and are centred on zero. The number of families in each panel was down-sampled so the same number of families is shown for each virus. Coevolved host-virus associations are in red. **B)** The genetic variance in  $\log_2$  viral load was estimated from the between family variance assuming that all genetic variance is additive. The bars are 95% credible intervals. Posterior probabilities for significantly different genetic variances are shown in grey (see Table S1 and S2).

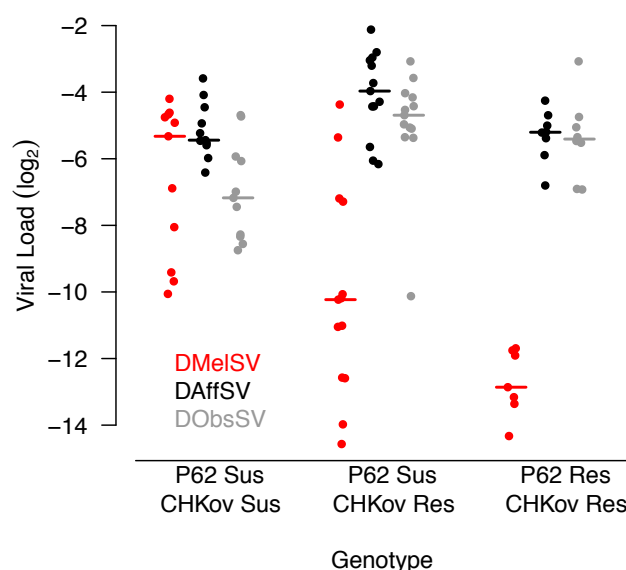
# **Major-effect genetic variants that are known to provide resistance to DMelSV do not protect against other viruses**

To examine whether the genetic basis of resistance to coevolved and non-coevolved viruses was different we estimated the genetic correlations ( $r_g$ ) in their viral loads. In *D. melanogaster* these were 0.40 for DMelSV-DAffSV, (95% CIs: 0.20, 0.61) and 0.25 for DMelSV-DObsSV (95% CIs: -0.01, 0.47). In the other species our estimates sometimes had wide credible intervals, but the genetic correlations between coevolved and non-coevolved viruses were mostly below 0.5 (Table S3). Therefore if natural selection increases genetic variation in susceptibility to a natural pathogen, there is expected to be a smaller effect on non-coevolved viruses.

In *D. melanogaster* a substantial proportion of the genetic variance in susceptibility to DMelSV is explained by major-effect variants in the genes *CHKov1* and *p62* (also known as *Ref(2)P*) [11, 24, 25, 36, 48]. The resistant allele in each of these genes has arisen recently by mutation and been driven up in frequency by natural selection, presumably due to the presence of DMelSV in natural populations [24, 25]. Therefore, if these genetic variants confer resistance to DMelSV but not the other sigma viruses, then this may explain the differences in genetic variance that we observed.

To examine whether *CHKov1* or *p62* contributed to the differences in genetic variance we observed in *D. melanogaster*, we genotyped the parents of the full sib families for variants that confer resistance (Table S4). Assuming the effects of the resistant alleles are additive, we estimated that the load of the coevolved virus DMelSV was more than halved in homozygous *CHKov1* resistant flies compared to susceptible flies (reduction in  $\log_2$  viral load = 1.2, 95% CI = 0.6, 1.8). In contrast we found no significant effect of this gene on loads of the non-coevolved viruses (DAffSV = -0.2, 95% CI = 0.2, -0.8; DObsSV = -0.4, 95% CI = -0.04, 1.0). The resistant allele of *p62* was present at such a low frequency (1.5%) in the population that we lacked of statistical power to investigate its effects.

To confirm these results we infected 1869 flies from 32 inbred *D. melanogaster* lines [49] that had known *CHKov1* or *p62* genotypes. The effect of these variants was greater on the naturally occurring virus than the viruses from other species (Figure 3; effect genotype on DMelSV load:  $F_{2,28}=13.2$ ,  $P=0.00001$ ; DAffSV:  $F_{2,29}=4.9$ ,  $P=0.01$ ; DObsSV:  $F_{2,29}=5.7$ ,  $P=0.01$ ).

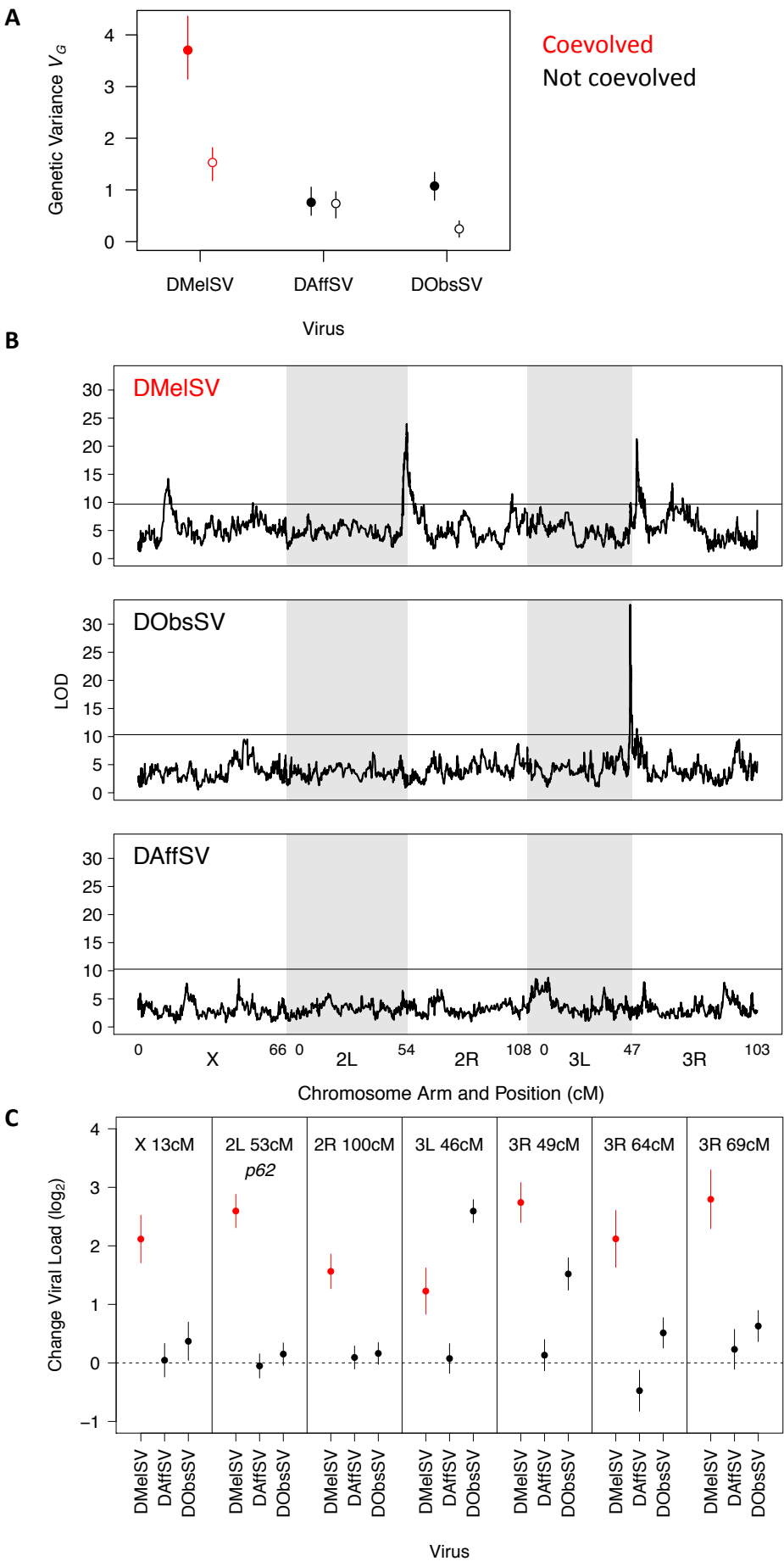


**Figure 3. Viral load in *D. melanogaster* lines carrying different alleles of *CHKov1* and *p62*.** Each point is the viral load of a separate inbred fly line carrying the resistant (R) or susceptible (S) allele of *P62* or *CHKov1*. Horizontal bars are medians. Viral load was measured 15 days post infection by quantitative RT-PCR relative to a *Drosophila* housekeeping gene (*RpL32*).

### There are a greater number of major-effect variants in coevolved host-virus associations

To investigate how coevolution shapes the genetics of resistance in an unbiased way, we mapped loci controlling resistance using a *D. melanogaster* advanced intercross population (the DSPR panel [50]). This population samples genetic variation in a small number of genotypes from around the world (the experiments above sampled many genotypes from a single location). It was founded by allowing two sets of 8 inbred founder lines to interbreed in large populations for 50 generations, then creating recombinant inbred lines (RILs) whose genomes are a fine-scale mosaic of the original founder genomes. We used 377 RILs from these populations, which have up to 15 alleles of each gene (one founder line is shared between the two populations). We infected 15,916 flies across 1362 biological replicates with DMelSV, DAffSV or DObsSV and measured viral load as above.

We first estimated the genetic variance in viral load within our mapping population. The results recapitulated what we had found above in a natural population of flies — there was considerably more genetic variation in susceptibility to the coevolved virus than the non-coevolved viruses (Figure 4A, filled circles). Therefore, our earlier result from a single population holds when sampling flies from across 6 continents, although the magnitude of the effect is considerably greater in this mapping population.





**Figure 4. The genetic architecture of resistance to coevolved and non-coevolved viruses in *D. melanogaster*.** A) The genetic variance in viral load within the mapping population (filled circles). The open circles are estimates of the genetic variance after accounting to the effects of the QTL in panel C. Error bars are 95% credible intervals. (B) QTL affecting viral load. The horizontal line shows a genome-wide significance threshold of  $P < 0.05$  that was obtained by permutation. (C) The effect of the seven QTL detected on the load of the three viruses. Only QTL that remained were significant following multiple regression with all the loci are shown. The coevolved virus is shown in red.

To examine the genetic basis of virus resistance, we looked for associations between genotype and viral load across the genome (Figure 4B). In the coevolved association (DMelSV) we identified seven QTL associated with resistance, compared to one that affects DObsSV and none affecting DAffSV (Table S5; this excludes one DMelSV on the X chromosome that did not remain significant after accounting for the other QTL). The QTL affecting DObsSV also has a significant effect on DMelSV. One of the QTL corresponded to *p62* (2L 53cM). The susceptible allele of *CHKov1* was not present in the fly lines assayed.

To examine the effect that the QTL have on viral load, we first split the founder alleles into a resistant class and a susceptible class (see methods) and then estimated the difference in viral load between the functionally distinct alleles. Six of the seven QTL resulted in greater reductions in the load of the coevolved virus (DMelSV) than the viruses isolated from other species (Figure 4C). There were only two cases where there was substantial cross-resistance to multiple viruses—3R 49cM confers strong resistance to DMelSV and weak resistance to DObsSV, while 3L 46cM confers weak resistance to DMelSV and strong resistance to DObsSV.

Together, this modest number of loci with substantial effects on resistance explains most of the high genetic variance in resistance to the coevolved virus (Figure 4A, filled versus open circles). Individually resistant alleles cause an approximate 3-7 fold reduction in viral load (Figure 4C), and together they explain 59% of the genetic variance in susceptibility to DMelSV, 77% for DObsSV and 3% for DAffSV (Figure 4A, filled versus open circles). However, even after accounting for these genes there remains a significantly higher genetic variance in the viral load of the coevolved virus (Figure 4A, open circles, non-overlapping 95% CI).

## Discussion

We have found greater genetic variation in susceptibility to pathogens that naturally coevolve with a host compared to those that do not, suggesting that selection by pathogens acts to increase the amount of genetic variation in susceptibility. This effect was largely caused by a modest number of major-effect genes that explain over half of the genetic variance in resistance.

There are several different ways in which selection by pathogens could increase genetic variation. Negative frequency dependent selection can maintain genetic variation if the alleles conferring resistance are costly or increase susceptibility to other pathogen

genotypes [18]. Directional selection can cause transient increases in genetic variance if it increases the frequency of rare or *de novo* resistance mutations [23]. Temporal or geographical variation in the prevalence of pathogens could maintain both resistant and susceptible alleles [20, 21].

As the genetic variants in *p62 (Ref(2)P)* and *CHKov1* that confer resistance to DMelSV have been identified, this has previously allowed us to use patterns of DNA sequence variation to infer how selection has acted on resistance in *D. melanogaster*. In both these genes the resistant alleles have arisen relatively recently by mutation and natural selection has pushed them rapidly up in frequency, leaving a characteristic signature of elevated linkage disequilibrium and low genetic diversity around the variant causing resistance [24, 25, 36]. There is no indication of negative frequency dependant selection, and these polymorphisms appear to have arisen from partial selective sweeps [24, 25].

At equilibrium directional selection on a trait is not expected to affect its genetic variance (relative to a population under mutation-drift balance; [17]). However, the genetic variance will transiently increase if the variants under selection are initially at low frequency [23], as was the case for both *p62* and *CHKov1* [24, 25]. A particular feature of pathogens is that the direction of selection is likely to continually change, as new pathogens appear in populations or existing pathogens evolve to overcome host defences. For example, in France and Germany in the 1980's DMelSV evolved to largely overcome the effects of the resistant allele of *p62* [51, 52]. Similarly, DImmSV has swept through European populations of *D. immigrans* in the last ~16 years and DObsSV through UK populations of *D. obscura* in the last ~11 years [28, 31]. If selection by pathogens continually changes and resistance evolves from new mutations, then this may cause a sustained increase in genetic variance in susceptibility to infection.

Association studies in humans have revealed that quantitative traits are typically controlled by a very large number of genetic variants, each of which tends to have a very small effect [53, 54]. However, in humans some have advocated the view that susceptibility to infectious disease is qualitatively different from other traits and has a much simpler genetic basis [55]. Our results support this, with seven polymorphisms explaining over half the genetic variance. This confirms our previous work in *D. melanogaster* showing a simple genetic basis of virus resistance [11, 56]. As these genetic variants mostly only affect the naturally occurring pathogen of *D. melanogaster*, our results suggest that not only is selection by pathogens increasing the genetic variance but it is also altering the genetic architecture of the trait by introducing major-effect variants into the population. One explanation for this observation is that most quantitative traits are under stabilising selection, so major effect variants will tend to be deleterious and removed by selection [57]. In contrast, selection by pathogens likely changes through time and populations may be far from their optimal level of resistance. If this is the case, Fisher's geometric model predicts that major effect variants will be favoured by directional selection [58]. Similar arguments may apply if resistance polymorphisms are under negative frequency dependant selection.

A major source of emerging infectious disease is pathogens jumping into novel hosts where they have no co-evolutionary history [59, 60]. Our results suggest that when a pathogen infects a novel host species, there may be far less genetic variation in susceptibility among individuals than is normally the case. This may create a 'monoculture effect' [6, 61, 62], which could leave populations vulnerable to epidemics of pathogens that have previously circulated in other host species. Longer term, low levels of pre-standing genetic variation may also slow down the rate at which the new host can evolve resistance to a new pathogen.

In conclusion, we have demonstrated that selection by pathogens can increase the amount of genetic variation in host susceptibility and result in a simple genetic architecture where resistance is largely controlled by major effect resistance loci.

## Methods

### *Virus extraction and infection*

We extracted the sigma viruses DAffSV, DIImmSV, DMelSV and DObsSV from infected stocks of *D. affinis* (line: NC10), *D. immigrans* (line: DA2), *D. melanogaster* (line: E320 Ex) and *D. obscura* (line: 10A) respectively [28-31, 44]. These infected lines were collected from the wild between 2007-2012 (all from the UK, bar *D. affinis* which was collected in the USA). Infected fly stocks were checked for infection by exposing the flies to 100% CO<sub>2</sub> at 12°C for 15mins then paralysed flies were collected 30mins later. The DIImmSV infected line does not show CO<sub>2</sub> sensitivity and so was confirmed to have a high level of infection using RT-PCR. Infected flies were frozen at -80°C and later homogenised in ringers solution (2.5µl per fly) and centrifuged at 13,000g for 10min at 4°C. The supernatant was collected, 2% v/v FBS was added then virus solutions were aliquoted and stored at -80°C.

### *Experimental design*

We set up a common garden experiment to measure genetic variation in susceptibility to natural and non-natural viruses across four host species. For each species we infected it with its own virus, as well as two viruses that do not infect that host species (see Figure 1). All fly stocks used were tested for existing sigma virus infection using RT-PCR over two generations. For all species we collected flies from the wild and we used a full-sib mating design. The progeny of these crosses was infected by injecting them with the viruses intrathoracically and measuring viral RNA loads 15 days post infection, as in [63]. This time point was selected as RNA viral load tends to plateau from around day 15 post infection and there is no significant mortality from infection in this period. The specifics for each host species are detailed in the supplementary methods.

### *Known resistance genes in D. melanogaster*

We genotyped parents of each *D. melanogaster* full sib family from the experiment above for two resistance alleles that are known to confer protection against DMelSV; *p62 (Ref(2)P)* and *CHKov1*. We genotyped parental flies using PCR assays that produce different sized products depending on whether flies carry resistant or susceptible alleles. Information on these PCRs and primer sequences can be found in table S3. We then calculated the number of resistance alleles in each family by summing the number of alleles from both mothers and fathers. We produced genotype information for 230 of the 255 families.

Another resistance allele has been identified in the gene *Ge-1* [34]. However, this allele has been found to occur at a low frequency in wild populations. We genotyped a subset of 184 parental flies from our experiment and found the resistant allele was not present, suggesting it is rare or absent.

We further examined the effect of alleles known to affect susceptibility to DMelSV on all three viruses. Firstly, we infected 32 lines from the Drosophila Genetic Reference Panel (DGRP) [49] that were susceptible for both *p62* and *CHKov1* ( $n=11$  lines), were resistant for *CHKov1* only ( $n=13$  lines), or were resistant for both genes ( $n=8$  lines) with DMelSV, DAffSV and DObsSV. No lines in the panel were resistant for *p62* and susceptible for *CHKov1*. We infected a mean of 18 flies per line (range = 3-22).

### Mapping resistance genes in *D. melanogaster*

We used 377 DSPR lines (154 from panel A and 223 from panel B, <http://FlyRILs.org> [50, 64], kindly provided by S.J. Macdonald, University of Kansas) to carry out a Quantitative Trait Locus (QTL) study to examine the genetic basis of resistance to DAffSV, DMelSV and DObsSV in *D. melanogaster*.

3 females and 3 males from each DSPR line were placed into yeasted cornmeal vials and allowed to lay for 3-4 days, at 25°C. Male offspring were collected at 0-4 days post-eclosion, and placed at 18°C for 4-6 days. Flies were then injected with DMelSV, DAffSV or DObsSV as described above. Injected males were maintained on unyeasted cornmeal at 18°C, and frozen on day 15 post-infection as above.

Injections were carried out over 13 weeks. Each day of injection a mean of 47 unique lines (range 20-60) and 51 replicate vials were injected with 1-3 different viruses. In total we assayed 377 DSPR lines (108 lines had 2 biological replicates). Each replicate vial contained a mean of 12 flies (range 1-22). In total, 15,916 flies were injected across both panels of DSPR lines. We injected 319 lines with all 3 viruses, 38 with 2 viruses and 20 with 1 virus. The order of injection of lines and of viruses, was randomised across injection days. Independent biological replicates were injected on different days. Panel A and Panel B lines were assayed in two overlapping blocks.

### Measuring viral load

We measured the change in RNA viral load using qRT-PCR. The viral RNA load was expressed relative to the endogenous control housekeeping gene *RpL32* (*Rp49*). RNA was extracted from flies homogenised in Trizol and reverse transcribed with GoScript reverse transcriptase (Promega) and random hexamer primers, and then diluted 1:10 with nuclease free water. The qRT-PCR was performed on an Applied Biosystems StepOnePlus system using Sensifast Hi-Rox Sybr kit (Bioline) with the following PCR cycle: 95°C for 2min followed by 40 cycles of: 95°C for 5 sec followed by 60°C for 30 sec. Two qRT-PCR reactions (technical replicates) were carried out per sample with both the viral and endogenous control primers. Each qRT-PCR plate contained three standard samples, and all experimental samples were split across plates in a blocked design. A linear model was used to correct the cycle threshold (Ct) values for differences between qRT-PCR plates. Primer sequences are in Table S6.

To estimate viral load, we calculated  $\Delta Ct$  as the difference between the qRT-PCR cycle thresholds of the virus and the endogenous control. Viral load calculated without using the endogenous control is strongly correlated to  $\Delta Ct$  for all species.

### Statistical analysis full-sib experiments

We used a linear mixed model to examine the amount of genetic variation in susceptibility to the different viruses. We used a trivariate model with the load of the three viruses as the response variable. For each species the model was structured as:

(1)

$$y_{vfi} = \beta_{l,v} + u_{v:f} + u_{v:d} + e_{vfi}$$

Where  $y_{vfi}$  is the  $\log_2$  viral load of the  $i^{th}$  biological replicate of full-sib family  $f$  infected with virus  $v$ .  $\beta$  are the fixed effects, with  $\beta_1$  being the mean viral load of each virus.  $u$  are the random effects for full-sib families ( $f$ ) and for the day of injection ( $d$ ),  $e$  are the residuals. By assuming that all the genetic variation in the population is additive [65], we estimated the genetic variance ( $V_G$ ) of the viral load as twice the between-family variance [14, 15]. Both empirical data and theory suggest additive genetic variation makes up large proportion of the total genetic variance [65].

In addition, for *D. melanogaster* we ran a further model that included the additional fixed effects  $\beta_{2,v}$  and  $\beta_{3,v}$  that are the linear effects of the *CHKov1* and *p62 (Ref(2)P)* resistance alleles. We assumed these genes had additive effects, and modelled their effects simply as the proportion of resistant alleles in a family (if one parent was heterozygous and the other homozygous susceptible, the value is 0.25).

The model was fitted using the MCMCglmm package in R [66]. The random effects (and residuals) are assumed to be multivariate normal with zero mean and covariance structure  $\mathbf{V} \otimes \mathbf{I}$ .  $\mathbf{I}$  is an identity matrix, and  $\mathbf{V}$  a matrix of estimated variances and covariances. For the random effects  $\mathbf{V}$  is a 3x3 covariance matrix describing the variances for each virus and the covariances between them. The off-diagonal elements of  $\mathbf{V}$  for the residual were set to zero because the covariances between traits at these levels are not estimable by design.

Diffuse independent normal priors were placed on the fixed effects (means of zero and variances of  $10^8$ ). Parameter expanded priors were placed on the covariance matrices resulting in scaled multivariate  $F$  distributions which have the property that the marginal distributions for the variances are scaled (by 1000)  $F_{1,1}$ . The exceptions were the residual variances for which an inverse-gamma prior was used with shape and scale equal to 0.001. The MCMC chain was ran for 130 million iterations with a burn-in of 30 million iterations and a thinning interval of 100,000.

We confirmed the results were not sensitive to the choice of prior by also fitting models with inverse-Wishart and flat priors for the variance covariance matrices (described in [63]), as well as fitting the models by REML in ASReml R [67]. These analyses all gave qualitatively similar results (data not shown).

### Statistical analysis of QTL experiment

Based on genotyping data, the probability that each Recombinant Inbred Line (RIL) in the DSPR panel was derived from each of the 8 founder lines has been estimated at 10kB intervals across the genome (King et al. 2012). To identify QTL affecting viral load we first calculated the mean viral load ( $\Delta Ct$ ) across the biological replicates of each RIL. We then regressed the mean viral load against the 8 genotype probabilities and calculated logarithm of odds (LOD) scores using the DSPRqtl package in R (King et al. 2012). These LOD scores were calculated separately for DSPR Panel A and Panel B, and then summed at each genomic location. To obtain a significance threshold, we permuted our mean viral load estimates across the RILs within each panel, repeated the analysis above and recorded the highest LOD score across the entire genome. This process was repeated 1000 times to obtain a null distribution of the maximum LOD score.

To estimate the effect of each QTL we assumed that there was a single genetic variant affecting viral load, so the founder alleles could be assigned to two functionally distinct allelic classes. First, we regressed the mean viral load against the genotype probabilities (as described above), resulting in estimates of the mean viral load of each founder allele in the dataset. The two DSPR panels had one founder line (line 8) in common. For QTL where the line 8 allele was present in both panels, this analysis included data from both panels, and 'panel' was included as a fixed effect in the analysis. When this was not the case, we analysed only data from the panel where the QTL was most significant. We then ranked the founder alleles by viral load estimate, and split this ranked list into all possible groups of two alleles. For each split the genotype probabilities in the first group of founder alleles were summed. We then regressed mean viral load against each of these combined genotype probabilities. The regression model with the highest likelihood was taken as the most likely classification into allelic classes. The effect size of the QTL was then estimated from this model.



To estimate the genetic variance in viral load within the DSPR panels we modified the model described in Equation 1 as follows.  $\gamma_{vfi}$  is the  $\log_2$  viral load of the  $i^{th}$  biological replicate of each RIL  $f$  infected with virus  $v$ . There was a single fixed effect,  $\beta$ , of the panel the line is from.  $u$  are the random effects for each RIL ( $f$ ). The day of injection ( $d$ ) was omitted. As all the RILs are homozygous, we estimated the genetic variance in viral load ( $V_G$ ) as half the between-RIL variance. This assumes all the genetic variation is additive.

To estimate the proportion of the genetic variance that is explained by the QTL we identified, we repeated this analysis but included the 7 QTL we identified as fixed effects in the model. Each QTL was included by estimating the probability that each line carried the resistant allele of the QTL and adding this as a fixed effect to the model. The between-RIL variance then allowed us to estimate the genetic variance in viral load after removing the effects of the QTL.

### Data availability

Datasets and R code for estimating the amount of genetic variation in susceptibility

<https://doi.org/10.6084/m9.figshare.6743339>

DGRP dataset <https://doi.org/10.6084/m9.figshare.6743354>

DSPR dataset and R code <https://doi.org/10.6084/m9.figshare.7195751>

### Acknowledgements

Many thanks to: Alastair Wilson and Jarrod Hadfield for useful advice and discussion; Stuart Macdonald for providing DSPR lines, Trudy Mackay for providing the DGRP fly lines and Kelly Dyer with help collecting *D. affinis*; Darren Obbard for providing photographs of *Drosophila* species; Camille Bonneaud, Katherine Roberts, Ryan Imrie and the Unckless lab group for constructive comments on this work.

### Competing interests

We declare no competing interests.

### Funding

Natural Environment Research Council grant (NE/L004232/1 <http://www.nerc.ac.uk/>)

European Research Council grant (281668, *Drosophila* Infection, <http://erc.europa.eu/>)

Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 109356/Z/15/Z)

# References

1. Chapman S.J., Hill A.V. 2012 Human genetic susceptibility to infectious disease. *Nat Rev Genet* **13**(3), 175-188. (doi:10.1038/nrg3114).
2. Bangham J., Kim K.W., Webster C.L., Jiggins F.M. 2008 Genetic variation affecting host-parasite interactions: Different genes affect different aspects of sigma virus replication and transmission in *Drosophila melanogaster*. *Genetics* **178**(4), 2191-2199.
3. Hammond-Kosack K.E., Jones J.D. 1997 Plant Disease Resistance Genes. *Annu Rev Plant Physiol Plant Mol Biol* **48**, 575-607. (doi:10.1146/annurev.arplant.48.1.575).
4. Lazzaro B.P., Scurman B.K., Clark A.G. 2004 Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science* **303**(5665), 1873-1876.
5. Cooke G.S., Hill A.V. 2001 Genetics of susceptibility to human infectious disease. *Nat Rev Genet* **2**(12), 967-977. (doi:10.1038/35103577).
6. King K.C., Lively C.M. 2012 Does genetic diversity limit disease spread in natural host populations[quest]. *Heredity* **109**(4), 199-203.
7. Alonso-Blanco C., Mendez-Vigo B. 2014 Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. *Curr Opin Plant Biol* **18**, 37-43. (doi:10.1016/j.pbi.2014.01.002).
8. Burgner D., Jamieson S.E., Blackwell J.M. 2006 Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *The Lancet Infectious diseases* **6**(10), 653-663. (doi:10.1016/s1473-3099(06)70601-6).
9. Tinsley M.C., Blanford S., Jiggins F.M. 2006 Genetic variation in *Drosophila melanogaster* pathogen susceptibility. *Parasitology* **132**, 767-773. (doi:10.1017/S0031182006009929).
10. Obbard D.J., Dudas G. 2014 The genetics of host-virus coevolution in invertebrates. *Current opinion in virology* **8**, 73-78. (doi:10.1016/j.coviro.2014.07.002).
11. Magwire M.M., Fabian D.K., Schweyen H., Cao C., Longdon B., Bayer F., Jiggins F.M. 2012 Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *Plos Genet* **8**(11), e1003057. (doi:10.1371/journal.pgen.1003057).
12. Beerntsen B.T., James A.A., Christensen B.M. 2000 Genetics of mosquito vector competence. *Microbiol Mol Biol Rev* **64**(1), 115-137.
13. Maori E., Tanne E., Sela I. 2007 Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes (vol 362, pg 342, 2007). *Virology* **368**(1), 218-218. (doi:10.1016/j.virol.2007.07.013).
14. Falconer D.S. 1960 *Introduction to Quantitative Genetics*. Edinburgh/London, Oliver & Boyd.
15. Falconer D.S., Mackay T.F.C. 1996 *Introduction to Quantitative Genetics*,. Longmans Green, Harlow, Essex, UK.
16. Blows M.W., Hoffmann A.A. 2005 A reassessment of genetic limits to evolutionary change. *Ecology* **86**(6), 1371-1384. (doi:10.1890/04-1209).
17. Hill W.G. 1982 Rates of change in quantitative traits from fixation of new mutations. *Proc Natl Acad Sci U S A* **79**(1), 142-145.
18. Woolhouse M.E.J., Webster J.P., Domingo E., Charlesworth B., Levin B.R. 2002 Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32**(4), 569-577.
19. Haldane J.B.S. 1949 Disease and evolution. *Ric Sci Suppl* **19**, 68-76.
20. Nuismer S.L., Thompson J.N., Gomulkiewicz R. 2003 Coevolution between hosts and parasites with partially overlapping geographic ranges. *Journal of Evolutionary Biology* **16**(6), 1337-1345. (doi:10.1046/j.1420-9101.2003.00609.x).

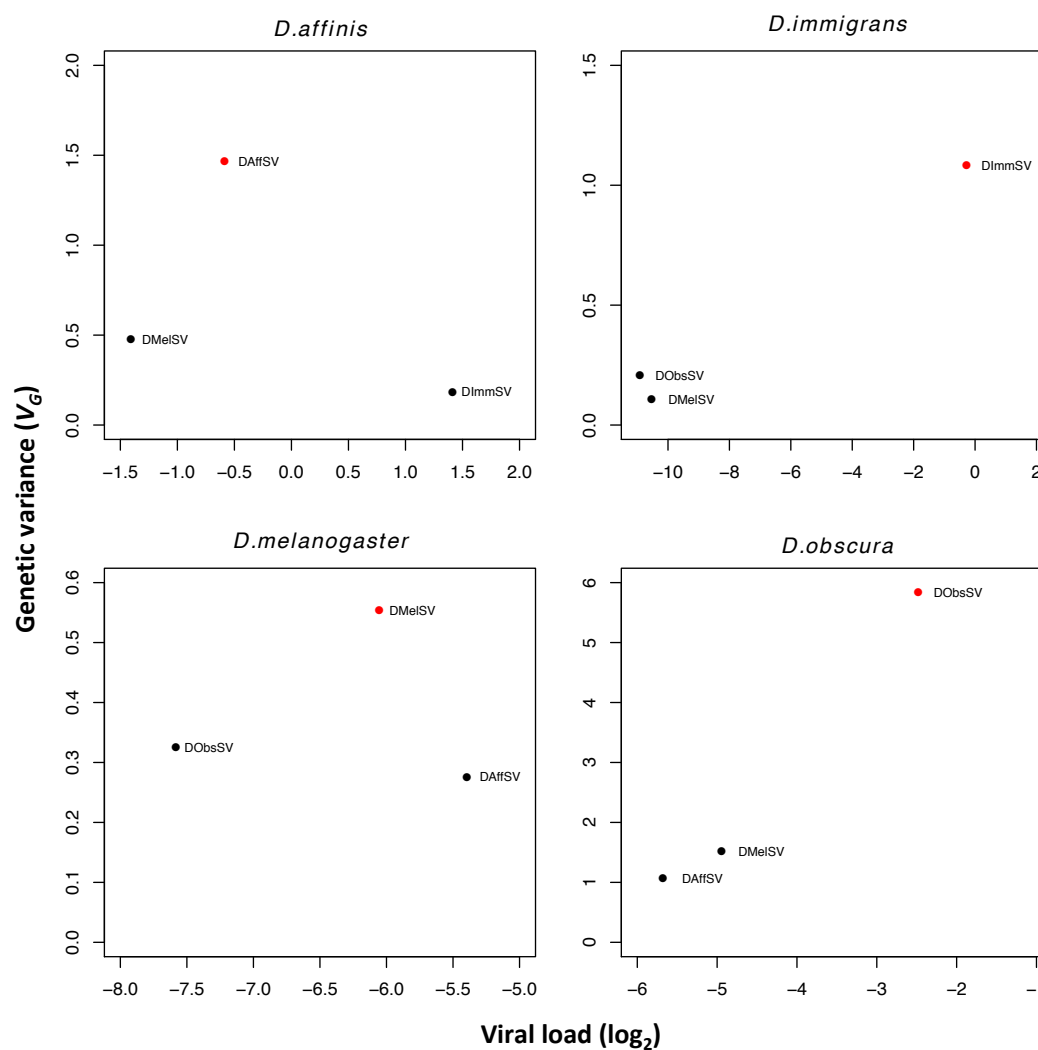


21. Thompson J.N. 1999 Specific hypotheses on the geographic mosaic of coevolution. *American Naturalist* **153**, S1-S14. (doi:10.1086/303208).
22. Koskella B. 2018 Resistance gained, resistance lost: An explanation for host-parasite coexistence. *PLoS Biol* **16**(9), e3000013. (doi:10.1371/journal.pbio.3000013).
23. Barton N.H., Turelli M. 1987 Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genet Res* **49**(2), 157-173.
24. Bangham J., Obbard D.J., Kim K.W., Haddrell P.R., Jiggins F.M. 2007 The age and evolution of an antiviral resistance mutation in *Drosophila melanogaster*. *Proceedings of the Royal Society B-Biological Sciences* **274**(1621), 2027-2034.
25. Magwire M.M., Bayer F., Webster C.L., Chuan C., Jiggins F.M. 2011 Successive Increases in the Resistance of *Drosophila* to Viral Infection through a Transposon Insertion Followed by a Duplication. *Plos Genet* **7**(10), e1002337. (doi:10.1371/journal.pgen.1002337).
26. Longdon B., Jiggins F.M. 2012 Vertically transmitted viral endosymbionts of insects: do sigma viruses walk alone? *Proceedings of the Royal Society B* **279**(1744), 3889-3898. (doi:10.1098/rspb.2012.1208).
27. Longdon B., Wilfert L., Jiggins F.M. 2012 *The Sigma Viruses of Drosophila*. Norfolk, UK, Caister Academic Press.
28. Longdon B., Day J.P., Schulz N., Leftwich P.T., de Jong M.A., Breuker C.J., Gibbs M., Obbard D.J., Wilfert L., Smith S.C., et al. 2017 Vertically transmitted rhabdoviruses are found across three insect families and have dynamic interactions with their hosts. *Proc Biol Sci* **284**(1847). (doi:10.1098/rspb.2016.2381).
29. Longdon B., Murray G.G., Palmer W.J., Day J.P., Parker D.J., Welch J.J., Obbard D.J., Jiggins F.M. 2015 The evolution, diversity, and host associations of rhabdoviruses. *Virus Evolution* **1**(1), vev014.
30. Longdon B., Obbard D.J., Jiggins F.M. 2010 Sigma viruses from three species of *Drosophila* form a major new clade in the rhabdovirus phylogeny. *Proceedings of the Royal Society B* **277**, 35-44. (doi:10.1098/rspb.2009.1472).
31. Longdon B., Wilfert L., Obbard D.J., Jiggins F.M. 2011 Rhabdoviruses in two species of *Drosophila*: vertical transmission and a recent sweep. *Genetics* **188**(1), 141-150. (doi:10.1534/genetics.111.127696).
32. Wilfert L., Jiggins F.M. 2013 The dynamics of reciprocal selective sweeps of host resistance and a parasite counter-adaptation in *Drosophila*. *Evolution* **67**(3), 761-773. (doi:10.1111/j.1558-5646.2012.01832.x).
33. Yampolsky L.Y., Webb C.T., Shabalina S.A., Kondrashov A.S. 1999 Rapid accumulation of a vertically transmitted parasite triggered by relaxation of natural selection among hosts. *Evol Ecol Res* **1**(5), 581-589.
34. Cao C., Magwire M.M., Bayer F., Jiggins F.M. 2016 A Polymorphism in the Processing Body Component Ge-1 Controls Resistance to a Naturally Occurring Rhabdovirus in *Drosophila*. *PLoS Pathog* **12**(1), e1005387. (doi:10.1371/journal.ppat.1005387).
35. Bangham J., Knott S.A., Kim K.W., Young R.S., Jiggins F.M. 2008 Genetic variation affecting host-parasite interactions: major-effect quantitative trait loci affect the transmission of sigma virus in *Drosophila melanogaster*. *Molecular Ecology* **17**(17), 3800-3807. (doi:10.1111/j.1365-294X.2008.03873.x).
36. Wayne M.L., Contamine D., Kreitman M. 1996 Molecular population genetics of ref(2)P, a locus which confers viral resistance in *Drosophila*. *Molecular Biology and Evolution* **13**(1), 191-199.
37. Fleuriet A. 1988 Maintenance of a Hereditary Virus - the Sigma-Virus in Populations of Its Host, *Drosophila-Melanogaster*. *Evolutionary Biology* **23**, 1-30.
38. Wilfert L., Jiggins F.M. 2010 Host-parasite coevolution: genetic variation in a virus population and the interaction with a host gene. *J Evol Biol* **23**(7), 1447-1455. (doi:10.1111/j.1420-9101.2010.02002.x).

39. Wilfert L., Jiggins F.M. 2014 Flies on the move: an inherited virus mirrors *Drosophila melanogaster*'s elusive ecology and demography. *Molecular Ecology* **23**(8), 2093-2104. (doi:10.1111/mec.12709).
40. Obbard D.J., MacLennan J., Kim K.-W., Rambaut A., O'Grady P.M., Jiggins F.M. 2012 Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Molecular Biology and Evolution* **29**(11), 3459-3473.
41. Tamura K., Subramanian S., Kumar S. 2004 Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* **21**(1), 36-44. (doi:10.1093/molbev/msg236 msg236 [pii]).
42. Longdon B., Wilfert L., Osei-Poku J., Cagney H., Obbard D.J., Jiggins F.M. 2011 Host switching by a vertically-transmitted rhabdovirus in *Drosophila*. *Biology Letters* **7**(5), 747-750. (doi:10.1098/rsbl.2011.0160).
43. Bregliano J.C. 1970 Study of Infection of Germ Line in Female *Drosophila* Infected with Sigma Virus .2. Evidence of a Correspondance between Ovarian Cysts with Increased Virus Yield and Stabilized Progeny. *Annales De L Institut Pasteur* **119**(6), 685-704.
44. Brun G., Plus N. 1980 The viruses of *Drosophila*. In *The genetics and biology of Drosophila* (eds. Ashburner M., Wright T.R.F.), pp. 625-702. New York, Academic Press.
45. Ohanessian-Guillemain A. 1963 Etude De Facteurs Genetiques Controlant Les Relations Du Virus Sigma Et De La *Drosophile* Son Hote. *Annales De Genetique* **5**(1-2), G1-G64.
46. Wilfert L., Jiggins F.M. 2010 Disease association mapping in *Drosophila* can be replicated in the wild. *Biol Lett* **6**(5), 666-668. (doi:10.1098/rsbl.2010.0329).
47. Longdon B., Hadfield J.D., Day J.P., Smith S.C., McGonigle J.E., Cogni R., Cao C., Jiggins F.M. 2015 The Causes and Consequences of Changes in Virulence following Pathogen Host Shifts. *PLoS Pathog* **11**(3), e1004728. (doi:10.1371/journal.ppat.1004728).
48. Contamine D., Petitjean A.M., Ashburner M. 1989 Genetic resistance to viral infection: the molecular cloning of a *Drosophila* gene that restricts infection by the rhabdovirus sigma. *Genetics* **123**(3), 525-533.
49. Mackay T.F., Richards S., Stone E.A., Barbadilla A., Ayroles J.F., Zhu D., Casillas S., Han Y., Magwire M.M., Cridland J.M., et al. 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**(7384), 173-178. (doi:10.1038/nature10811).
50. King E.G., Macdonald S.J., Long A.D. 2012 Properties and power of the *Drosophila* Synthetic Population Resource for the routine dissection of complex traits. *Genetics* **191**(3), 935-949. (doi:10.1534/genetics.112.138537).
51. Fleuriet A., Periquet G. 1993 Evolution of the *Drosophila-Melanogaster* Sigma Virus System in Natural-Populations from Languedoc (Southern France). *Archives of Virology* **129**(1-4), 131-143.
52. Fleuriet A., Sperlich D. 1992 Evolution of the *Drosophila-Melanogaster*-Sigma Virus System in a Natural-Population from Tubingen. *Theoretical and Applied Genetics* **85**(2-3), 186-189.
53. Shi H., Kichaev G., Pasaniuc B. 2016 Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* **99**(1), 139-153. (doi:10.1016/j.ajhg.2016.05.013).
54. Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., et al. 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**(7), 565-569. (doi:10.1038/ng.608).
55. Hill A.V.S. 2012 Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**(1590), 840-849. (doi:10.1098/rstb.2011.0275).

56. Cogni R., Cao C., Day J.P., Bridson C., Jiggins F.M. 2016 The genetic architecture of resistance to virus infection in *Drosophila*. *Mol Ecol* **25**(20), 5228-5241. (doi:10.1111/mec.13769).
57. Gibson G. 2018 Population genetics and GWAS: A primer. *PLoS Biol* **16**(3), e2005485. (doi:10.1371/journal.pbio.2005485).
58. Fisher R.A. 1930 *The genetical theory of natural selection*. Oxford, UK, Clarendon Press.
59. Longdon B., Brockhurst M.A., Russell C.A., Welch J.J., Jiggins F.M. 2014 The Evolution and Genetics of Virus Host Shifts. *PLoS Pathog* **10**(11), e1004395. (doi:10.1371/journal.ppat.1004395).
60. Parrish C.R., Holmes E.C., Morens D.M., Park E.C., Burke D.S., Calisher C.H., Laughlin C.A., Saif L.J., Daszak P. 2008 Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* **72**(3), 457-470. (doi:10.1128/mmbr.00004-08).
61. Lively C.M. 2010 The effect of host genetic diversity on disease spread. *Am Nat* **175**(6), E149-152. (doi:10.1086/652430).
62. Ostfeld R.S., Keesing F. 2012 Effects of Host Diversity on Infectious Disease. *Annu Rev Ecol Evol S* **43**, 157-182. (doi:10.1146/annurev-ecolsys-102710-145022).
63. Longdon B., Hadfield J.D., Webster C.L., Obbard D.J., Jiggins F.M. 2011 Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathogens* **7**((9)), e1002260. (doi:10.1371/journal.ppat.1002260).
64. Long A.D., Macdonald S.J., King E.G. 2014 Dissecting complex traits using the *Drosophila* Synthetic Population Resource. *Trends Genet* **30**(11), 488-495. (doi:10.1016/j.tig.2014.07.009).
65. Hill W.G., Goddard M.E., Visscher P.M. 2008 Data and theory point to mainly additive genetic variance for complex traits. *Plos Genet* **4**(2). (doi:ARTN e1000008 10.1371/journal.pgen.1000008).
66. Hadfield J.D. 2010 MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**(2), 1-22.
67. Gilmour A., Gogel B., Cullis B., Welham S., Thompson R. 2002 ASReml User Guide Release 1.0.

719 **Supplementary materials**



720  
721 **Figure S1. Estimates of genetic variance plotted against mean viral load for each species-**  
722 **virus.**  
723

Host	Virus comparison	95% CIs	
<i>D. affinis</i>	<b>DAffSV</b> -DImmSV	0.862	1.699
	<b>DAffSV</b> -DMelSV	0.532	1.510
	DImmSV-DMelSV	-0.606	0.004
<i>D. immigrans</i>	<b>DImmSV</b> -DMelSV	0.630	1.331
	<b>DImmSV</b> -DObsSV	0.500	1.240
	DMelSV-DObsSV	-0.365	0.167
<i>D. melanogaster</i>	<b>DMelSV</b> -DAffSV	0.080	0.485
	<b>DMelSV</b> -DObsSV	0.012	0.486
	DAffSV-DObsSV	-0.264	0.149
<i>D. obscura</i>	<b>DObsSV</b> -DAffSV	3.219	6.184
	<b>DObsSV</b> -DMelSV	2.786	5.817
	DAffSV-DMelSV	-0.996	0.102

**Table S1. Credible intervals of the differences between estimates of genetic variation in susceptibility across host species and viruses.** The natural virus for each host is in red and bold. 95% CIs show differences in estimates of genetic variation for different host-virus combinations, intervals that do not cross zero represent statistically significant differences.

Host	Virus	Mean	95% CIs	
<i>D.affinis</i>	<b>DAffSV</b>	1.47	1.12	1.92
<i>D.affinis</i>	DImmSV	0.18	0.05	0.32
<i>D.affinis</i>	DMelSV	0.48	0.21	0.77
<i>D.immigrans</i>	<b>DImmSV</b>	1.08	0.81	1.39
<i>D.immigrans</i>	DMelSV	0.11	0.00	0.27
<i>D.immigrans</i>	DObsSV	0.21	0.00	0.41
<i>D.melanogaster</i>	<b>DMelSV</b>	0.55	0.39	0.73
<i>D.melanogaster</i>	DAffSV	0.28	0.16	0.39
<i>D.melanogaster</i>	DObsSV	0.33	0.15	0.48
<i>D.obscura</i>	<b>DObsSV</b>	5.84	4.39	7.23
<i>D.obscura</i>	DAffSV	1.07	0.75	1.44
<i>D.obscura</i>	DMelSV	1.52	1.01	1.98

**Table S2. Estimates of genetic variation for each host virus combination.** The natural virus for each host is in red and bold.

736

Host	Viruses	Mean	95% CI
<i>D. affinis</i>	<b>DAffSV</b> -DImmSV	0.50	-0.01, 0.92
	<b>DAffSV</b> -DMelSV	0.05	-0.54, 0.52
	DMelSV-DImmSV	0.02	-0.62, 0.57
<i>D. immigrans</i>	<b>DImmSV</b> -DMelSV	-0.32	-0.91, 0.33
	<b>DImmSV</b> -DObsSV	-0.22	-0.70, 0.26
	DMelSV-DObsSV	0.29	-0.43, 0.94
<i>D. melanogaster</i>	<b>DMelSV</b> -DAffSV	0.40	0.20, 0.61
	<b>DMelSV</b> -DObsSV	0.25	-0.02, 0.48
	DAffSV-DObsSV	0.36	0.10, 0.62
<i>D. obscura</i>	<b>DObsSV</b> -DAffSV	0.21	-0.03, 0.46
	<b>DObsSV</b> -DMelSV	0.40	0.16, 0.63
	DAffSV-DMelSV	0.68	0.49, 0.84

737 **Table S3. Genetic correlations ( $r_g$ ) in viral loads across host species after infection by**  
738 **coevolved and non-coevolved viruses** The coevolved (natural) virus for each host is in red  
739 and bold. Models ran using REML gave similar estimates.

740  
741

Primer	Sequence (5'-3')	Primer information
Ge-1_Indel_1F Ge-1_Indel_1R	AGCGTCAAGCTTTTCCTTCA CACCAGCGGTCAGGATAGAT	<i>Ge-1</i> genotyping primers. Product with susceptible allele is 155bp and 77bp with resistance allele.
Ref2p_P1_F Ref2p_PS1_R Ref_a1_F Ref_a1_R	CTCACCAGCTGCACTTGTA TGTTGCAATCTTTGCGACTC GGATGCCCTCCCAGAATTA CGACGCAATRYGGTGTATCC	<i>p62 (Ref(2)P)</i> genotyping primers. Multiplex all four primers. Susceptible allele produces a 235bp product, resistant allele a 289bp product.
CHK_F Doc1420_F CHK_R	CTCTTGCTCCAAACGTGAC CTTGTTACATTGTCGCTGAG AAGGCAAACGACGCTCTT	<i>CHKov-1</i> genotyping primers. Use reverse primer (R) in singleplex with each of the forward (F) primers, if resistant Doc1420_F + CHK_R primers produce 560bp product, if susceptible CHK_F + CHK_R produce 634bp product

742 **Table S4. Primers for genotyping *D. melanogaster* resistance genes *Ge-1*, *p62 (Ref(2)P)* and**  
743 ***CHKov-1*.** PCRs were carried out using a touchdown PCR cycle (95°C 30sec, 62°C (-1°C per  
744 cycle) 30sec, 72°C 1min; for 10x cycles followed by; 95°C 30sec, 52°C 30sec, 72°C 1min; for a  
745 further 25x cycles).

746

**Table S5 QTL and their locations**

QTL	Virus <sup>1</sup>	Chromosome	Position <sup>2</sup>	Map Position	LOD	Confidence Interval <sup>2,3</sup>
X_13	DMelSV	X	5670000	13.3	14.2	X:5530000..5740000
2L_53	DMelSV	2L	19720000	53.7	23.9	2L:19540000..19870000
2R_100	DMelSV	2R	18710000	100.9	11.5	2R:18650000..18770000
3L_46	DMelSV & DObsSV <sup>4</sup>	3L	20080000	46.1	33.5	3L:19990000..20140000
3R_49	DMelSV & DObsSV <sup>5</sup>	3R	5550000	49	21.3	3R:5420000..5930000
3R_64	DMelSV	3R	14960000	64.9	13.4	3R:14900000..15060000
3R_69	DMelSV	3R	16640000	69.6	10.8	3R:16570000..16840000

<sup>1</sup>The virus used to identify the QTL

<sup>2</sup>Release 5 of the *D. melanogaster* genome

<sup>3</sup>Confidence interval on location based on a LOD drop of 2

<sup>4</sup>QTL significant for both DMelSV and DObsSV, but showing location from DObsSV

<sup>5</sup>QTL significant for both DMelSV and DObsSV, but showing location from DMelSV

Primers	Forward	Reverse
D.mel_ <i>RpL32</i>	TGCTAAGCTGTCGCACAAATGG	TGCGCTTGTTTCGATCCGTAAC
D.imm_ <i>RpL32</i>	TGCTAAGTTGTCGCACAAATGG	TACGCTTGTTGGAGCCATAAC
D.aff_ <i>RpL32</i>	TGCCAAGTTGTCGCACAAATGG	TGCGCTTGTTGGAGCCATAAC
D.obs_ <i>RpL32</i>	TGCTAAGTTGTCGCACAAATGG	TGCGCTTGTTGGAACCGTAAC
DMelSV	CCGACTACAAATGCTATATG	CAGGTATTAGAGGCTTCTTA
DImmSV	CAGAGTATTTGATCAAGGAG	TGGCAATACAGATGTTTTG
DAffSV	CCAGGTGACTAACGATTC	GGAAGAGGAGATATTCAAG
DObsSV	GGCATCTGTTACCGTTAA	ACTCGGAATTTTCTGACC

**Table S6. Primers for qRT-PCR (5'-3').**



## Supplementary Methods

All lines were screened for their retrospective sigma virus over two generations by RT-PCR, and infected isofemale lines discarded prior to the experiment.

### *Drosophila melanogaster*

We created an outcrossed population by combining 150 isofemale lines of *D. melanogaster* (collected in Accra, Ghana (5.593, -0.188) in 2014) in a population cage. The population was maintained throughout the experiment with a large population size (~1500-2000 flies), with eggs collected from the population cage used to set up each subsequent generation. All rearing was carried out on cornmeal medium (recipe below) sprinkled with live yeast ('yeasted') at 25°C.

Virgin flies were collected daily from bottles set up at a controlled egg density. Full-sib families were set up using crosses of single male and female virgins placed in the same vial and aged for 3 days. Each of these families was tipped onto fresh food daily for 5 days to create replicate vials. After 5 days the adult flies were frozen for later genotyping. 12 days after laying, male offspring were collected from each replicate vial and split into two vials of cornmeal medium without any yeast on the surface ('unyeasted') and placed at 18°C. After 5 days these flies were injected with 69nl of virus extract intra-abdominally using a Nanoject II micro-injector (Drummond scientific). Injected flies were kept at 18°C and tipped onto fresh unyeasted cornmeal every 5 days, before being homogenised in Trizol (Invitrogen) and frozen at -80°C on day 15 post injection for later RNA extraction and qRT-PCR.

Injections were carried out over 25 overlapping blocks. Each block consisted of 10 families, and each day, 2 vials per family were injected with 2 different viruses. Each replicate vial contained a mean of 14 flies (range 3-28 flies). In total we measured 255 families over 1567 biological replicates. We aimed to carry out a minimum of 2 replicates of each virus per family, but where possible we carried out 3 or 4 replicates (92 virus-family combinations had 1 replicate, 555 had 2 replicates, 82 had 3 replicates and 25 had 4 replicates.). 248 families had replicates for all 3 viruses. Blocks were staggered to overlap with at least 20 families being infected on any one day. The order families were injected in was randomised, and the order the different viruses were injected was blocked across days.

### *D. immigrans*

92 *D. immigrans* lines were collected from Madingley, Cambridge, UK (52.225, 0.043) in 2012 and 2015. Full-sib families were set up using crosses between the 92 isofemale lines of *Drosophila immigrans*. Flies were reared on malt food (recipe below) at 18°C. Crosses were between different isofemale lines (i.e. excluding reciprocal crosses) and maximising the number of lines used. Families were established from 2-4 day old single female and male virgin flies placed in the same vial for 7 days. These crosses were tipped onto fresh food every 7 days to generate replicate vials of each family. Eclosed males were collected 27-34



days after initial egg laying and injected with DImmSV, DMelSV or DObsSV 1-3 days post-collection, then maintained and frozen on day 15 post-infection as above.

Injections were carried out over 18 overlapping blocks. Each block consisted of an average of 19 families and 46 replicate vials. Each replicate vial contained a mean of 14 flies (range: 4-26). In total we assayed 341 families over 812 biological replicates. We aimed to have a minimum of 2 replicates per virus per family (235 virus-family combinations had 1 replicate, 270 had 2 replicates, 11 had 3 replicates and 1 had 4 replicates). 140 families had replicates across 2 different viruses and 18 families had replicates for all 3 viruses. Blocks were staggered to overlap with a mean of 39 families being infected on any one day. The order families were injected in was randomised, and the order the different viruses were injected was blocked across days.

# *D. affinis*

Site	N
Athens, Georgia, USA, (33.946, -83.384) in 2012	13
Great Smokey Mountain National Park, Gatlinburg, USA (35.698, -83.613) in 2015	23
Rochester, New York, USA (43.135, -77.599) in 2012	4

Full-sib families were set up using crosses between 40 isofemale lines of *Drosophila affinis* (see above for collection details) collected in the U.S. Flies were reared on malt food (recipe below) at 18°C. Crosses were between different isofemale lines (i.e. excluding reciprocal crosses) and maximising the number of lines used. Families were established from 6 day old single female and male virgin flies placed in the same vial for 7 days. These crosses were tipped onto fresh food every 7 days to generate replicate vials of each family. Eclosed males were collected 35-42 days after initial egg laying and then injected with DAffSV, DImmSV or DMelSV 1-3 days post-collection, then maintained and frozen on day 15 post-infection as above.

Injections were carried out over 27 overlapping blocks. Each block consisted of an average of 19 families and 28 replicate vials. Each replicate vial contained a mean of 11 flies (range: 3-23). In total we assayed 520 families over 1003 biological replicates. We aimed to have a minimum of 2 replicates per virus per family (336 virus-family combinations had 1 replicate, 286 had 2 replicates and 30 had 3 replicates). 109 families had replicates across 2 different viruses and 12 families had replicates for all 3 viruses. Blocks were staggered to overlap with a mean of 23 families being infected on any one day. The order families were injected in was randomised, and the order the different viruses were injected was blocked across days.

839 ***D. obscura***  
840

Site	N
Derbyshire Site A , UK (52.978, -1.440) in 2012	4
Derbyshire Site C, UK (52.903, -1.374), in 2012	1
Les Gorges du Chambon, France (45.622, 0.555) in 2012	1
Madingley, Cambridge, UK (52.226, 0.046) in 2014	15

841  
842 *D. obscura* were collected in the United Kingdom and France (see above). Males and females  
843 were separated, and females were placed in vials to establish isofemale lines. Full-sib  
844 families were set up using crosses between 21 isofemale lines of *Drosophila obscura*  
845 collected in the UK. Flies were reared on banana food (recipe below) at 18°C. Crosses were  
846 between different isofemale lines (i.e. excluding reciprocal crosses) and maximising the  
847 number of lines used. Families were established from 6 day old single female and male virgin  
848 flies placed in the same vial for 7 days. These crosses were tipped onto fresh food every 7  
849 days to generate replicate vials of each family. Eclosed males were collected 35-42 days  
850 after initial egg laying and then injected with DAffSV, DMelSV, or DObsSV 1-3 days post-  
851 collection, then maintained and frozen on day 15 post-infection as above.

852  
853 Injections were carried out over 25 overlapping blocks. Each block consisted of a mean of 16  
854 families with a mean of 76 vials being injected each day for 12 days. Each replicate vial  
855 contained a mean of 8 flies (range: 1-15). In total we assayed 320 families over 913  
856 biological replicates. We aimed to have a minimum of 2 replicates per virus per family (126  
857 virus-family combinations had 1 replicate, 314 had 2 replicates and 49 had 3 replicates and 3  
858 had 4 replicates). 94 families had replicates across 2 different viruses and 39 families had  
859 replicates for all 3 viruses. Blocks were staggered to overlap with at least 10 families being  
860 infected on any one day. The order families were injected in was randomised, and the order  
861 the different viruses were injected was blocked across days.

862  
863 **Sample size estimation**  
864

865 The number of full-sib families required for estimating genetic variance in susceptibility was  
866 determined by simulation using previous estimates of genetic variation to DMelSV in *D.*  
867 *melanogaster* [11]. After carrying out the full-sib experiment in *D. melanogaster*, we then  
868 down-sampled this data to calculate the minimum number of families required to provide  
869 accurate estimates for the other species. Sample sizes for the DSPR experiment were based  
870 on previous data [56].

871  
872  
873  
874

# **Food recipes**

876

877 Banana:

878

879 Mixture 1

880 1000ml water

881 30g yeast

882 10g agar

883

884 Mixture 2

885 20ml Nipagin

886 150g pureed banana

887 50g corn syrup

888 30g malt powder

889

890 Bring mixture 1 to the boil for 3-4 minutes, whisk constantly. Add to Mixture 2 to Mixture 1.

891 Whisk constantly and simmer for 5 minutes.

892

893 Cornmeal:

894

895 1200ml water

896 13g agar

897 105g dextrose

898 105g maize

899 23g yeast

900

901 Combine and bring to a boil for 5mins, cool to 70°C before adding 35ml Nipagin (10%)

902

903 Malt:

904

905 1000ml water

906 10g agar

907 60g semolina

908 20g yeast

909 80g malt extract

910

911 Combine and bring to a boil for 5mins, cool to 70°C and then add 14ml Nipagin (10%) and

912 5ml proprionic acid.

913

914

915

916

917

918