The evolution and genetics of *Drosophila melanogaster* and the sigma virus

Jennifer A Carpenter

PhD
The University of Edinburgh
2008

For my mother, who never doubted I would read (eventually); and my father, for inspiring me to study biology

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text.

This work has not been submitted for any other degree or professional qualification except as specified.

Jennifer A Carpenter, August 3, 2009.

Acknowledgments

This work would not have been possible without the help and encouragement of a large number of people. I am grateful to the following:

Frank Jiggins for his supervision, encouragement and useful comments on manuscripts and thesis chapters, for his unyielding patience and for always, without exception, having his office door open to visitors with questions. Tom Little for additional supervision, invaluable discussion and for putting up with me warbling on about flies. Graham Stone for guidance and a stiff drink when I needed one. Mark Blaxter for allowing me to work in his lab for three years. Marta Wayne, Sergey Nuzhdin, Peter Christian, Dieter Contamine, Ian Gordon, and members of their groups for letting me work in their laboratories during fieldwork. Katelyn Fenn for my molecular training and encouragement throughout. Jarrod Hadfield, Matthew Robinson and Ian White for statistical analysis. Darren Obbard for helpful discussions. Dan Halligan and Jonathan Bollback for the LaTeX 2_ε used to typeset this thesis and Kim Bissix for help creating images. Kang-Wook Kim, Claire Colon, Robert Young, Mike Magwire, Lena Bayer-Wilfert and Claire Webster for their help with lab work. Penny Haddrill and Kelly Dyer for help in the lab when experimental work grew unmanageable and for their constant encouragement throughout. Andrea Betancourt for illuminating discussion about evolution and population genetics and for her back of the serviette diagrams. Natasa Fytrou for her help with fly work late into the night and for her laughter in the lab. Jenny Bangham for teaching me how to see the good in a failed experiment, for her patient editing and inspiring me to look beyond my own work. Thanks to the British Broadcasting Corporation, Canadian Broadcasting Corporation and National Public Radio, especially This American Life, Radio Lab and The Archers—this body of work would not have been completed, if not for them.

Thanks to my friends in Edinburgh, especially Sarah Bangham, Helen McElhinney,

Susan Williamson, Andy Gardener, Zeynep Arman and Tabitha Innocent for giving me perspective and friends further away in other labs, Richard Naylor and Jonathan Linklater, thanks for their understanding and mutual cursing of midnight virgin collections, and thanks to Cynthia Telfer and Akivah Starkman for their support from the conception of the project to the last typos.

Finally thanks to Elie Dolgin for his sandwich making and companionship and to my family, especially my parents, whose unquestioning belief in me has carried me through many moments of doubt.

This work was supported by a Natural Environment Research council postgraduate studentship number NER/S/A/2004/12278.

Abstract

Parasites shape the evolution of their hosts and hosts shape the evolution of their parasites. Understanding how these reciprocal selective forces drive evolutionary dynamics is crucial in the fight against infectious disease, and requires knowledge of the genetics of both the host organisms and parasites. In this thesis, I use phylogenetic, molecular, and quantitative genetic techniques to explore the impact of coevolutionary genetic change using the fruit fly, *Drosophila melanogaster*, and the sigma virus—a negative sense RNA virus that occurs in natural *D. melanogaster* populations throughout the world.

Following a general introduction in Chapter 1 and a description of the general materials and methods used throughout this work (Chapter 2), I describe, in Chapter 3, the isolation and characterisation of new sigma-viral isolates collected from Europe and North America. With these new isolates, I show that the sigma virus has very low levels of viral genetic diversity across Europe and North America compared to other RNA viruses. Based on laboratory measurements of the viral substitution rate, I suggest that most European and North American viral isolates shared a common ancestor approximately 200 years ago, and offer two possible explanations for this: the first is that *D. melanogaster* has recently acquired the sigma virus; the second is that a single viral type has recently swept through *D. melanogaster* populations. I go on to examine the population structure of these new viral isolates and find that in contrast to *Drosophila* populations, the sigma viral populations are highly structured. This is surprising for a vertically transmitted pathogen that has a similar migration rate to its host. I suggest that the low structure in the viral populations can be explained by the smaller effective population size of the virus.

In Chapter 4, I examine the susceptibility of *D. melanogaster* to five of the viral isolates described in Chapter 3 to investigate whether specificity exists in this system, and if it does, whether it generates tradeoffs between resistance against different pathogen genotypes, thereby maintaining variation. To investigate this, I measured the transmission rate of five viral isolates in fly lines with different first, second and

third chromosomes. I found significant genetic variation in resistance against all five viruses on each of the three different chromosomes. Most of this resistance is general: acting equally against all five viruses. This result suggests that there is little constraint on flies evolving resistance to all five viruses, and that trade-offs between resistance against the five viruses is unlikely to explain why variation in susceptibility exists in wild populations of *D. melanogaster*.

Bacterial and fungal infections induce a potent immune response in *D. melanogaster*, but it is unclear whether viral infections induce a similar immune response. In Chapter 5, I investigate *D. melanogaster's* immune response against the sigma virus. In the first experiment I show that sigma-viruses increases the susceptibility of flies to *Beauveria bassiana*—a fungus that commonly infects insects in the wild. This could have profound effects in the wild where flies are constantly exposed to bacteria and fungus during feeding. One interpretation of the increased susceptibility of sigma-infected flies, is that the sigma virus is suppressing the Toll-pathway—an important component of the innate immune system that is involved in immune defences against fungal infections. However, I found no evidence for viral suppression of the Toll-pathway, nor did I find evidence that flies mount a Toll-dependent immune response against the sigma virus. This suggests that either *Drosophila* do not mount an immune response against the sigma virus, or that the immune response is controlled by other pathways.

Finally, in Chapter 6, I describe the hypermutation of adenosines to guanosines in the genome of the sigma virus. The clustering of these mutations, and the context in which they occur, indicates that they have been caused by ADAR—RNA editing enzymes that target double stranded RNA. However, ADAR editing of viral RNA is either rare or edited viral RNA are rapidly degraded, as I only detected evidence for editing in one of infected viral strains I studied. This is the first evidence that ADARs target viruses outside of mammals, and it raises the possibility that ADARs could play a role in the antiviral defences of insects.

Contents

Abbreviations **1**

1 Introduction **2**

1.1 Importance of host-parasite coevolution . . . 2
1.2 Host-parasite coevolution and the maintenance of genetic variation . . 3
1.3 *Drosophila* viruses . . . 6
1.4 The sigma virus . . . 7
1.4.1 Virion and genome structure . . . 8
1.4.2 The CO_2 symptom . . . 9
1.4.3 Virus transmission . . . 9
1.4.4 Prevalence and host range . . . 10
1.4.5 The genetics of resistance to the sigma virus . . . 10
1.4.6 *ref(2)P* . . . 11
1.4.7 *ref(2)P* alleles . . . 12
1.4.8 Evolutionary genetics of resistance to the sigma virus . . . 14
1.5 Aims of this study . . . 16

2 General materials & methods **18**

2.1 RNA extraction techniques . . . 18
2.1.1 Isolation of RNA from *Drosophila* with kit . . . 18
2.1.2 Isolation of RNA from *Drosophila* with Trizol® . . . 19
2.2 RNA sequencing techniques . . . 20
2.2.1 Reverse transcription . . . 20
2.2.2 Standard polymerase chain reaction (PCR) . . . 21
2.2.3 Agarose gel electrophoresis . . . 22
2.2.4 EXOSAP cleanup . . . 22
2.2.5 Sequencing reaction . . . 22
2.2.6 Pyrosequencing . . . 23
2.2.7 Cloning . . . 24

2.2.8 Fly culturing techniques 24
2.2.9 Fly culturing techniques 24

3 Sigma virus phylogenetics 26
3.1 Introduction 26
3.2 Materials and methods 27
3.2.1 Collection 27
3.2.2 Sequencing and sequence analysis 28
3.2.3 Estimation of mutation rates and phylogenetic reconstruction . 28
3.3 Results 30
3.3.1 Viral prevalence 30
3.3.2 Recombination 32
3.3.3 Viral sequence variation and population structure 32
3.3.4 Estimating substitution rate 34
3.3.5 Age of viral spread 35
3.4 Discussion 35
3.4.1 Low viral diversity 35
3.4.2 Population structure 37

4 Genotype-by-genotype interactions between *Drosophila melanogaster* and the sigma virus 40
4.1 Introduction 40
4.2 Material and methods 42
4.2.1 Experimental overview 42
4.2.2 Stocks, viral isolates and general methods 42
4.2.3 Experimental methods 44
4.2.4 Experiment measuring transmission in second and third chromosome-substitution lines 46
4.2.5 Experiment measuring transmission in first chromosome-substitution lines 47
4.3 Statistical anaylsis 48
4.3.1 Estimating variance and covariance 50
4.3.2 Eigenanaylsis 51
4.4 Results 52

4.4.1 Genetic variance in maternal transmission to homozygous offspring . . . 52
4.4.2 Genetic variance in maternal transmission to hemizygous offspring . . . 52
4.4.3 Genetic variance in paternal transmission to heterozygous offspring . . . 52
4.4.4 Genetic covariance between viruses in rates of transmission . 54
4.4.5 Response to selection on each of the 5 viruses . . . 58
4.4.6 The effect of *ref2p* on paternal transmission . . . 64
4.5 Discussion . . . 64
4.5.1 Transmission rates of viruses . . . 64
4.5.2 Patterns of genetic variation . . . 65
4.5.3 What is maintaining variation in resistance against the sigma virus? . . . 66

5 The antiviral role of the Toll pathway against sigma 68
5.1 Introduction . . . 68
5.2 Materials and methods . . . 71
5.2.1 Infection methods . . . 71
5.2.2 Measuring susceptibility of sigma-infected flies to fungal infection . . . 72
5.2.3 Measuring the suppression of Toll-pathway in sigma-infected flies . . . 73
5.2.4 Measuring susceptibility to sigma infection in flies without a functional toll-pathway . . . 75
5.3 Results . . . 76
5.3.1 Fungal infection experiment . . . 76
5.3.2 Activation of the Toll-pathway experiment . . . 81
5.3.3 Toll-pathway mutant experiment . . . 83
5.4 Discussion . . . 84
5.4.1 Sigma-induced susceptibility to fungal infection . . . 85
5.4.2 No activation of the Toll-pathway by sigma . . . 85
5.4.3 Toll-pathway not involved in an antiviral response against sigma virus . . . 86

6 ADAR-induced hypermutation in the sigma virus 89
6.1 Introduction . . . 89
6.2 Materials and methods . . . 91
6.2.1 Detecting hypermutation in the sigma virus . . . 91
6.2.2 Detecting suppression of ADAR-editing by the sigma virus . . . 92
6.2.3 Statistical analysis . . . 93
6.3 Results . . . 94
6.3.1 Evidence for hypermutation . . . 94
6.3.2 No evidence for ADAR-editing in cloned virus . . . 96
6.3.3 No evidence for suppression of ADAR-editing by viruses . . . 97
6.4 Discussion . . . 98
6.4.1 Evidence for ADAR-editing . . . 98
6.4.2 No evidence for suppression of ADAR-editing . . . 101

7 Discussion and conclusions 102
7.1 Summary . . . 102
7.2 Conclusions . . . 105
7.3 Future directions . . . 107

Bibliography 109

A Appendix 122

B Appendix 125
B.1 Is the sigma virus processed by RNAi? . . . 125
B.2 Introduction . . . 125
B.3 Methods and materials . . . 126
B.3.1 RNA extraction . . . 126
B.3.2 siRNA analysis . . . 126
B.4 Results . . . 127

C Publications 129

Abbreviations

List of commonly used abbreviations:

bp	Base Pairs
dATP	Deoxyadenosine 5-triphosphate
dCTP	Deoxycytosine 5-triphosphate
dGTP	Deoxyguanosine 5-triphosphate
DNA	Deoxyribose Nucleic Acid
dTTP	Deoxythymidine 5-triphosphate
kb	Kilobase
μl	Microlitre
ml	Millilitres
mRNA	Messenger RibNucleic Acid
PCR	Polymerase Chain Reaction
roH_2O	Reverse osmosis sterile H_2O
SD	Standard Deviation
SE	Standard Error
Taq	*Thermus aquaticus* polymerase

All other abbreviations are for chemical formulae or are detailed in the main text.

1 Introduction

1.1 Importance of host-parasite coevolution

Pathogens affect host survival and reproduction, and so have the potential to drive the evolution of many host traits. This, in turn, puts pathogens at the heart of many biological phenomena, potentially influencing host population dynamics, phylogenetic patterning, speciation and the evolution of sex (Haldane 1949, 1954). Understanding the reciprocal nature of host-parasite coevolution will help us to recognise the damaging effects of the diseases caused by pathogens, and the impact of disease on the genetics of host populations. What's more, parasitism is thought to be at the heart of many lingering problems in evolutionary biology: why there is considerable genetic variation in susceptibility to pathogens; how many genes underlie this variation, which genes are responsible, how big are the effects of these genes and whether these genes offer general or specific resistance to pathogens?

Most of the work into the genetics of host-parasite coevolution has focused on understanding the molecular mechanisms of host resistance and the genes underlying this resistance (Lazzaro *et al.* 2004, Tinsley *et al.* 2006). This work has resulted in a better understanding of the mechanisms and pathways involved in invertebrate immunity, and along with a comparative approach between vertebrate and invertebrate innate immune systems, has increased our understanding of the genes and mechanisms involved in human immunity. However, this approach has largely ignored the role of the pathogen's genetics. Until recently, research into the invertebrate immune system has mostly involved challenging the insect immune system by introducing non-coevolved pathogens directly into the body cavity (Siva-Jothy *et al.* 2001). Resistance to coevolved pathogens entering through natural routes may be missed with such an approach. Despite this, few studies have attempted to characterise the relationships between coevolved pathogens and their hosts (Carius *et al.* 2001, Dybdahl & Lively 1998, Ferrari *et al.* 2001). Still fewer studies have tackled the genetic underpinnings of these relationships (Lazzaro *et al.* 2004, Kraaijeveld & van Alphen 1995, Tinsley

et al. 2006).

Carefully chosen model systems could allow both halves of the co-evolutionary interaction to be studied simultaneously. *Drosophila* and their viruses offer a particularly tractable genetic model; viruses occur naturally, are often pathogenic and can be easy to assay. And new research at the molecular level has led to the discovery of novel strategies used by insect hosts to protect themselves against viral infections (Ding & Voinnet 2007, Zambon *et al.* 2005). By understanding better the antiviral mechanisms of *Drosophila* and the viruses that naturally infect them, we can piece together both halves of the host-parasite co-evolutionary interaction to understand the genetics of both host and parasite, which has implications for the study of epidemiology, immunology and evolutionary biology.

1.2 Host-parasite coevolution and the maintenance of genetic variation

Many studies have found genetic variation affecting disease resistance in wild populations. Evidence for this genetic variation has been found in a wide range of taxa: from humans (Hirschhorn & Daly 2005) to plants (Holub 2001, Burdon 1987, Chaboudez & Burdon 1995, Dinoor 1977), and invertebrates (Carius *et al.* 2001, Dybdahl & Lively 1998, Ferrari *et al.* 2001, Henter 1995, Henter & Via 1995, Kraaijeveld & van Alphen 1995, Kraaijeveld & Godfray 1997, Lazzaro *et al.* 2004, Riehle *et al.* 2006, Tinsley *et al.* 2006). However, rather few studies have formally estimated how much genetic variation exists in natural populations for resistance to pathogens (Frank 1994, Bergelson *et al.* 2001).

Understanding how much genetic variation exists in wild populations and why it exists has important implications for managing and treating diseases: genetic variation affects whether disease vectors transmit pathogens to humans, and how populations respond to disease (Gooding 1996). Genetic variation in disease resistance is also important for agriculture: genetic variation affects how crops and livestock respond to disease, and how they respond to selection during breeding programs. Furthermore, by understanding how many genes underlie variation, how big an effect each gene has on an individual's resistance and which genes are responsible, we should also begin to understand the molecular mechanisms that make one individual more susceptible

to pathogens than another.

In addition to these applications, understanding why genetic variation is maintained is also crucial for understanding the evolution of sex (Bell & Smith 1987). Haldane (1949) suggested that parasites may ultimately be responsible for the maintenance of the large amounts of genotypic variation that we observe in natural populations (and which otherwise would be quickly eroded by natural selection). If they are responsible, sexual reproduction may have been favoured because it creates this variation, explaining why sexual reproduction is maintained over clonal alternatives (Haldane 1949, 1954).

By finding out why variation in resistance to pathogens is maintained we can gain insights into the models of coevolution. A number of models have been proposed to explain this variation. The first class of models suggest that variation could be transient, and exist because a selective sweep is in progress. Under this scenario, variation is maintained because although these alleles confer resistance, they never reach fixation because the pathogen—the target of their resistance—is continually evolving, and so the direction in which selection is acting is continuously shifting. Under this model, genes involved in the immune system are expected to evolve rapidly and show evidence of natural selection fixing large numbers of amino acid substitutions. There is evidence for this in vertebrates (Hughes & Yeager 1997), and in many components of the invertebrate immune system (pathogen-recognition proteins, signal transduction proteins, or antimicrobial peptides), that evolve faster than the genome as a whole, and show evidence of natural selection (Jiggins & Hurst 2003, Jiggins & Kim 2005, Schlenke & Begun 2003). The most exceptional of these, are genes involved in an antiviral RNAi function (Dcr2, R2D2, and Ago2), which are among the fastest evolving 3% of all *Drosophila* genes (Obbard *et al.* 2006).

Although it is clear from these studies that directional selection is common, it is not known whether this contributes to genetic variation in the population. This is because selective sweeps happen fast, so the polymorphisms may be short lived. One of the few cases showing that a resistance polymorphism has resulted from a partial selective sweep is in the gene ref(2)P (Bangham *et al.* 2007).

The second class of models suggest that variation is maintained by negative frequency-dependent selection. Under this scenario, the fitness of a genotype is dependent on its frequency relative to other genotypes in the population. So new

resistant alleles increase in frequency in the population as long as resistance is beneficial. But over time, the frequency of the pathogen will decline such that, eventually, this pathogen is so rare that resistance against it is no longer advantageous. After the resistant alleles cease to be beneficial, they are expected to remain at high frequencies or drift to fixation, except if they are costly.

Costs associated with resistance result in selection against these resistant alleles, preventing them from going to fixation and maintaining variation. Costs can exist either as a trade-off between resistance against different pathogen genotypes or as a trade-off with other components of fitness.

Two influential population genetic models have described how the different costs of resistance maintain variation in host susceptibility in natural populations. The first set of models propose that each host genotype is better than other genotypes at resisting a particular pathogen genotype, but is worse at resisting other pathogen genotypes. This is because trade-offs exist between resistance against different pathogen genotypes, and prevent any one host genotype resisting everything—known as 'matching-allele concept' (Agrawal & Lively 2002, Howard & Lively 1994). The second set of models proposes that some host genotypes are intrinsically better than others at resisting pathogens but this resistance is costly and these costs prevent these genotypes from going to fixation—the so-called 'gene-for-gene concept' (Flor 1955, Agrawal & Lively 2002).

Evidence for gene-for-gene and matching-allele models exists, (gene-for-gene: Flor (1955), McVey (1990), Webster *et al.* (1986), and matching-allele: Carius *et al.* (2001), Lambrechts *et al.* (2005), Salvaudon *et al.* (2007)). However, solely based on observed polymorphisms, it is difficult to infer whether host-pathogen genetics follow strict gene-for-gene or matching-allele models (Frank 1996). Measuring the strength of the trade-offs (as I do in Chapter 4) is important, because, if the trade-offs between resistance against different pathogen genotypes are small enough—even if they result in specific interactions between host and parasite—they are unlikely to prevent the evolution of general resistance, and so are unlikely to maintain the variation that we see in natural populations. If they are large, however, most of the genetic variation affecting susceptibility to pathogens will be specific to a pathogen genotype, and so there is little potential for the population to evolve general resistance to all pathogen genotypes over time.

1.3 *Drosophila* viruses

Early surveys using electron microscopy and DNA hybridization show that 30-40% of *D. melanogaster* populations are infected with at-least one virus (Brun & Plus 1998, Christian 1987). However, few studies have investigated the distribution and prevalence of these viruses in natural populations of *Drosophila* (Carpenter *et al.* 2007, Christian 1987). In total, seven viruses have been isolated from wild populations of *Drosophila*; *Drosophila* viruses C, A and the sigma virus are the most common, having been isolated from more than ten different geographical locations (Brun & Plus 1998, Christian 1987). *Drosophila* viruses P and F have each been found in the wild on at least three occasions, while viruses C, A, P and Nora have also been recovered repeatedly from laboratory strains (Brun & Plus 1998, Christian 1987, Habayeb *et al.* 2006, 2007). Most of these viruses, which are picorna-like viruses (Christian 1987, Plus & Duthoit 1969), except for one Reovirus (F) (Plus *et al.* 1975), and one Rhabdovirus (sigma) (Fleuriet 1976*b*). Although the taxonomic relationships between *Drosophila* viruses remain unresolved, these viruses seem to be both phylogenetically and biological diverse.

Understanding how common these viruses are and whether they are host-specific, will lead to greater understanding of the type and strength of selection pressures that viruses exert on their *Drosophila* hosts. Specialist parasites, unlike generalists, are likely to be involved in a tight co-evolutionary arm-race with the host's immune system, and so be an important driving force in the co-evolutionary adaptation of the innate immune system. However, despite the importance of specialist viruses, few studies have examined the host-specificity of *Drosophila* viruses.

The most well-studied virus isolated in *Drosophila* is the sigma-virus—a naturally occurring parasite of *D. melanogaster*. The *Drosophila*-sigma system offers a tractable model system for studying the evolutionary interactions between *Drosophila* and its coevolved parasite. Parasitoid wasps are the only other coevolved parasite of *Drosophila* that have been studied in detail (Kraaijeveld *et al.* 1998).

In the following sections I will describe the virology of the sigma virus, including the structure and organisation of its genome, how the virus is transmitted through its host and how it causes CO_2 sensitivity. I will then describe what is known about the loci involved in the fly's resistance against the sigma virus, concentrating on *ref(2)P*—

a polymorphic gene that has a major effect on the fly-sigma interactions. Finally, I will discuss the evolutionary genetics of the resistance against the sigma virus and examine the evidence that suggests that *ref(2)P* is under selection.

1.4 The sigma virus

The sigma virus is a rhabdovirus that commonly infects wild populations of fruit flies and is transmitted vertically, from parent to offspring. Sigma virus first caught the interest of biologists in 1937 when a study by P. L'Heritier and G. Tessier described how a line of flies collected in France were irreversibly paralyzed when exposed to a CO_2, which is commonly used as an anesthetic in fly genetics. This CO_2-induced paralysis was shown to be heritable, and at first it was thought that this trait was transmitted like a chromosome. Believing that they had discovered another heritable element, the researchers called it a 'genoide'. However, in subsequent studies, L'Heritier and Tessier discovered that this trait could be transmitted by injection and once injected into a fly, was passed to offspring through both males and females. It soon became clear that the 'genoide' was in fact a virus, and was renamed sigma.

A number of studies went on to characterize the CO_2 sensitivity caused by sigma virus infection. They found that, whereas uninfected flies recover rapidly from the effects of CO_2 exposure upon return to a normal atmosphere, flies infected with sigma virus remain irreversibly paralyzed and eventually die. They described how this paralysis was specific to CO_2, and sensitive to changes in the gas concentration and temperature. For example, paralysis appears at 10°C with CO_2 concentrations greater than 50%, while CO_2 concentrations must exceed 75% to induce paralysis at 16°C (L'Heritier 1948).

This CO_2 sensitivity has been observed among both wild flies and laboratory strains of flies, and has been shown in several *Drosophila* species (Brun & Plus 1998, Williamson 1961). In three of these species, *D. melanogaster, D. affinis*, and *D. athabasca*, the narcotic effect of CO_2 has been shown to be the consequence of sigma virus infection (Williamson 1961).

1.4.1 Virion and genome structure

Sigma virus is a negatively sense, single-stranded RNA virus that is encapsulated into a bullet-shaped particle that is approximately 75 by 140–200 nm. The sigma virions closely resemble the virions of other rhabdoviruses (Richardmolard *et al.* 1984), and phylogenetic analyses, based on the polymerase gene, indicate that sigma is clustered with vesiculoviruses (Hogenhout *et al.* 2003).

The sigma virus contains six genes, five of which are arranged in the same order as other rhabdoviruses ($3'-N-P-PP3-M-G-L-5'$) with the additional sixth gene—*PP3*, which encodes reverse transcriptase—between the *P* gene and the *M* gene. This sixth gene is found in a small number of rhabdovirues (haematopoitetic necrosis virus, IHNV (Kurath *et al.* 1985); Flanders virus of birds (Boyd & Whitakerdowling 1988); and plant sonchus yellow net virus, SYNV, (Heaton *et al.* 1989)), however it is more similar to the reverse transcriptases of DNA viruses, such as retroviruses and retrotransposons, than it is to the reverse transcriptases of other RNA viruses in the Mononegavirales group (Landesdevauchelle *et al.* 1995). *Drosophila* contains many retroviruses and retrotransposons (Kim *et al.* 1994) and the presence of the *PP3* gene in sigma virus could result from assortment between a retrovirus and the sigma virus.

Five of sigma's six genes have been sequenced, while the sequence of the polymerase gene (*L*), remains largely unknown. The remaining genes, *N, P, PP3, M* and *G*, were identified either by their sequence identity to the genes of other rhabdoviruses or by their structural similarities to rhabdovirus proteins.

Sigma's *N* gene encodes the nucleocapsid and is most similar to vesoculo- and lyssa-virues (Bras *et al.* 1994). The *P* gene encodes the polymerase-associated *P* protein, and is the most variable of all the Mononegavirales proteins, and so there is little sequence similarity between the sigma virus and other rhabdoviruses for this gene. Even so, sigma's *P* protein exhibits enough similar acidity and charge distribution to *P* proteins of other rhabdoviruses to confirm its role as a polymerase-associated protein. The *M* protein is involved in assembling the nucleocapsid in a tightly coiled structure, inhibiting transcription and interacting with cellular factors required for efficient viral synthesis (Bras *et al.* 1994). Like, the *P* protein, the charge and size of sigma's *M* protein is conserved among *M* proteins of vesiculoviruses, and its main domains (basic domain, proline-rich domain, hydrophobic domain) are found in *M* proteins of other rhabdoviruses. Sigma's *G* gene encodes the glycoprotein

that is involved in the budding of virions, and mediates fusion between the virus and target membranes, allowing virion adsorption and release of nucleocapsid into the cell. Sigma's *G* gene is most similar to the glycoproteins of vesiculoviruses, sharing ~20% of its amino acids, compared to only ~13% shared with the rabies virus (Teninges & Brasherreng 1987).

1.4.2 The CO_2 symptom

As mentioned above, sigma virus causes flies to become paralysed when exposed to CO_2. CO_2-induced paralysis is correlated with the presence of sigma virus in the nerve ganglia, and the disruption to nerve ganglia occurs only when the concentration of viral particles exceeded a threshold (Teninges & Brasherreng 1987). It is likely that when sigma virus is injected into the fly's abdominal wall, the virus undergoes several cycles of replication before it is at high enough concentration within the nerve ganglia to cause paralysis of the fly (L'Heritier 1948). This would account for the 15 day delay in the expression of the CO_2-induced paralysis in flies injected with sigma extract. Either increasing the temperature that flies are maintained at, or increasing the concentration of the viral dose, can reduce this delay. CO_2-induced paralysis occurs as quickly as three days post injection, if the virus is injected directly into the nerve ganglia.

Interestingly, other rhabdoviruses also cause paralysis when injected into *Drosophila* (for example, vesicular stomatitis virus, Chandirpura and Piry viruses (Bussereau 1973, 1975); for a more complete list see Teninges & Brasherreng (1987)).

1.4.3 Virus transmission

The sigma virus is transmitted only vertically, from parent to offspring in the cytoplasm of the gametes (Fleuriet 1988). There is no evidence that the sigma virus can be transmitted horizontally, either by direct contact between insects, or vectored through parasitoids or mites. The sigma virus is transmitted at a high rate through females (who usually pass it to 100% of offspring), and at a lower rate through males (who pass it to between 0% and 100% of offspring). It is likely that transmission through males is, on average, lower than through females, because male gametes contain less cytoplasm and so act as a poorer vector for the virus between generations.

Variation in paternal transmission is crucial to whether the virus can invade and maintain itself in a population (Fleuriet 1991). This is because sigma virus infection is harmful to flies, reducing both egg viability and survival overwinter (Fleuriet 1981). Therefore, if the sigma virus was transmitted solely through females, it would be lost from the population, and so a sufficient rate of paternal transmission is required to overcome the costs of infection imposed on the host.

1.4.4 Prevalence and host range

Sigma virus is found in natural populations of *D. melanogaster* all over the world, at frequencies of 0%-15% (Carpenter *et al.* 2007). A few studies have shown that sigma is more common in Europe than in North America and Africa, although these findings are based on limited sampling in Africa (Brun & Plus 1998, Carpenter *et al.* 2007). The prevalence of sigma has been found to vary widely between collection sites, even those only a few kilometres apart (Carpenter *et al.* 2007, Felix *et al.* 1971).

Has sigma virus been found in other species? There have been a number of reports of CO_2 sensitivity in North American species of *Drosophila: D. affinis, D. athabasca* and *D. tolteca* (Felix *et al.* 1971, Williamson 1961), suggesting that sigma can infect other species of *Drosophila*. More recently, sigma infection has been verified in *D. affinis* and *D. subobscura* molecularly (B. Longdon, personal communication).

1.4.5 The genetics of resistance to the sigma virus

Up to seven different *D. melangaster* loci have been be shown to affect replication and transmission of the sigma virus; one locus has been approximately mapped to chromosome one, two loci to chromosome two and four loci to chromosome three (Bangham *et al.* 2007, 2008*b*, Gay 1978). One of these loci (located on chromosome three) affects the transmission of the sigma virus from males to their progeny (Gay 1978), while the other loci are known to affect replication of the sigma virus after injection (Bangham *et al.* 2008*a*, Brun & Plus 1998). These loci are polymorphic, and their alleles can be put into one of two categories: alleles that are 'resistant' to the sigma virus and alleles that are 'susceptible' to the sigma virus. The genes (although only one has been precisely mapped) that underlie these loci have been called 'ref'

genes. 'ref' refers to the refractory nature of these genes to sigma infection, the number in parentheses refers the chromosome on which the gene is located, the following letter is the particular name of the gene, and alleles are indicated by an exponent (Gay 1978).

1.4.6 *ref(2)P*

The most extensively studied of these genes is *ref(2)P*, which maps to the left arm of chromosome two of *D. melanogaster* [cytogenic region 37E3-37F3]. Cloning of *ref(2)P* (Contamine *et al.* 1989) and sequencing of one of its alleles (Dezelee *et al.* 1989) shows that *ref(2)P* is a protein encoding gene (3.1 kbp long) that is divided into three exons and is transcribed into two mRNAs of ~2300 and 2400 nucleotides. Males contain roughly equal amounts of both mRNAs, while females contain more of the longer mRNA, which is the only type found in the ovaries (Contamine *et al.* 1989).

ref(2)P encodes a protein that sits within the Toll pathway—an important component of the innate immune system. Although it is not fully known what *ref(2)P* does there, its structural similarity to a scaffold protein—*p62*—that is involved in the mammalian Toll pathway, has focused research to look for an analogous role of *ref(2)P* in the *Drosophila* Toll pathway (Avila *et al.* 2002). This similarity between *ref(2)P* and human *p62* (~24% of its amino acids) is comparable to the homology found between other immune pathway signalling molecules (human *TRAF6* and *Drosophila DTRAF2*, ~31%; and human *MyD88* and *Drosophila DMyD88*, ~22%) (Avila *et al.* 2002). Over expressing *ref(2)P* in cell lines activates a promoter protein, just upstream of Drosomycin—an antimicrobial peptide gene in the Toll pathway. Moreover, depletion of *ref(2)P* in cell lines, leads to a reduction in Drosomycin transcription (Avila *et al.* 2002). *ref(2)P* was not shown to be involved in any other innate immune pathways (Avila *et al.* 2002).

What is *ref(2)P*'s role in the Toll pathway? A number of studies have shown that *ref(2)P* interacts with both *Drosophila* atypical protein kinase C (aPKC) and *Drosophila* tumor necrosis factor receptor-associated factor 2 (dTRAF2) (Avila *et al.* 2002). Both aPKC and dTRAF2 are involved in the Toll pathway; aPKC is an isozyme that activates the NF-κB complex—a protein complex that is a transcription factor—that when active moves across the nuclear membrane to transcribe antimicrobial

peptides, and dTRAF2 is a signalling protein just upstream of NF-κB complex (Sanz *et al.* 1999, Shen *et al.* 2001). Further evidence that these interactions are genuine comes from studies that show *ref(2)P*'s homolog p62 interacts with both with mammalian aPKCs and TRAF-6 (Sanz *et al.* 1999). What *ref(2)P* is doing in these interactions is not known, but it most likely fulfils a similar function to its mammalian homolog, anchoring aPKC and the NF-κB complex together to the intercellular membranes.

In addition to the role of *ref(2)P* in the Toll pathway, *ref(2)P* also affects sperm development. The absence of *ref(2)P* causes the breakdown of mitochondria in the spermatids resulting in non-motile sperm (Contamine *et al.* 1989, Dezelee *et al.* 1989). *ref(2)P* does not affect female fertility (Contamine *et al.* 1989), although its role in reducing the transmission of sigma virus through females suggests that *ref(2)P* is carrying out an important function in the ovaries.

1.4.7 *ref(2)P* alleles

Natural populations of *D. melanogaster* contain both the susceptible and resistant alleles of *ref(2)P* (Contamine *et al.* 1989, Fleuriet 1988). The resistant allele was first discovered in Paris, France by Gay (1968*a*,*b*) and Ohanessian-Guillemain (1963), and was shown to reduce the rate at which some strains of sigma replicate within the fly. In contrast to this, flies with the susceptible allele never experience a reduction in sigma virus replication. This polymorphism is found worldwide and the resistant allele occurs on an average of 20% of chromosomes (Christian Schlötterer and Pablo Orozco, unpublished data on over 2000 alleles from 21 populations).

Previous studies have shown that there are six mutational differences between the resistant and susceptible alleles, however, only one of these differences—a complex mutation in exon 1 of *ref(2)P*, in which CAG-ATT (glutamine-asparagine) has changed to GGA (glycine)—has been shown to account for the difference in viral replication rates between the two alleles (Dru *et al.* 1993, Wayne *et al.* 1996).

Studies investigating the genetics of the resistant and susceptible alleles have shown that these alleles are co-dominant, since heterozygotes are intermediate between the two homozygotes in their resistance to sigma infection. These studies have also shown that the two *ref(2)P* alleles are antimorphic, because they cause opposing

phenotypes; this is compared to mutants, lacking *ref(2)P*, that have no effect on the phenotype of the fly. This finding is confirmed by the observation that individuals that are homozygous for the resistant allele are more susceptible to the sigma virus, if they also have a copy of the susceptible allele (Nakamura 1978). In addition to this, there is also evidence that the resistant allele displays haplo-insufficiency, since flies that are hemizygous—with only one copy of the resistant allele—are more susceptible to sigma virus infection than homozygous flies (Nakamura 1978). This is not true for the susceptible allele, since flies that are hemizygous and homozygous for the susceptible alleles are equally susceptible to the virus. This information is summarised below:

$$P^O/P^O = P^O/Del > P^O/P^p > P^p/P^p/P^O > P^p/Del > P^p/P^p \tag{1.1}$$

Summary of the different genotypes and their susceptibility to sigma virus. Resistant alleles =P^p*, susceptible alleles =* P^O*. Del = a deletion in the ref(2)P gene responsible for the difference between the two alleles. Susceptibility goes from left to right.*

It is worth remembering that mutations in *ref(2)P* affect both the replication of the virus within the fly and transmission of the virus to progeny. A recent study shows that *ref(2)P* resistant mutation strongly affects transmission of the virus from females to their offspring, but plays a more minor role in transmission of the virus from males to their progeny (Bangham *et al.* 2008*b*). This suggests that regardless of whether transmission occurs through the egg or sperm, the *ref(2)P* polymorphism affects the susceptibility of the zygote to infection with the sigma virus, rather than the ability of the parents to transmit the virus to their progeny. The authors of this study suggest that the zygote receives a very small amount of virus from its parent and so genetic variation that affects the clearance of the virus from the zygote can be very effective, perhaps totally clearing the viral infection before it takes hold.

To try to understand how *ref(2)P* affects viral replication and transmission, studies have looked for interactions between viral proteins and *ref(2)P* proteins(Avila *et al.* 2002). These studies have shown that *ref(2)P* interacts directly with viral proteins; forming complexes with both the *N* and *P* viral proteins. This interaction seems to be specific to sigma virus, as no complex was observed between *ref(2)P* and vesticular stomatitis virus—a close relative of sigma virus. This study suggests that *ref(2)P* targets the *P* protein, which is involved in viral replication, so that it might, at

least partially, control virus replication. Interestingly, antibodies used to target *ref(2)P* also associate with the *N* protein, suggesting that this nucleocapsid protein has many structural similarities to *ref(2)P*. This could be caused by a true association between *ref(2)P* and the *N* protein, however, molecular mimicry by the virus can not be ruled out. Viruses can mimic host proteins to avoid detection by the host's immune system (Srinivasappa *et al.* 1986).

1.4.8 Evolutionary genetics of resistance to the sigma virus

A number of studies have looked for evidence that *ref(2)P* is under selection, and if it is, what type of selection is acting on it. The earlier studies showed that selection has promoted amino-acid polymorphism within *ref(2)P* (Wayne *et al.* 1996). These studies identified an excess of amino-acid polymorphism among lines (relative to between species) at the 5' region of the gene (where the complex mutation occurs). This is consistent with both an arms race, in which a resistant allele is currently sweeping through (so we expect to see some variation), and frequency dependent selection maintaining variation in this gene. A reduction in variation among the resistant haplotypes would be good evidence that this allele is not being maintained as a balanced polymorphism but is instead sweeping through the population, however, sample sizes in earlier studies were too small to test this. In a recent study, Bangham *et al.* (2007) examined a larger sample of second chromosomes and found significantly less variation among the resistant haplotypes compared to the susceptible haplotypes than expected by chance. This finding suggests that this mutation has a selective advantage and has increased in frequency, along with linked sites either side of it, reducing overall diversity among the resistant haplotypes. Although these findings do not totally rule out the role of frequency-dependent selection in maintaining the resistant polymorphism, they do provide strong evidence that selection is acting on *ref(2)P*.

Estimating the age of the resistant *ref(2)P* mutation would indicate whether it has existed as a long-term balanced polymorphism or whether it has recently swept through *Drosophila* populations. In a recent study, the resistant *ref(2)P* mutation was estimated to be several thousand years old (Bangham *et al.* 2007). This result, together with the low variation among resistant haplotypes, suggests that the resistant mutation has slowly increased in frequency as part of a selective sweep. This puts the

spread of the resistant mutation much earlier than several previous studies, that show an increase in infective viral genotypes in both French and German populations of *D. melanogaster* during the 1980s. This indicates a very recent sweep of a resistant *ref(2)P* mutation, followed by a sweep of 'infective' viruses, that can infect flies that have the resistant *ref(2)P* allele.

It is possible to reconcile the findings of these two studies with a number of arguments. First, it is worth remembering that the two studies looked at different populations that might be experiencing different selective sweeps—Fleuriet (1990), Fleuriet *et al.* (1990) examined European populations, while Bangham *et al.* (2007) examined populations in the USA. In support of this idea, 'infective' viral genotypes are most common in France (85%), declining in the rest of Europe (30%) to become very rare in Africa and USA. Therefore, *ref(2)P* might be exerting greater selection pressure in European *Drosophila* populations because the virus is at higher frequency in Europe.

Second, Bangham *et al.* (2007) have evidence to show that the resistant mutation is recessive, which means it could take thousands of years to reach the current frequency. At low frequency, the resistant mutation would not exert much of a selective pressure on the virus and so it perhaps only recently became frequent enough to select for counteradaptation in the virus. In support of this, the frequency of the resistant mutation in samples collected across three continents has never exceeded 23%. This means that homozygous flies—that are resistant to sigma infection—are rare (~5%).

Is *ref(2)P* responsible for controlling sigma infection in wild *Drosophila* populations? Early work suggested that the *ref(2)P* resistant allele is responsible for sigma's low frequency in natural populations (0-15%). Evidence for this came from the results of cage experiments where the virus, introduced in the absence of the resistant allele, always spread to 100% (Fleuriet 1978). However, when populations were set up with the resistant-susceptible polymorphism at natural frequency, or near enough (resistant allele = 30%), and sigma at natural levels (10%), it was found that although the resistant allele maintains itself at 30%, the frequency of sigma infected flies increased from 10% to 90%. This trend continued irrespective of the density at which the resistant allele was introduced. Furthermore, the resistant allele is maintained in a population irrespective of whether the virus is present or not, suggesting that heterosis

might be maintaining this polymorphism at its natural level.

However, a great deal of data indicates that the role of *ref(2)P* resistant mutation in affecting infection levels in wild populations of *Drosophila* should not be discounted. If the virus confers some cost to its carriers, and this cost is not experienced by flies within the cages, then sigma infection would exceed natural levels within these experiments, irrespective of influence of the resistant *ref(2)P* mutation.

1.5 Aims of this study

This thesis explores the genetics of the *Drosophila*—sigma virus system using both experimental and phylogenetic approaches. Four different studies have been carried out, all of which address questions related to the co-evolutionary dynamics between *D. melanogaster* and the sigma virus.

The interactions between viruses and their hosts have been largely under-exploited by *Drosophila* biologists studying immune systems—despite viruses commonly occurring in natural populations of *Drosophila*. For this reason, in Chapter 3, I describe new sigma isolates that were collected from populations of *Drosophila* from different places in the world. These viral isolates allow a number of key questions to be answered: what is the prevalence of the sigma virus; how does the prevalence vary between populations; how long has the virus persisted in *Drosophila* populations; and how is genetic variation in the virus distributed across *Drosophila* populations? To answer these questions I compare levels of viral genetic diversity in the sigma virus across Europe and North America to those found in other RNA viruses. Based on laboratory measurements of the viral substitution rate, I estimate a common ancestor for European and North American sigma viral isolates, and examine the population structure of the sigma virus.

An insect's ability to first evade infection, to recognise and suppress the infection, and finally eliminate the infection, is dependent on the genetics of both the pathogen and the host. And yet, few studies have examined both the genetics of the pathogen and host within one system. In Chapter 4, I take several of the new viral isolates and test for genotype-specific interactions in *D. melanogaster*. In this Chapter, I aim to investigate whether different host genotypes differ in their susceptibility to viral isolates; whether viral isolates differ in their ability to infect different host genotypes

and ultimately whether there are host-viral genotype interactions? To test this, I measure the transmission rates for five different viral isolates collected from around the world in flies lines that differ for their first, second and third chromosome.

The Toll pathway is an important component of the innate immune system that has been shown to be activated by other *Drosophila* viruses (Zambon *et al.* 2005). In Chapter 5, I describe several experiments that examine whether the Toll pathway is involved in an antiviral response against the sigma virus. Testing both whether flies that lack a Toll pathway are more susceptible to sigma-infection and whether in turn, the sigma virus has evolved to suppress this immune response, as would be expected under coevolutionary theory. I also investigate whether the sigma virus can cause flies to be more susceptibility to fungal infections. This could be important in the wild where flies are exposed to a variety of different pathogens.

Finally, in Chapter 6, I describe the first evidence from outside mammals that viruses can be hypermutated by host ADARs. Adenosine deaminases that act on RNA (ADARs) are RNA-editing enzymes that target regions of dsRNA and cause hyper-editing. Typically, they are involved in post-transcriptional editing of host genes, but there is evidence that they also edit viruses. I examine other viral isolates from around the world for evidence of hypermutation caused by ADARs, and investigate the role of ADAR as an antiviral mechanism by looking to see whether the sigma virus has evolved to suppress ADAR editing.

2 General materials & methods

The standard experimental techniques and solutions used throughout the course of this work are detailed in this Chapter.

2.1 RNA extraction techniques

This section provides details of techniques used to purify genomic viral RNA. Two methods for extracting RNA were used, the method described in section 2.1.1 refers to Chapter 3 and 6 and section 2.1.2 refers to Chapter 4.

2.1.1 Isolation of RNA from *Drosophila* with kit

Total template viral RNA and *Drosophila* RNA was obtained from multiple flies from a single line using the protocol below and a Total RNA Isolation System (Promega, WI, USA), which contains RNAagents® Denaturing Solution, 2M Sodium Acetate Solution, Phenol:Chloroform:Isoamyl Alcohol and Isopropanol. Forty flies were used to ensure that we were able to isolate sufficient quantities of viral genomic RNA.

1. Preparation: 0.1% NaOH solution was used to wipe over counter and pipettes, followed by 70% ethanol (30% DEPC treated water). The microcentrifuge tube pestle were left in 1% DEPC-treated water overnight and then autoclaved and dried.
2. Tissue homogenizing: 300μl of chilled Denaturing Solution was added to forty flies that had been chilled in 1.5ml eppendorf tube on ice. The flies were then homogenised thoroughly using a microcentrifuge tube pestle and the sample placed back on ice.
3. RNA extraction. 30μl of Sodium Acetate Solution was added to the tube and mixed carefully by inversion 5 times. 300μl Phenol:Chloroform:Isoamyl Alcohol was then added to the sample and mixed by inverting 5 times before

shaking vigorously for 10 seconds before the sample was put on ice for 15 minutes. The sample was then centrifuged at 13,000 rpm for 20 minutes so that the fly material forms a tight pellet. The supernatant containing the RNA was then carefully removed and transferred to a fresh DEPC-treated 1.5ml eppendorf tube, leaving behind the organic phase and the interface. Care was taken to avoid taking material from the interface, which contains the genomic DNA.

4. RNA resuspension. 50μl of RNA storage solution was added to the dried pellet and placed in a water bath at 50°C for 10 minutes before gently mixing the sample with a pipette to allowed the RNA to resuspend. Resuspended RNA was stored at -80°C.

2.1.2 Isolation of RNA from *Drosophila* with Trizol®

Genomic template viral RNA and *Drosophila* RNA was obtained from multiple flies from a single line using the Trizol® based protocol (Invitrogen, Paisley, UK) below:

1. Tissue Homogenizing: 250μl of Trizol® solution was added to ten flies that had been chilled in 1.5ml eppendorf tube on ice. The flies were then homogenised thoroughly using a microcentrifuge tube pestle and the sample was mix by inverting the tube and incubated at room temp for 5 minutes.
2. RNA extraction. The sample was then centrifuged at 13,000 rpm at 4°C for 10 minutes so that the fly material forms a tight pellet. The supernatant containing the RNA was then carefully removed and transferred to a fresh DEPC-treated 1.5ml eppendorf tube. 80μl of chloroform solution was added to the tube and mixed well by shaking the tubes vigorously for 15 seconds by hand. The samples were then incubated at room temperature for 3 minutes before being centrifuged 13,000 rpm at 4°C for 10 minutes. The supernatent was then transferred to a fresh DEPC-treated 1.5ml eppendorf tube, leaving behind the organic phase and the interface. Care was taken to avoid taking material from the interface, which contains the genomic DNA.
3. RNA precipitation. 200μl of Isopropanol was added to the supernatant and the sample was mixed by inverting the tube gently and then incubating the sample at room temperature for exactly 10 minutes to precipitate the RNA. The sample was then centrifuged at 13,000 rpm at 4°C for 10 minutes. The Isopropanol was

then poured out of the tube leaving behind the pelleted RNA. The pellet was then washed by adding 1ml of ice-cold 75%ethanol:25% DEPC-treated water and gently flicking the tube. The sample was then centrifuged again at 13,000 rpm at 4°C for 10 minutes. The ethanol was then removed and the pellet air-dried in a clean environment for 15 minutes.

4. RNA resuspension. 50μl of RNA storage solution was added to the dried pellet and placed in a water bath at 50°C for 10 minutes before gently mixing the sample with a pipette to allowed the RNA to rehydrate. Rehydrated RNA was stored at -80°C.

2.2 RNA sequencing techniques

This section provides details of techniques used to amplify and sequence genomic viral RNA.

2.2.1 Reverse transcription

Genomic RNA was turned into cDNA by reverse transcription. Reverse transcription (RT) was carried out in a MJ Research DNA Engine DYAD (Essex, UK). All RT reagents were obtained from Promega (WI, USA). Primers were designed using Primer 3 Rozen & Skaletsky (2000).

1. Reaction mix: for a single 20μl reaction, the following were mixed in 0.5ml tubes on ice:

10μl	Template RNA
2μl	2mM Primers (3 for *Drosophila* template RNA and 5 for viral template RNA)
1μl	10μM dNTP

The tubes were then heated to 65°C for 4 minutes, then cooled immediately on ice for 5 minutes. The tubes are then centrifuge briefly (to bring the reagents to the bottom of the tube) and add the following reagents:

1μl	RNasin® Plus RNase Inhibitor
4μl	5× M-MLV RT Buffer
2μl	0.1M DTT
20μl	

The tubes were then heated to 37°C for 2 minutes and 1μl M-MLV RT was added before the tubes were incubated at 42°C for 50 minutes and then heat to 70°C for 15 minutes.

2.2.2 Standard polymerase chain reaction (PCR)

PCR reactions were carried out in a MJ Research DNA Engine DYAD (Essex, UK). All PCR reagents were obtained from Sigma (Dorset, UK). Primers were designed using Primer 3 Rozen & Skaletsky (2000).

1. Reaction mix: for a single 20μl reaction, the following were mixed in stripe tubes on ice:

10.9μl	roH_2O
2μl	10× buffer
2μl	2mM dNTP
2μl	25mM $MgCl_2$
0.1μl	Taq DNA polymerase
1μl	10μM Forward Primer
1μl	10μM Reverse Primer
1μl	Template DNA
20μl	

2. PCR program:

95°C for 3 mins
95°C for 30 secs
62°C for 30 secs } 10 cycles, dropping annealing temperature 1°C every cycle
72°C for 2 mins
95°C for 30 secs
52°C for 30 secs } 35 cycles
72°C for 2 mins
72°C for 5 mins

2.2.3 Agarose gel electrophoresis

Successful amplification of the target DNA was checked by running samples on 1% agarose gel. Samples to be loaded on to the gel were mixed with an equal volume of 1.5× loading dye and were loads alongside a size marker (1kb DNA ladder, Promega, WI, USA). Horizontal gel electrophoesis was carried out in a 1× TBE buffer in Bio-Rad gel tanks (Hercules, CA). Gels were run at 100V. After electrophoresis, results were visualised and photographed under UV light using a camera (Genetics Research Instrumentation Ltd., Essex, UK).

2.2.4 EXOSAP cleanup

1. To sequence, unincorporated dNTPs and primers were digested by adding the following to each reaction on ice:

1μl	1u/μl SAP
0.075μl	20u/μl EXO
1.425μl	Dilution Buffer
4μl	cDNA Template

2. EXOSAP program:

   37°C for 47 mins
   95°C for 15 mins

2.2.5 Sequencing reaction

BigDye® Version 3 (Applied Biosystems, Foster City, CA 94404) was used for all sequencing reactions:

1. Reaction mix: the following reaction mix was prepared in sterile 0.5ml micro-tubes on ice.

   1.5μl BigDye® terminator ready reaction mix
   1μl Primer (3.2μM)
   1.2μl 5× Buffer
   3.3μl roH_2O
   4μl PCR template

2. Sequencing program:

95°C for 30 secs	25 cycles
50°C for 25 secs	
60°C for 4 mins	

2.2.6 Pyrosequencing

Pyrosequencing was used to identify a single base change in *Drosophila* RNA.

1. Sample preparation: RNA was obtained from 10 *Drosophila* using a Promega RNA isolation kit (as in section 2.1.1) and cDNA was synthesised with a standard RT procedure (as in section 2.2.1). cDNA was then amplified using the following program:

95°C for 15 secs	
95°C for 15 secs	35 cycles
63.7°C for 15 secs	
72°C for 15 secs	
72°C for 5 mins	

2. The samples were then prepared for pyrosequencing by mixing the following in a 96 well plate:

2μl	Bead solution (Streptavidin Sepharose)
38μl	Binding Buffer
10μl	Template cDNA
30μl	roH_2O

The plates were sealed and placed in a shaker for 10 minutes to anneal cDNA to the beads. The beads are suctioned into a lower vacuum preparation tool and washed with 70% ethanol for 5 seconds, then denaturing solution for 5 seconds and then washing buffer for 5 seconds. The beads were then returned to a 96 well plate and the sequencing primer (0.3μM) was added. The samples were then run through the pyrosequencer to generate the sequence reads.

2.2.7 Cloning

All cloning was done using a TOPO TA cloning® kit (Invitrogen, Paisley, UK). template PCR product was dA-tailed.

1. Ligation: 2μl of template cDNA was added to a stripe-tube containing 0.5μl salt solution and 0.5μl TOPO vector and left to incubate at room temperature for 30 minutes before being placed on ice.
2. Transformation: 2μl of the ligated vector was added to a tube containing chemically competent E. coli cells and incubated on ice for 20 minutes. The samples were then placed in a water bath at 42°C for 30 seconds before being returned to ice. 250μl of SOC medium was then added to the samples and agitated in a 37°C water bath for an hour.
3. Culturing: 50-100μl of the transformed samples were plated out on amphicillin plates and incubated for 12 hours at 37°C.

2.2.8 Fly culturing techniques

This section describes techniques used to culture *Drosophila* melanogaster.

1. All stocks were maintained on standard *Drosophila* media made up by mixing the following:

6L	water
41g	agar
562g	sugar
415g	maize
112g	yeast

   The mixture was then heated to boiling and simmered for 10 minutes and then allowed to cool until around 70°C before 92ml of Nipagin solution was added. Sodium hydroxide was added to bring the media to Ph 7.0.

2.2.9 Fly culturing techniques

Flies were collected from patches of banana mixed up with yeast or rotting fruit. Nets were swept over patches and single flies pootered into a tube or eppendorf. When

collecting isofemale lines, care was taken to collect from a number of different patches to avoid collecting siblings emerging from one patch of fruit.

3 Sigma virus phylogenetics

The work described in this Chapter has been recently published (Carpenter *et al.* 2007).

3.1 Introduction

Drosophila melanogaster is a model system for studying innate immune systems. Studies in *Drosophila* have made important contributions to our understanding of how the insect immune system recognizes and responds to micro-organisms. Many of the most influential studies of the innate immune system have involved challenging flies with general immunoelicitors, such as bacterial endotoxin lipolysccharide (LPS) (a component of bacterial cell walls) or bacteria that would not naturally infect *Drosophila*, such as *E. coli* or *M. luteus*. In these studies, infections are often established by introducing the pathogen directly into the body cavity of the fly. Although these studies provide a model of infection following septic injury, they do not examine specific defences against natural pathogens. Therefore, studies of natural host-pathogen interactions are needed to help us to understand how hosts evolve specific defences against parasites and how parasites evolve to suppress and evade those defences.

Recent studies have investigated how much genetic variation exists in natural populations of *D. melanogaster* for susceptibility to fungus (Tinsley *et al.* 2006), bacteria (Lazzaro *et al.* 2004) and viruses (Bangham *et al.* 2007, 2008*b*). To date, however, there has been very little research conducted to determine the frequency, identities, and virulence of microbes that infect wild *D. melanogaster*, and still fewer studies have examined how much variation for infectivity or virulence exists in these microbe populations.

In particular, the interactions between viruses and their hosts have been largely under-exploited by *Drosophila* biologists studying immune systems (Cherry & Sil-

verman 2006). Seven RNA viruses have been isolated in natural populations of *D. melanogaster* (Berkalof *et al.* 1965, Habayeb *et al.* 2006, Plus & Duthoit 1969, Plus *et al.* 1976, Teninges & Plus 1972). However, few studies have investigated how common these viruses are in wild populations or how prevalent they are within populations. Determining how viruses vary spatially and temporally in the wild is important for understanding how strong a selection pressure viruses exert on their hosts.

In this Chapter, I investigate the phylogenetics of the sigma virus in wild populations of *D. melanogaster*. The mechanisms of sigma-virus transmission in the laboratory have been described (Brun & Plus 1998), and natural populations of *D. melanogaster* have been investigated for evidence of variation in resistance to sigma infection (Fleuriet 1986). However, very little is known about the biology of the sigma virus in natural populations (Fleuriet 1976*a*). For this reason, I have collected new viral isolates from global populations of *Drosophila* to allow a number of questions to be addressed: what is the prevalence of the virus; how does the prevalence vary between populations; how long has the virus persisted in *Drosophila* populations; and how is genetic variation in the virus distributed across *Drosophila* populations?

3.2 Materials and methods

3.2.1 Collection

Sigma virus isolates were collected by isolating single female *Drosophila* from wild populations, allowing them to lay in a vial on standard *Drosophila* food, and exposing a proportion (50%) of their offspring to pure CO_2 at 12°C for 15 minutes. The flies were then allowed to recover at room temperature. Infected flies are sensitive to CO_2—they die or become severely paralysed in contrast to uninfected flies that fully recover (L'Heritier & Teissier 1937). Only CO_2 sensitive lines were kept and each isofemale line is termed a viral isolate. The flies were collected from vineyards and fruit farms from a number of locations (see Table 3.1). Flies were collected over 1-7 days from patches of fruit within 1-2 miles of each other. Didier Contamine supplied two viral lines (A3 and A3E55) that had been maintained in the laboratory. These viral lines are a single wild-collected isolate which was split between two separate fly lines and maintained at 20°C. Unfortunately, due to lost records, the precise age at

which the lines were split is unknown, however I know it is more than 10 years and less than 20 years.

3.2.2 Sequencing and sequence analysis

PCR primers were designed from the published sigma sequence from GenBank (x91062) using the program PRIMER 3. I sequenced one large fragment from three wild-collected isolates and two viral lines maintained in the laboratory, and two shorter fragments from a larger sample of viral isolates from five wild-collected populations (see Figure 3.1). Both viral and fly RNA were extracted using a Total RNA Isolation System (Promega, Wisconsin). The PCR primers were used to reverse transcribe the genome using M-MLV reverse transcriptase. Fragments were then amplified by PCR. Prior to sequencing, unused PCR primers and dNTPs were digested with exonuclease 1 and shrimp alkaline phosphase. The PCR products were then sequenced directly using the PCR primers and Big Dye reagents (ABI, Foster City) on an ABI capillary sequencer. In cases where PCR products could not be sequenced directly, fragments were cloned using the TOPO TA Cloning kit (Invitrogen, California). Multiple clones were mixed to avoid PCR errors and sequenced. Sequences were initially assembled using Sequencher 4.5 (Gene Codes Corporations) and chromatograms were inspected by eye to confirm the legitimacy of polymorphisms between viral lines. The sequences were then aligned using CLUSTAL W and genes were identified with reference to published sigma sequences.

3.2.3 Estimation of mutation rates and phylogenetic reconstruction

Sequences for two viral lines (A3 and A3 E55) that shared a common ancestor 10 – 20 years ago were aligned. Each mutation was then assigned to either the A3 lineage or the A3 E55 lineage using the sequences from a divergent isolate as an out-group (Essex line: E26). The substitution rate per site per year was then estimated independently for each lineage. I calculated K_S, the number of synonymous substitutions per synonymous site and π, the average pair-wise difference between two sequences using DNAsp 4.10. The Nei and Gojobori (Nei & Gojobori 1986) method was used to calculate the synonymous substitution rate.

An alignment was created of concatenated sequences from the polymerase-

associated gene and outer-coat protein gene. Nucleotide sequences were used to reconstruct phylogenies by maximum likelihood in PAUP* v.4.0b10 (Swofford 2002). The HKY85 with gamma distribution rate heterogeneity between sites (Hasegawa *et al.* 1985) was selected as the appropriate model of sequence evolution by comparing models using likelihood ratio tests in Modeltest 3.7 (Posada & Crandall 2001). Trees were constructed using a heuristic search algorithm and optimisation was performed by branch swapping using nearest-neighbour interchanges.

I performed three tests for recombination within the sigma virus genome. First, I tested for recombination between the sequenced genes by constructing maximum-likelihood trees separately for each gene, and then forcing each gene to take the topology of the other gene. Recombination was indicated if the forced topology had a significantly lowered likelihood relative to the gene's own maximum likelihood tree using a one-tailed Shimodaira-Hasegawa likelihood ratio test (Shimodaira & Hasegawa 1999*a*). Second, I also tested for recombination using the maximum chi-squared test (Maynard-Smith, 1992) and the Reticulate test (Jakobsen & Easteal 1996), which were performed in the program RDP (Martin & Rybicki 2000). The maximum chi-squared test identifies potential recombinant events between two sequences and a putative derived sequence. I used a sliding-window analysis (50bp, 1bp steps). At each step, the number of variable sites was compared in the left and right halves of the window using chi-squared test. Potential breakpoints correspond to peaks in the values of chi-squared. Third, I used the Reticulate test, that identifies regions of sequence within an alignment that have phylogenetic relationships that are incompatible with each other. The test then estimates whether these regions are longer than would be expected by chance. The test statistic is the Neighbour Similarity Score (NSS), which is the average proportion of times a region is compatible (shares a phylogenetic history) with a neighbouring region. The null distribution of both the chi-squared and NSS statistics were generated by recalculating them 104 times from datasets where the order of sites had been permuted.

I tested for population structure among the viral isolates using Hudson's (Hudson 2000) nearest-neighbour statistic (S_{nn}) estimated in DNAsp 4.10. To assess the significance of observed (S_{nn}) sequences were randomly assigned to localities, maintaining the same number of sequences in each locality as in the original sample. The proportion of permutated samples with (S_{nn}) larger or equal to the observed value is the estimated P value. I also report average values of K_{ST} (an analog of F_{ST}; (Hudson

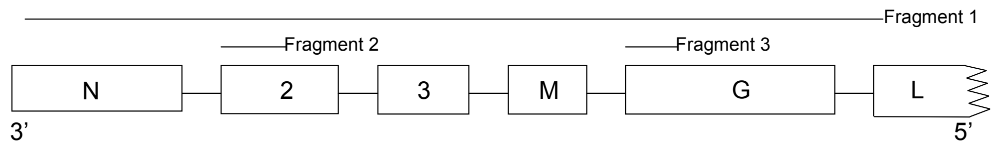


Figure 3.1: Schematic of the sigma virus. The six genes shown encode the following proteins: N, nucleocapsid protein; 2, polymerase-associated protein; 3, PP3; M, matrix protein; G, outer-coat protein; and L, polymerase protein. The L gene is incomplete. The first fragment covers the five genes and part of the polymerase gene; the second fragment covers 67% of the polymerase associated gene (636bp); and the third covers 36% of the outer-coat protein (589bp).

et al. 1992) across all the populations, and estimate a P value from the proportion of time that the observed K_{ST} value is greater than an estimated value of K_{ST} based on randomly partitioning the dataset among localities.

3.3 Results

3.3.1 Viral prevalence

I estimated the prevalence of sigma virus in a population by measuring the proportion of infected isofemale lines established from wild-caught females. The number of infected individuals varied greatly between populations ($\chi_2 = 47.55, d.f. = 9, P < 0.001$) (Table 3.1). I found the highest prevalence of sigma virus in some of the European populations (0-15%), it was lower in North America (0-6%) and I failed to find sigma virus in African populations (Table 3.1). The prevalence of sigma virus varied widely, even between neighbouring collection sites: for example, viral prevalence differed dramatically between populations collected at Essex and Kent, not more than 100 kilometres apart (Table 3.1). The virus was not found in other species of *Drosophila* collected alongside *D. melanogaster* (Table 3.2).

Table 3.1: Incidence of CO_2 sensitivity in wild-caught isofemale lines of *Drosophila melanogaster*

Population	No. of lines	No. of infected lines	Percentage infected	Date Collected
Apshawa, Fl, USA	65	1	1.5	March 2005
Wildwood, Fl, USA	32	0	0	March 2005
Georgia, USA	32	2	6.2	September 2005
New York, USA	16	0	0	June 2006
Nairobi, Kenya	125	0	0	May 2005
Athens, Greece	97	17	14.9	June 2005
Essex, UK	211	16	7.0	June 2005
Kent, UK	125	0	0	June 2005
Galicia, Spain	175	8	4.3	October 2005
Tenerife	198	0	0	June 2006
Vienna, Austria	13	0	0	September 2005
French Polynesia	10	0	0	June 2005

Table 3.2: Incidence of CO_2 sensitivity in wild-caught isofemale lines of *Drosophila simulans*

Population	No. of lines	No. of infected lines	Percentage infected	Date Collected
California, USA	42	0	0	March 2005
Tenerife	241	0	0	June 2006

3.3.2 Recombination

I sequenced two regions from a large sample of viral isolates: the first encompasses the polymerase-associated protein; the second includes the *G* gene, encoding the outer-coat protein (590bp and 637bp respectively; Figure 3.1). The sequences come from virus isolated in *D. melanogaster* lines collected in Europe (Greece, UK and Spain) and North America (Georgia and Florida, USA). These sequences are mostly protein coding but include small intergenic regions. I compared the tree topology for the two sequenced regions and found no significant conflict between genealogies using a one-tailed Shimodaira-Hasegawa likelihood ratio test ($P > 0.05$) (Shimodaira & Hasegawa 1999*b*). In the absence of recombination, these genes should share evolutionary histories, and so the lack of conflict among the genealogies suggests no recombination has occurred. I further tested for recombination between viral lines using a maximum χ_2 test and found no evidence for any significant breakpoints in our sequences ($P = 0.29$). Further support for a lack of recombination in the viral sequences comes from the Reticulate test. A NSS of 0.99 ($P = 1.0$) for the sequences indicates that neighbouring sites share similar phylogenetic histories as often as distant sites, as would be expected when there is no recombination. The lack of recombination between sequences allows the concatenation of the two sequenced regions (the polymerase associated protein and outer-coat protein), affording a larger data set for phylogenetic analysis.

3.3.3 Viral sequence variation and population structure

I sequenced 1224bp from between two and ten isolates from each of five populations. I analyzed sequence data for levels of variability and found that within-population variation is similar in all populations and is extremely low (Essex: $\pi = 0.00062$; Spain: $\pi = 0.00654$; Greece: $\pi = 0.00059$). I found that one more divergent American isolate, collected in Florida, differs considerably from all other European isolates ($K_S = 0.395$) while the other American isolate, collected in Georgia, is more similar ($K_S = 0.0475$) (Figure 3.2).

The sigma virus population is highly structured; I found that each European viral population forms a separate monophyletic group (Figure 3.2). This is evident in the high value of Hudson's nearest neighbour statistic (S_{nn}) (Hudson 2000), which

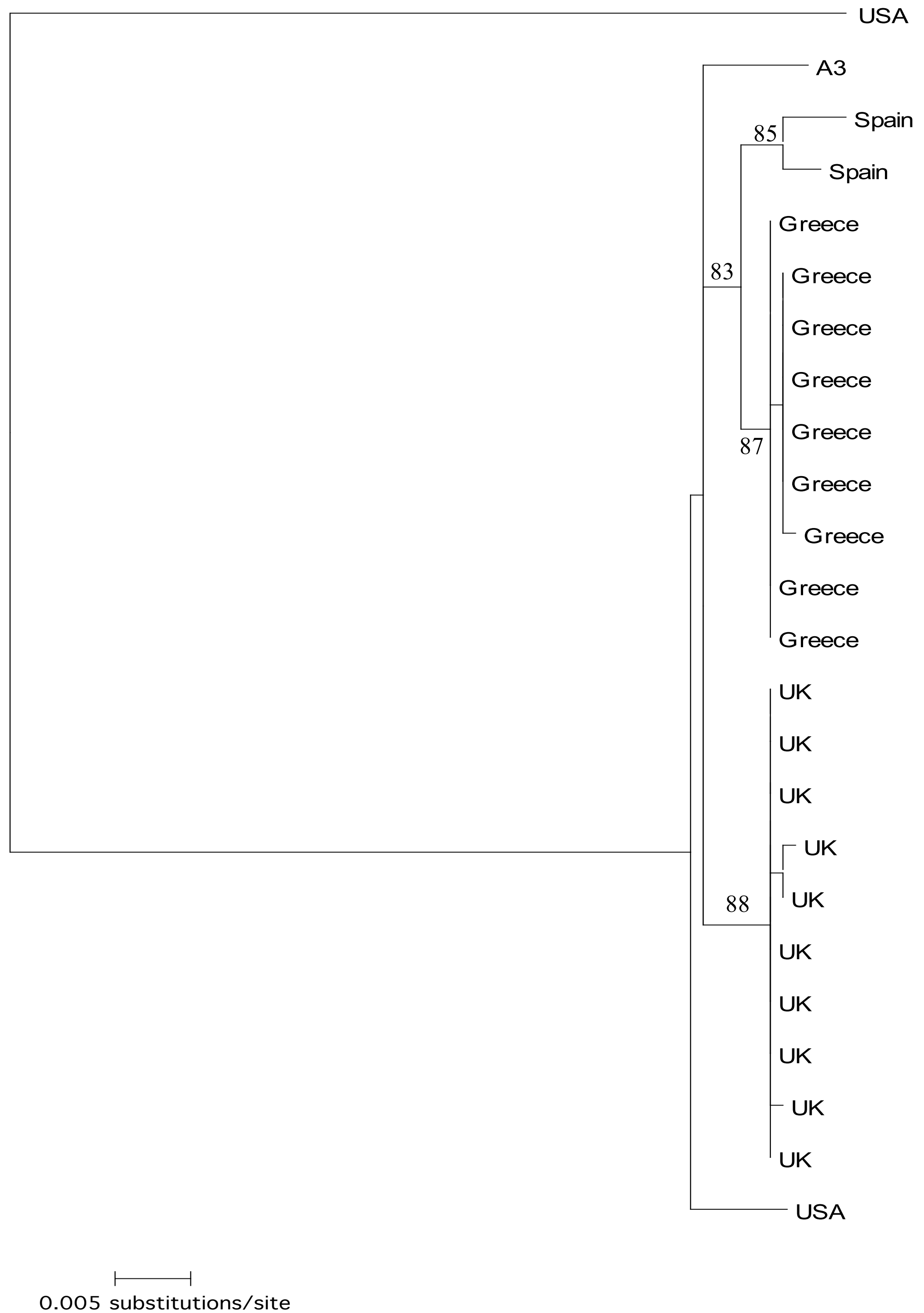


Figure 3.2: Maximum likelihood tree of the concatenated P and G genes of the sigma virus based on all sites. Bootstrap values are based on maximum-likelihood analysis (1000reps).

measures the proportion of times that nearest neighbours (in terms of genetic distance between sequences) are found in the same population (S_{nn} = 0.964; $P < 0.001$). The population-structuring is also reflected by high average K_{ST}, which compares genetic diversity within and between populations (K_{ST} = 0.730, $P < 0.001$) (Hudson *et al.* 1992). American viral isolates were excluded from population structure analysis due to limited sampling.

3.3.4 Estimating substitution rate

I estimated the substitution rate in the laboratory by sequencing 5744bp of the viral genome from two viral lines (A3 and A3 E55) split from each other and maintained at 20°C for 10 to 20 years. I found 25 substitutions between these two viral lines. Using an out-group, I assigned these mutations to either the lineage leading to A3 or A3 E55. I found that A3 E55 accumulated significantly more substitutions than A3 (χ_2 = 11.58, d.f. = 1, $P < 0.001$), with 21 of the 25 substitutions unique to A3 E55. The high substitution rate in the A3 E55 lineage was caused by hypermutation in a small region of the genome. These data are discussed in Chapter 6. I failed to find any evidence of similar patches of hypermutation in natural isolates of the virus. I therefore estimated viral substitution rate independently for the two lineages, and used a substitution rate based solely on data from the A3 lineage to date phylogenies. I estimate that the A3 lineage accumulated 4.6×10^{-5} substitutions per site per year (4 mutations have occurred in 5744bp of sequence in c.15 years). Assuming the substitutions follow a Poisson process I calculated the 95% confidence intervals around this rate (1.8×10^{-4} – 9.5×10^{-6} substitutions per site per year). As the exact time since the viruses split is unknown, I conservatively used 10 years and 20 years for upper and lower limits respectively. I calculated the synonymous substitution rate to be 1.0×10^{-4} (2 mutations have occurred in 1249 synonymous sites in c.15 years) (95% confidence intervals: 5.8×10^{-4} – 9.7×10^{-6} substitutions per synonymous site per year). When compared to a phylogeny of the natural isolates based only on synonymous sites, I found that A3 did not have an unusually long or short branch, suggesting that A3 is accumulating changes at a rate similar to these natural isolates, and so it is appropriate to extend the estimate of substitution rate based on A3 to wild viral populations.

3.3.5 Age of viral spread

The low sequence diversity across all viral populations is consistent with the hypothesis that a single viral type has recently spread through European *D. melanogaster* populations and across the Atlantic to North America. Using the substitution rate determined in the laboratory I was able to estimate the time since the viral isolates shared a common ancestor. I found that all the viral isolates, except the most divergent Florida isolate, shared a common ancestor 214 years ago (95% CI on the substitution rate estimate: 55 years–1036 years). All sequences collected (including the Florida isolate) shared a common ancestor 2,106 years ago (95% CI on the substitution rate estimate: 538 – 10,196 years ago). The genetic distance between sequences was taken from the mean depth of the tree shown in (Figure 3.2).

3.4 Discussion

3.4.1 Low viral diversity

The high mutation rate of RNA viruses typically leads to the existence of high levels of standing genetic variation. However, I detected very low sequence diversity across European and North American sigma populations. The average genetic diversity (π) within populations of the sigma virus is between 0.05% and 0.6%, lower than that reported for other RNA viruses (Garcia-Arenal *et al.* 2001). One cause of this low diversity is that sigma has a low viral substitution rate (4.6×10^{-5} substitutions/site/year) compared to many other RNA viruses (Davis *et al.* 2005). Sigma might experience low substitution rates because it is slow to replicate, or this may simply be a property of negatively stranded RNA viruses as sigma's substitution rate is comparable to rates in other rhabdoviruses (European bat lyssavirus (EBLV): 5×10^{-5} and rabies virus (RV): 2.3×10^{-4} substitutions/site/year) (Davis *et al.* 2005, Hughes *et al.* 2005).

The second cause of the low viral diversity is that most of the viral isolates collected shared a common ancestor as recently as two hundred years ago. Why do these viral sequences share such a recent ancestor? And why is one of our viral isolates much more divergent than the others? There are two possible explanations: the first is that *D. melanogaster* has recently acquired the sigma virus; the second is that a single

viral type has recently swept through *D. melanogaster* populations.

This first possibility is that the sigma virus has recently invaded *D. melanogaster* from another species and has spread rapidly through its novel host. It is likely that the biparentally transmitted sigma virus will spread quickly through an uninfected host population. Under this hypothesis the more divergent viral isolate collected in Florida may represent a separate invasion event. But how did the sigma virus, normally transmitted only vertically, first infect *D. melanogaster*? If other *Drosophila* are infected with sigma, perhaps a cross species transfer could occur in nature, with parasitic mites or wasps acting as carriers. Mites have been implicated as potential carriers of other vertically transmitted parasitic elements (Houck *et al.* 1991). In support of this idea, we know that sigma virus readily infects other *Drosophila* species when viral particles are directly microinjected into adult flies (Brun & Plus 1998). Further to this, sigma infection may be the cause of the CO_2 sensitivity observed in other species of *Drosophila* (though we didn't find any) and this CO_2 sensitivity is transmittable by injection (Williamson 1961). Also, field collections of both *D. affinis* and *D. athabasca* have shown CO_2 sensitivity characteristic of sigma virus infections, however these reports have not been confirmed molecularly (Williamson 1961).

The spread of a recently acquired viral infection through populations of *D. melanogaster* might be analogous to the spread of other parasitic elements through *Drosophila* populations. One example is the recent invasion of P elements—discovered in studies of hybrid dysgenesis—into the *D. melanogaster* genome (Kidwell *et al.* 1977). From an analysis of P-elements in the genus *Drosophila*, it was found that P-elements were transmitted from *D. willistoni* to *D. melanogaster*, mostly likely vectored through mites. This transmission could have occurred as recently as *D. melanogaster* (an Old World species) and *D. willistoni* (a New World species) became sympatric only 300 years ago, when *D. melanogaster* was introduced to the New World by humans (Lachaise & Silvain 2004). Another example of horizontal transmission associated with the migration of *D. melanogaster* to the New World is the introduction of Spiroplasma—a male-killing bacterium—into *D. melanogaster*. The transmission of the bacterium was mostly likely from *D. nebulosa*, a New World species in the willistoni group. In a final example of a parasitic element spreading through *Drosophila* populations, the cytoplasmic incompatibility (CI) causing Wolbachia to spread through populations of *D. simulans* in western USA over a stretch of ten years (Turelli & Hoffmann 1991). In this case, long distance

dispersal saw infections spreading at a rate of more than 100km per year to reach levels of 80% infection in areas that previously had none. The emerging picture suggests that host-parasite associations are very dynamic, with parasites constantly gained and lost from host populations.

The second possibility for the low levels of diversity among sigma viruses is that these infections represent ancient host-parasite interactions that have undergone a recent selective sweep. In support of this, we know that sigma populations are dynamic—some European populations have undergone successive replacement of a viral type that was sensitive to a host resistant gene by an insensitive type during the eighties (Fleuriet 1990, Fleuriet & Sperlich 1992). If this is true, the more divergent viral isolate collected in Florida may be a remnant viral type from a past sweep.

In summary, I offer two alternatives to explain the low levels of viral diversity in this study—recent invasion and selective sweep. If I had seen recombination in the sequences it would have been possible to separate these two explanations. This is because a selective sweep reduces the diversity around a single locus that is under selection, while the invasion of the sigma virus from elsewhere will affect diversity levels across the whole genome. However, I found no evidence of recombination in the sigma virus. Therefore, all sites remain in complete linkage to a site under positive selection, making positive selection on this site indistinguishable from demographic effects. The discovery of a closely related sigma virus in a relative of *D. melanogaster* would offer one possible approach to separating these hypotheses.

3.4.2 Population structure

The sigma virus shows extremely high levels of population structure within Europe, with each population sample forming a separate monophyletic group. This indicates that there has been no migration between populations of the virus, which is reflected in the high value for $K_{ST}*$ ($K_{ST}* = 0.730$). I might have expected the virus to mirror its host's population structure—the sigma virus is vertically transmitted and so is constrained within lineages of *D. melanogaster* and therefore experiences a similar migration rate to its host. European *D. melanogaster* populations show moderate genetic differentiation—in a study of 48 microsatalite loci, significant but low F_{ST} values were found across all six European populations studied ($F_{ST} = 0.053$) (Caracristi & Schlotterer 2003). By contrast, the spatial structure of viral populations

is much more pronounced.

How can we explain this discrepancy? If we estimate the size of the virus's effective population from its F_{ST} value, we can compare this estimate with the virus's true effective population size. This allows us to assess whether the virus's smaller effective population can explain the structure that we see in the virus. I did this by assuming that in the absence of mutation, the degree of structure is determined by the product of the migration rate and the effective population size ($F_{ST} = 1/(1 + 4N_e m)$, where N_e is the effective population size and m the migration rate). I found no evidence in my RNA sequences of multiple viral strains infecting the same fly, so the virus can be regarded as a haploid fly gene. Therefore, the effective population size of the virus is the product of the effective population size of *Drosophila* and the proportion of flies that are infected. Based on an estimate of F_{ST} in fly populations (Caracristi & Schlotterer 2003), I calculated that a virus that infects 4% of European flies would be expected to have the same level of structure as I see in the samples. This closely matches the average 3% sigma infection rate I observed across Europe. Therefore, the high structure in the viral population may simply reflect the low viral prevalence.

I have made a number of assumptions in these calculations. First, I have assumed that the rate of transmission through males and females is the same. In reality, transmission occurs at a lower rate through males than females. This will further reduce the viral effective population size and therefore increase the level of structure. Second, I have ignored the effects of mutation. It is striking that the viral isolates within each population form a monophyletic group and so there is no evidence for migration between populations in our sample. In isolated populations, mutation can substantially reduce estimates of population structure (Hedrick 1999). If this is the case, then the low prevalence of the virus may not fully explain the extreme structuring of the viral population. Third, I have assumed that the effective population size of the virus is determined by the arithmetic mean of the prevalence across populations. However, the prevalence is actually very variable, which may reflect temporal fluctuations in prevalence of the virus. If this is the case, the virus's true effective population size is better reflected by the harmonic mean of the prevalence estimates—harmonic means are disproportionately affected by the small values and so better reflect populations that periodically experience very small population sizes. As the harmonic mean is lower than the arithmetic mean, this would tend to further

increase levels of structure. Finally, it is worth remembering that I did not sample the virus from the same populations as Caracristi & Schlotterer (2003) sampled their flies.

In conclusion, the data indicate that there is limited migration and genetic exchange between European populations of *D. melanogaster*. In studies of the hosts themselves, this is disguised by large populations, which prevents the appearance of genetic structure. Therefore, vertically transmitted pathogens with smaller populations and higher mutation rates have the potential to be useful tools in revealing structure and migration in the host population.

4 Genotype-by-genotype interactions between *Drosophila melanogaster* and the sigma virus

The experimental work and statistical analysis described in this Chapter were carried out by myself, with the exception of the analysis used to estimate variances and covariance and eigenanaylsis, which was written by Jarrod Hadfield.

4.1 Introduction

By reducing the fitness of their hosts, pathogens impose strong selection on hosts to resist pathogen infections. At the DNA level, this selection for resistance is expected to remove genetic variation affecting susceptibility to pathogens. Yet, nonetheless, there remains a great deal of genetic variation for susceptibility to pathogens in many natural host populations (Hirschhorn & Daly 2005, Holub 2001, Burdon 1987, Chaboudez & Burdon 1995, Dinoor 1977, Riehle *et al.* 2006, Henter 1995, Henter & Via 1995, Lazzaro *et al.* 2004, Tinsley *et al.* 2006, Kraaijeveld & van Alphen 1995, Kraaijeveld & Godfray 1997, Carius *et al.* 2001, Dybdahl & Lively 1998, Ferrari *et al.* 2001). How this variation is maintained remains a central question in evolutionary biology.

One potential source of variation is deleterious mutations: alleles destined to be lost from the population but maintained at any one time by mutation-selection balance. Alternatively, variation could be transient, and exist because a selective sweep is in progress. Under this scenario, variation is maintained because although these alleles confer resistance, they never reach fixation because the pathogen—the target of their resistance—is continually evolving, and so the direction in which selection is acting is continuously shifting.

Another possibility is that variation is maintained by negative frequency-dependent selection. Under this scenario, the fitness of a genotype is dependent on its frequency relative to other genotypes in the population. So new resistant alleles

increase in frequency in the population as long as resistance is beneficial. But over time, the frequency of the pathogen will decline such that, eventually, this pathogen is so rare that resistance against it is no longer advantageous. After the resistant alleles cease to be beneficial, they are expected to remain at high frequencies or drift to fixation, except if they are costly. Costs associated with resistance result in selection against these resistant alleles, preventing them from going to fixation and maintaining variation. Costs can exist either as a trade-off between resistance against different pathogen genotypes or as a trade-off with other components of fitness.

Two influential population genetic models have described how the different costs of resistance maintain variation in host susceptibility in natural populations. The first set of models propose that each host genotype is better than other genotypes at resisting a particular pathogen genotype, but is worse at resisting other pathogen genotypes. This is because trade-offs exists between resistance against different pathogen genotypes, and prevent any one host genotype resisting everything—known as the 'matching-allele concept' (Agrawal & Lively 2002, Howard & Lively 1994). The second set of models proposes that some host genotypes are intrinsically better than others at resisting pathogens but this resistance is costly and these costs prevent these genotypes from going to fixation—the so-called 'gene-for-gene concept' (Flor 1955, Agrawal & Lively 2002).

Evidence for gene-for-gene and matching-allele models exists (gene-for-gene: Flor (1955), McVey (1990), Webster *et al.* (1986) and matching-allele: Carius *et al.* (2001), Lambrechts *et al.* (2005), Salvaudon *et al.* (2007)). However, solely based on observed polymorphisms, it is difficult to infer whether host-pathogen genetics follow strict gene-for-gene or matching-allele models (Frank 1996). Measuring the strength of the trade-offs is important, because, if the trade-offs between resistance against different pathogen genotypes are small enough—even if they result in specific interactions between host and parasite—they are unlikely to prevent the evolution of general resistance, and so are unlikely to maintain the variation that we see in natural populations. If they are large, however, most of the genetic variation affecting susceptibility to pathogens will be specific to a pathogen genotype, and so there is little potential for the population to evolve general resistance to all pathogen genotypes over time.

In this chapter, I look to see whether specific interactions exist, and test the

strength of these trade offs. This will tell us whether these specific interactions might help to maintain genetic variation in susceptibility, by preventing the evolution of general resistance. To investigate this, I examine the variation in transmission of five different viral isolates in *Drosophila melanogaster* lines with different first, second and third chromosomes. My aim is three fold. First, to measure the amount of genetic variation affecting transmission rates for each chromosome separately. Second, to determine what proportion of this variation affects transmission of all five viruses, and what proportion affects transmission of particular viruses. Third, to estimate whether the evolution of general resistance is constrained by the specific interactions that I found.

4.2 Material and methods

4.2.1 Experimental overview

I examine the variation in transmission of five different viral isolates in *D. melanogaster* lines with different first, second and third chromosomes. To create these lines, first, second and third chromosomes collected from the wild were substituted onto a common isogenic background. Therefore, all first chromosomes had a common background, all second chromosomes shared a different common background to the first, and all third chromosomes were put on, yet again, a different common background. The transmission of five different viral isolates was measured from females to their offspring, and from males to their offspring, for all three chromosomes.

4.2.2 Stocks, viral isolates and general methods

The effect of the *D. melanogaster* first, second and third chromosome on sigma virus transmission was measured separately using chromosome-substitution lines (69 1st chromosomes; 77 2nd chromosomes and 67 3rd chromosomes). Each of the first chromosome-substitution lines has a different homozygous first chromosome that had been sampled from a population in California (USA) in 1998, and have been substituted into a common isogenic background. The same is true for the second chromosome (collected in Pennsylvania (USA) in 1998 and 1999) and the third

chromosome (collected in North Carolina (USA) in 1997).

A mutation in a gene called *ref(2)P* is already known to affect the susceptibility of *D. melanogaster* to the sigma virus (Bangham *et al.* 2007, Dru *et al.* 1993, Wayne *et al.* 1996). This mutation occurs on the 2nd chromosome and so, to take this into account in our experiments, 77 second chromosome-substitution lines were sequenced for this mutation. The lines that I refer to as resistant carry the d and p mutations decribed by Carre-Mlouka *et al.* (2007) and the lines I refer to as susceptible have neither of these mutations.

Because the sigma virus is only transmitted vertically, and because the chromosome-substitution lines were not infected with the sigma virus at the outset, I created fly stocks that were infected with the sigma virus, and used these to cross the virus into the fly lines. I used five different viral isolates, AP30 (collected in Florida, 2005), GC20 (collected in Spain, 2005), PF115 (collected in Greece, 2005), E27 (collect in UK, 2005) (collected in France) and described in Chapter 3, (Carpenter *et al.* 2007) and and Hap23 (supplied by D. Contamine). To create the infected stocks I took females from the five different wild caught lines infected with the virus and backcrossed them for six generations to either *SM5/Pm;spapol* males (for the second chromosome experiment) or *TM3/Tb* males (for the third chromosome experiment). For the first chromosomes experiment, I took males from the five different wild caught lines infected with the virus and crossed them to *X^X* females and backcrossed F1 females to a male (with a wild-type first chromosome) that is used to maintain the attached-X stock. *X^X* females have the same genetic background as the first-chromosome substitution lines; *SM5/Pm;spapol* has the same genetic background as the second chromosome substitution lines and is homozygous for the susceptible *ref(2)P* allele and *TM3/Tb* has the same genetic background as the third-chromosome substitution lines. I tested whether the backcrossed lines were infected with the virus after every generation by assaying half of the progeny (by gassing flies with CO_2 for 15 minutes at 12°C) and then set up the next generation from lines that were infected. After six generations of backcrossing I selected five lines that had been backcrossed to *X^X*, each line was infected with a different viral isolate. I did the same for the lines that had been backcrossed to *SM5/Pm;spapol* and *TM3/Tb*.

The isogenic P18 strain, used during the assays of transmission from homozygous parents, was generated by A. Fytrou from an isofemale line collected

in Pennsylvania and was made isogenic using standard crosses to a balancer stock (*SM1/Pm;TM6/Sb;spapol*). P18 is homozygous for the susceptible *ref(2)P* allele.

Throughout the experiments, flies were reared at a constant density. To produce the constant density bottle cultures, I washed eggs off apple juice agar-plates that had been left for 8 hours in cages with live yeast. I then pipetted 26μl (for balanced stocks) or 13μl (for other stocks) of eggs into half-pint bottles containing standard Drosophila media. The flies were reared at 25°C on a 12-hour light/dark cycle. When setting up crosses throughout the experiment, I achieved approximately standard densities of 50 offspring by keeping virgin females for a few days on food that had been lightly sprinkled with live yeast, and then setting up crosses with two females in a vial for two days, without additional yeast (2nd and 3rd chromosome-substitution lines) and lightly yeasted (1st chromosome-substitution lines that did not fair well if left totally un-yeasted).

To assay for infection by the sigma virus, adults were exposed to pure CO_2 for 15 minutes at 12°C. I then counted the number of flies that were dead two hours post-exposure. This is sufficient time that uninfected flies will have fully recovered from the anaesthesia, but flies that are infected with the sigma virus are dead or paralysed.

4.2.3 Experimental methods

I measured the variation in the transmission of five sigma viral isolates in males and females for each panel of chromosome-substitution lines separately. The sigma virus is vertically transmitted, so in each set of experiments, the panel of chromosome-substitution lines were each infected by crossing them to their respective balancer or attached-X stock (*X^X* for the 1st chromosomes-substituted lines, *SM5/Pm;spapol* for 2nd chromosomes-substituted lines and *TM3/Tb* for the third chromosomes-substituted lines). These stocks had previously been infected with the five different virus (see methods above). It is worth noting that for simplicity I refer to the viral infection as present or absent for the remainder of the methods, however, I measured the presence or absence of the five different viral isolates. These crosses are summarised in Figure 4.1, cross (a) and (b).

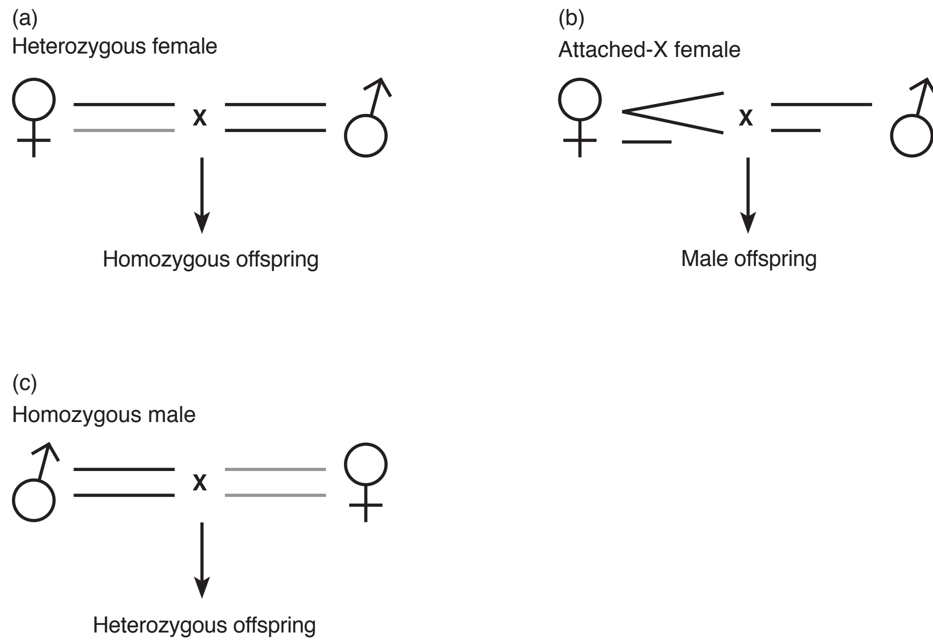


Figure 4.1: Schematic describing the three types of transmission experiment carried out. Maternal transmission in second and third chromosomes was carried out as follows: (a) Heterozygous females were crossed to males homozygous for the same wild-type chromosome. Maternal transmission in the first chromosome was carried out as follows: (b) Attached-X females were crossed to males with a wild-type chromosome. Paternal transmission from a male homozygous for the wild-type first, second or third chromosome was carried out as follows: (c) Homozygous males were mated to females from the P18 isogenic line. Grey and black are used to show when a fly is heterozygous or homozygous.

4.2.4 Experiment measuring transmission in second and third chromosome-substitution lines

The first set of experiments were carried out to assess transmission from heterozygous females to homozygous offspring, and from homozygous males to heterozygous offspring in second and third chromosome-substituted lines. Here I describe the experimental setup to measure transmission in the second chromosome-substitution lines. However, this setup is identical to the experiment measuring transmission in the third chromosome-substitution lines, except the balancer stock used to infect third chromosome-substitution lines was *TM3/Tb*, not *SM5/Pm;spapol*.

Infected *SM5/Pm;spapol* virgin females were collected from the standard-density bottle cultures. After three days, pairs of females were placed in vials with pairs of males from each of the second chromosome-substitution lines and allowed to lay for two days. Between two and four replicate crosses were set up for each second chromosome-substitution line, depending on the numbers available. After two days in the vial, the parents were removed from the vials and checked that the female parent was infected; if either female was uninfected the vial was discarded. The infected *SM5/+* female F1 progeny were aged for four days and then backcrossed to the chromosome-substitution line (+/+). Between one and four replicates were set up from each vial. As before, the flies were left to lay for two days and then the parents were removed and females checked that they were infected and vials were discarded if they were not. I collected the offspring of this cross that were homozygous for the wild-type chromosome two. These flies were genetically identical to the second chromosome-substitution lines, but have been infected with the virus from their mothers. Some males were put aside to be used to measure the rate of sigma virus transmission by infected males (see below). 15 days after the cross was set up, the remainder of the flies were tested for infection. This provided an estimate of the effect of the wild-type second chromosome on transmission from a heterozygous female (*SM5*/+) to homozygous offspring (+/+). In total I assayed 14,166 flies in 1021 vials.

To measure the rate at which infected homozygous males transmit the sigma virus to their offspring, 5-day-old infected males derived from the previous cross were mated to 4-day-old females from the isogenic P18 strain. For each vial from the previous generation, between one and four replicates were set up. After two days in

the vial, the parents were removed from the vial and checked that the female parent was infected and vials were discarded if either of the male parents were not infected. 15 days after the crosses were set up, the progeny were assayed for sigma infection. In total I assayed 57,559 flies in 1703 vials.

In the experiment measuring transmission in the third chromosome-substitution lines from a heterozygous female (*TM3/+*) to homozygous offspring (+/+), I assayed 7,756 flies in 645 vials, and from a homozygous males (+/+) to heterozygous offspring (P18/+) I assayed 31,791 flies in 1092 vials.

4.2.5 Experiment measuring transmission in first chromosome-substitution lines

I used an attached-X stock (*XˆX*) to cross the virus into the first chromosome-substitution lines, and because all of the female offspring produced from an attached-X female will also have attached-X chromosomes, I was unable to measure transmission from heterozygous females to homozygous progeny as I did for the other two chromosomes. Instead, I measured transmission from attached-X females crossed to males from each of the first chromosome-substitution lines and measured the proportion of the male offspring infected. Although this cross is uninformative about the effect of the first chromosome on maternal transmission (as all females are identical *XˆX*), it does provide a measure of the effect of wild-collected 1st chromosome on susceptibility of the zygote to infection, as male progeny from this cross have a wild-collected first chromosome. I then went on to measure transmission from homozygous males (+/+) to heterozygous progeny (P18/+) in the first chromosome-substitution lines. These crosses are summarised in Figure 4.1, cross (b) and (c).

To do this I collected infected *XˆX* virgin females from the standard-density bottle cultures. After four days, pairs of females were placed in vials with pairs of males from each of the chromosome-substitution lines and allowed to lay for two days. Between one and four replicate crosses were set up for each chromosome-substitution line, depending on the numbers available. After two days in the vial, the parents were removed from the vials and checked that the female parent was infected; if either female was uninfected the vial was discarded. I collected the male offspring from this cross. Some males were put aside to be used to measure the rate of sigma

virus transmission by infected males (see below). 15 days after the cross was set up, the remainder of the flies were tested for infection. This provided an estimate of the effect of the wild-type first chromosome on the susceptibility of hemizygous male offspring.

The infected male F1 progeny were aged for four days and then crossed to 5-day-old females from the isogenic P18 strain. For each vial from the previous generation, between one and eight replicates were set up. After two days in the vial, the parents were removed from the vial and checked that the female parent was infected and vials were discarded if either of the male parents were not infected. 17 days after the crosses were set up, the progeny were assayed for sigma infection. In total, from *X^X* females to hemizygous males I assayed 1,579 flies in 220 vials, and from homozygous males (+/+) to heterozygous offspring (P18/+) I assayed 71,440 flies in 1260 vials.

4.3 Statistical anaylsis

The statistical analysis was carried out using R (v.2.2.1) and ASReml software and language. Viral transmission rates were measured as a ratio of the number of dead flies to the number of surviving flies as a combined dependent variable and the data follow a binomial distribution. The factors affecting variation in transmission of the sigma virus were determined using a generalised linear mixed model (GLMM). GLMM are difficult to fit in a frequentist framework because the likelihood cannot be obtained analytically. So instead, the likelihood of the models was determined using Markov Chain Monte Carlo (MCMC) techniques. In short, each binomial data point is associated with a linear predictor which is the probability of death on the logit scale. These linear predictors are assumed to be normally distributed with means determined by the fixed effect and (co)variances determined by the random effects and a residual term that accounts for extra-binomial variation. Because no vial was treated with more than one virus, residual covariances were set to zero, and only the covariances due to line were estimated. These (co)variance matrices were estimated using the multivariate extension to the Gibbs sampling method of Garcia-Cortes & Sorensen (2001), where missing data were augmented to make the multivariate design balanced. The linear predictors were sampled using Metropolis-Hastings steps using a normal distribution multiplied by the binomial likelihood. Improper priors were used, and each chain (i.e. for each chromsome) was run for 1.8 million iterations with a

burn-in of 300,000 and a thinning interval of 1500.

Maternal transmission: When measuring maternal-transmission, I found almost no variation in the proportion of offspring infected with sigma virus from their mother for first and third chromosome lines across all viruses. Therefore no further analysis was carried out. For the 2nd chromosome the effect of *ref(2)P* on the mean transmission was determined using the Wilcoxon sign-rank test.

Paternal transmission: The paternal transmission datasets were analysed with a generalised linear mixed model as described above. I have used the following mixed-effects model to describe paternal transmission in first chromosome substitution lines:

Let $v_{i,j,k,l}$ be the ratio of dead to alive flies for virus treatment i on day j from line k.

$$v_{i,j,k,l} = \mu + \beta_i + \alpha_j + \kappa_k + (\beta\kappa)_{i,k} + \epsilon_{i,j,k,l} \tag{4.1}$$

where μ is the mean ratio of dead to alive flies, β_i represents the fixed effect of virus treatment $i = 1, \ldots, 4$, α_j represents the fixed effect of day $j = 1, \ldots, 4$, κ_k is a random variable representing the deviation for line k and $\epsilon_{i,j,k,l}$ is a random variable representing the deviation for observation l from virus i, day j and line k. The model allows an interaction between line and virus.

I have used the following mixed-effects model to describe paternal transmission in second chromosome substitution lines:

Let $v_{i,j,k,1,l}$ be the ratio of dead to alive flies for virus treatment i on day j for line k for *ref(2)P* allele 1.

$$v_{i,j,k,1,l} = \mu + \beta_i + \alpha_j + \kappa_k + \delta_1 + (\beta|\kappa)_{i,k} + \epsilon_{i,j,k,1,l} \tag{4.2}$$

where μ is the mean ratio of dead to alive flies, β_i represents the fixed effect of virus treatment $i = 1, \ldots, 4$, α_j represents the fixed effect of day $j = 1, \ldots, 5$, κ_k is a random variable representing the deviation for line k, δ_1 and δ_2 are the effects of the presence or absence of resistant *ref(2)P* allele respectively, and $\epsilon_{i,j,k,1,l}$ is a random variable

representing the deviation for observation l from virus i, day j, line k and allele 1. The model allows an interaction between line and virus. A separate model in which an interaction between *ref(2)P* and one of the viral lines was permitted to account for the variation in the Florida line. The other viral lines showed no sensitivity for the resistant allele.

I have used the following mixed-effects model to describe paternal transmission in third chromosome substitution lines:

Let $v_{i,j,k,m,l}$ be the ratio of dead to alive flies for virus treatment i on day j for line k for vial m.

$$v_{i,j,k,m,l} = \mu + \beta_i + \alpha_j + \kappa_k + \omega_m + (\beta|\kappa)_{i,k} + \epsilon_{i,j,k,m,l} \quad (4.3)$$

where μ is the mean ratio of dead to alive flies, β_i represents the fixed effect of virus treatment $i = 1, \ldots, 4$, α_j represents the fixed effect of day $j = 1, \ldots, 5$, κ_k is a random variable representing the deviation for line k, ω_m is a random effect representing the deviation for vial, and $\epsilon_{i,j,k,m,l}$ is a random variable representing the deviation for observation l from virus i, day j, line k and vial m. The model allows an interaction between line and virus.

4.3.1 Estimating variance and covariance

These models were used to estimate the mean transmission for each virus, and the genetic variance (V_g) and covariance (*Cov*) among the lines. These variances and covariances are used to create a symmetrical, square matrix with one row and one column per trait (in this study, one row and one column per viral isolate). The diagonals of this matrix indicate how much genetic variance underlies each trait (i.e., transmission of each virus), while the off-diagonals indicate the genetic covariance between each pair of traits (i.e., how transmission of one virus covaries with transmission of another virus). This genetic variance-covariance matrix is known

as a G matrix and is summarised below:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} \mathbf{E} = \begin{bmatrix} E_1 & & & & \\ & E_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & E_n \end{bmatrix} \mathbf{G} = \begin{bmatrix} V_{g1} & & & & \\ Cov_{1,2} & V_{g2} & & & \\ Cov_{1,3} & Cov_{2,3} & \cdot & & \\ Cov_{1,4} & Cov_{2,4} & Cov_{3,4} & \cdot & \\ Cov_{1,5} & Cov_{2,5} & Cov_{3,5} & Cov_{4,5} & V_{gn} \end{bmatrix} \quad (4.4)$$

β is the means of the traits, E is the residual error, G is the G-matrix, comprised of V_g the variance around each trait mean and the Cov, the covariance between each trait mean. Confidence limits for these estimates were calculated by running the models 1000 times and taking the 95% distribution of the limits using Markov Chain Monte Carlo methods.

4.3.2 Eigenanaylsis

When two or more traits are measured, it becomes increasingly difficult to interpret patterns of covariation among traits. In these instances, eigenanalysis (e.g. principle component analysis) is used to analyse G matrices. This involves generating two new variables: eigenvectors, which are the linear combinations of the original traits (i.e. the variance and covariances); and eigenvalues, which are the length of the eigenvectors (i.e. the amount of genetic variance associated with each vector). The motivation behind eigenanalysis is to reduce the dimensionality of G, which allows the significance of each of the eigenvectors to be assessed, which in turn, allows the influence of each of the traits on the phenotype to be inferred. Why is this important? Because it gives us some measure of the evolutionary potential of those traits. In effect, it illuminates whether these traits are selected in the same or opposing directions, and, hence, whether they ultimately lead to greater evolutionary trade-offs.

4.4 Results

4.4.1 Genetic variance in maternal transmission to homozygous offspring

I measured the transmission from an infected mother heterozygous for either a wild-type second or third chromosome to offspring that were homozygous for that chromosome (Figure 4.1). For the third chromosome, there is very low variation in the proportion of offspring infected with sigma virus across all viruses (a mean of 98% of offspring were infected, across all viruses). For the second chromosome, there is very low variation in the proportion of offspring infected for all viruses except for the Florida isolate, which has a bimodal pattern of transmission—lines with the *ref(2)P* susceptible allele had a mean rate of transmission of 94%, while the subset of lines carrying the *ref(2)P* resistant allele had a mean transmission rate of 7% (Wilcoxon signed-rank test $W = 4890.5, n_{susceptible} = 126, n_{resistant} = 41, P < 0.001$) (Figure 4.2).

4.4.2 Genetic variance in maternal transmission to hemizygous offspring

For the first chromosome, I measured transmission from an attached-X female to male offspring carrying the wild-type first chromosome (Figure 4.1). This estimates the effect of the wild-type first chromosome on susceptibility of the zygote to infection. Each of the viral strains tested, had a high transmission rate (a mean of 98% across all viruses, and for each of the viral strains, there was very low variation among chromosome lines (Figure 4.2).

4.4.3 Genetic variance in paternal transmission to heterozygous offspring

In the next experiment I examined viral transmission, for each of the five viruses, through sperm rather than eggs. I examined paternal transmission from males hemizygous for the wild-type first, and homozygous for the second or third chromosome to offspring that were heterozygous for that chromosome (Figure 4.1). As for maternal transmission, the resistant *ref(2)p* allele has a significant effect on the rate of transmission of the Florida lines—the *ref(2)P* susceptible allele had a mean rate of transmission of 23%, while the subset of lines carrying the *ref(2)P* resistant allele had a mean transmission rate of 8% (Wilcoxon signed-rank test $W = 2410, n_{susceptible} = 217, n_{resistant} = 16, P < 0.01$) (Figure 4.3).

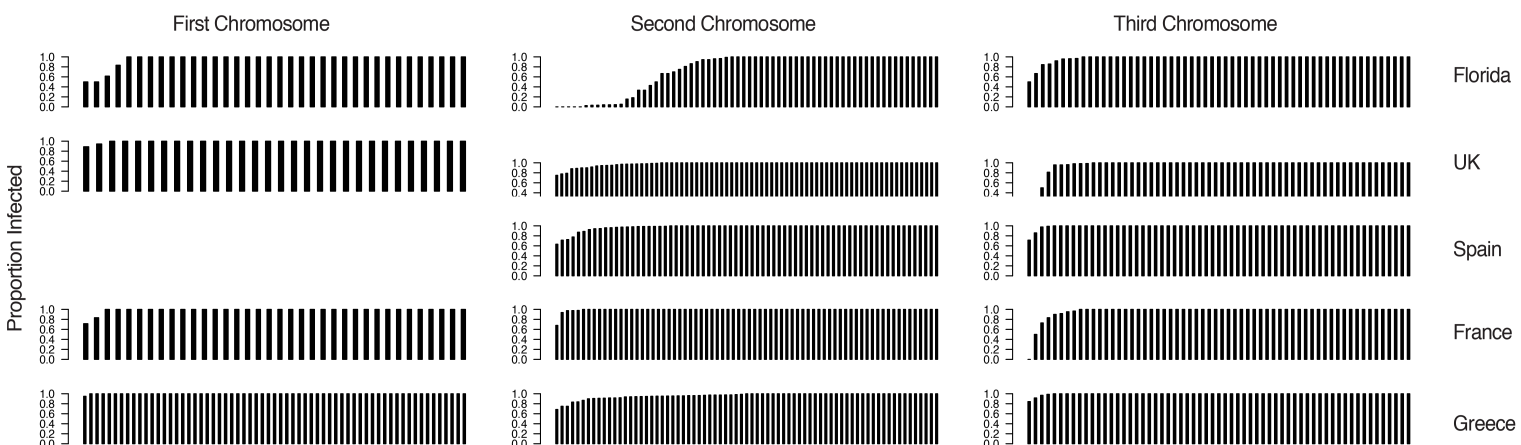


Figure 4.2: Sigma virus maternal-transmission rates across chromosome-extracted lines of *D. melanogaster* for five viruses. For the first chromosome transmission is from attached-X females to male offspring, while for the second and third chromosomes, transmission is from heterozygous females to homozygous offspring. The bars show the means with equal weights to each vial.

However, in contrast to maternal transmission, and aside from the effect of *ref(2)p* there was substantial variation in transmission rates among the different chromosome-substitution lines (Figure 4.3). This is reflected in the variances in Table 4.1, which shows that there is variation in susceptibility of flies to the different viruses (diagonals in matrices are significantly greater than 0). However, there is greater variation in transmission rates for the second and third chromosomes compared to the first chromosome. The first chromosome has about half the genes of chromosomes two and three, and if resistance were caused by many small-effect genes, evenly distributed along the chromosome, it would be expected to have lower variation than the other chromosomes.

From these genetic variance (V_g) and residual variance (V_r), I estimated heritability (the degree of genetic determination) for the paternal transmission rate using the following equation:

$$Heritability = \frac{V_g}{(V_g + V_r)} \tag{4.5}$$

The high heritability of the second and third chromosomes shows that a large proportion of the phenotypic variation measured in this experiment is determined genetically (Table 4.2).

4.4.4 Genetic covariance between viruses in rates of transmission

To investigate whether trade-offs exist between resistance against different viral isolates I examined the amount of genetic variation that is general to all the viruses by calculating eigenvalues for each of the five principle components. I found that for the second and third chromosomes most of the variation is explained by the first eigenvector (70% for 2nd and 80% for 3rd) and that the viruses are contributing more or less equally to this first vector (Table 4.3 and Figure 4.4). This indicates that most of the variation is general. In contrast, for the first chromosome, less of the variation is explained by the first eigenvector (only 50%) and the viruses are not contributing equally to each of the eigenvectors. In fact, a single virus is contributing a significant amount of the variation to each of the five eigenvectors in turn, indicating that most of the variation is not general, but is instead associated with a specific virus.

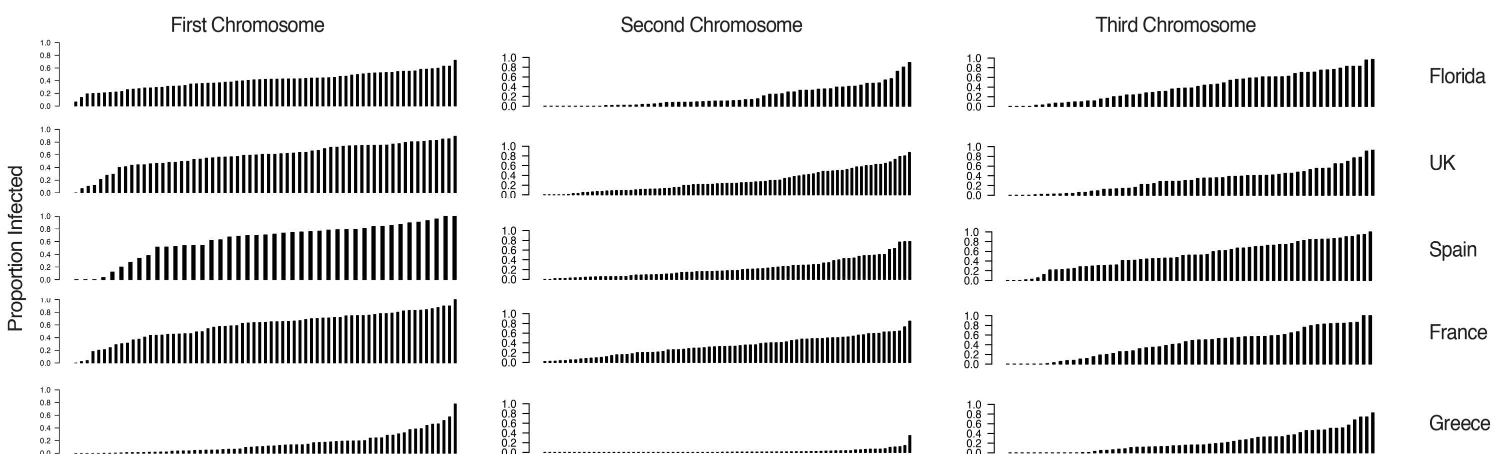


Figure 4.3: Sigma virus paternal-transmission rates across chromosome-extracted lines of *D. melanogaster* for five viruses. Transmission is from males to heterozygous offspring for all chromosomes. The bars show the means with equal weights to each vial.

Table 4.1: G-matrices for chromosome one, two and three. 95% confidence limits in brackets.

Chromosome one

	[USA]	[UK]	[SPA]	[FRA]	[GRE]
[USA]	**0.296** (0.138, 0.643)				
[UK]	0.066	**0.437** (0.181, 0.914)			
[SPN]	-0.011	-0.014	**0.513** (0.129, 1.800)		
[FRA]	-0.012	0.041	0.014	**0.493** (1.164, 1.122)	
[GRE]	0.094	0.196	0.125	0.193	**1.721** (0.777, 2.841)

Chromosome two

	[USA]	[UK]	[SPA]	[FRA]	[GRE]
[USA]	**3.589** (1.887, 6.553)				
[UK]	1.160	**1.877** (0.79, 3.577)			
[SPN]	0.995	1.380	**1.413** (0.692, 2.493)		
[FRA]	1.221	0.836	0.688	**0.718** (0.219, 1.687)	
[GRE]	1.272	0.944	0.815	0.606	**0.852** (0.234, 2.44)

Chromosome three

	[USA]	[UK]	[SPA]	[FRA]	[GRE]
[USA]	**3.149** (0.314, 8.641)				
[UK]	2.352	**2.767** (0.461, 7.355)			
[SPA]	2.181	2.113	**2.339** (0.449, 6.277)		
[FRA]	1.606	1.588	1.488	**1.529** (0.242, 4.539)	
[GRE]	1.469	1.510	1.345	1.122	**1.127** (0.269, 3.278)

Table 4.2: Heritability for the paternal transmission rate for chromosome one, two and three. 95% confidence limits in brackets.

Chromosome	USA	UK	SPA	FRA	GRE
First	0.243 (0.119, 0.415)	0.235 (0.105, 0.411)	0.152 (0.038, 0.423)	0.206 (0.068, 0.402)	0.418 (0.218, 0.575)
Second	0.827 (0.688, 0.9168)	0.738 (0.519, 0.854)	0.719 (0.506, 0.854)	0.363 (0.13, 0.585)	0.245 (0.063, 0.540)
Third	0.796 (0.244, 0.925)	0.752 (0.290, 0.898)	0.587 (0.191, 0.787)	0.602 (0.182, 0.839)	0.452 (0.148, 0.729)

Although it is not possible to visualise all five eigenvectors simultaneously, the first three eigenvectors and the eigenvalues associated with them are used to define a subspace in Figures 4.5 and 4.6. The plots show the subspace defined by a chromosome (red sphere), and present the variation for another chromosome that lies in the same direction as the vectors used to describe the subspace for the first chromosome (blue sphere).

4.4.5 Response to selection on each of the 5 viruses

By applying the multivariate breeder's equation, $\Delta z = G\beta$, where G is the genetic variance-covariance matrix (Table 4.1), β, is the vector of selection; and Δz, the vector of predicted response to selection, the response—the change in rate of transmission of each of the five viruses—can be investigated. This is done by exerting a selection pressure on any one virus and examining the change in the trait mean of both the virus under selection, and the correlated response to selection in the other viruses.

For the second and third chromosome, selection on each virus in turn (selection gradient = 1) indicates that no antagonism exists between resistance to each of the viruses, as all of the viruses show a positive change in their trait mean, although the relative response varies across viruses (Table 4.4). However, for each of the viruses in turn, if we compare the response of a virus when it is being directly selected on, to its response when another virus is being selected on, we can see that its correlated response is considerably less than expected from the direct response.

In contrast, the response to selection is generally smaller for the first chromosome compared to the second and third. This is reflected in the lower predicted response to selection for each virus for the first chromosome compared to the second and third, when all five viruses are selected on simultaneously (Table 4.5 and 4.6). The limited response of the first chromosome might reflect the smaller genetic variances for this chromosome. However, for a few of the viruses, the correlated responses to selection are negative, indicating that there may be some antagonism between resistance to each of the viruses.

Table 4.3: Eigenvectors, which are the linear combinations of the original traits (i.e. the variance and covariances) for each of the viruses for the three chromosomes separately. The columns are the first five principle components (,PC)

Chromosome one

	[,PC1]	[,PC2]	[,PC3]	[,PC4]	[,PC5]
[USA,]	-0.06586147	-0.15807610	0.14825350	-0.30826322	0.92394204
[UK,]	-0.14686333	-0.36319455	-0.02079739	-0.84952229	-0.35270416
[Spain,]	-0.09447109	0.91484807	-0.06352546	-0.38616455	0.03113957
[France,]	-0.14888591	-0.07761139	-0.97540043	0.02436363	0.14074756
[Greece,]	-0.97108271	-0.01145152	0.14881831	0.18321851	-0.03393119

Chromosome two

	[,PC1]	[,PC2]	[,PC3]	[,PC4]	[,PC5]
[USA,]	0.6730649	0.67538779	0.006387043	-0.101005827	-0.2838872
[UK,]	0.4437865	-0.56306384	-0.626952761	0.009063303	-0.3047306
[Spain,]	0.3769422	-0.47096755	0.740673912	-0.273456140	-0.1128187
[France,]	0.3075110	-0.01206742	-0.185852037	-0.405463821	0.8404460
[Greece,]	0.3367324	-0.06967345	0.154112355	0.866335102	0.3278259

Chromosome three

	[,PC1]	[,PC2]	[,PC3]	[,PC4]	[,PC5]
[USA,]	-0.5383481	0.8287809	0.03798865	0.1471808	-0.01407869
[UK,]	-0.5121456	-0.3263018	-0.75321029	0.1382943	-0.21161994
[Spain,]	-0.4676535	-0.1689422	0.25219773	-0.8288220	-0.04700145
[France,]	-0.3555755	-0.3499164	0.59620011	0.4766457	-0.41046181
[Greece,]	-0.3205535	-0.2359371	0.11033067	0.2123100	0.88562450

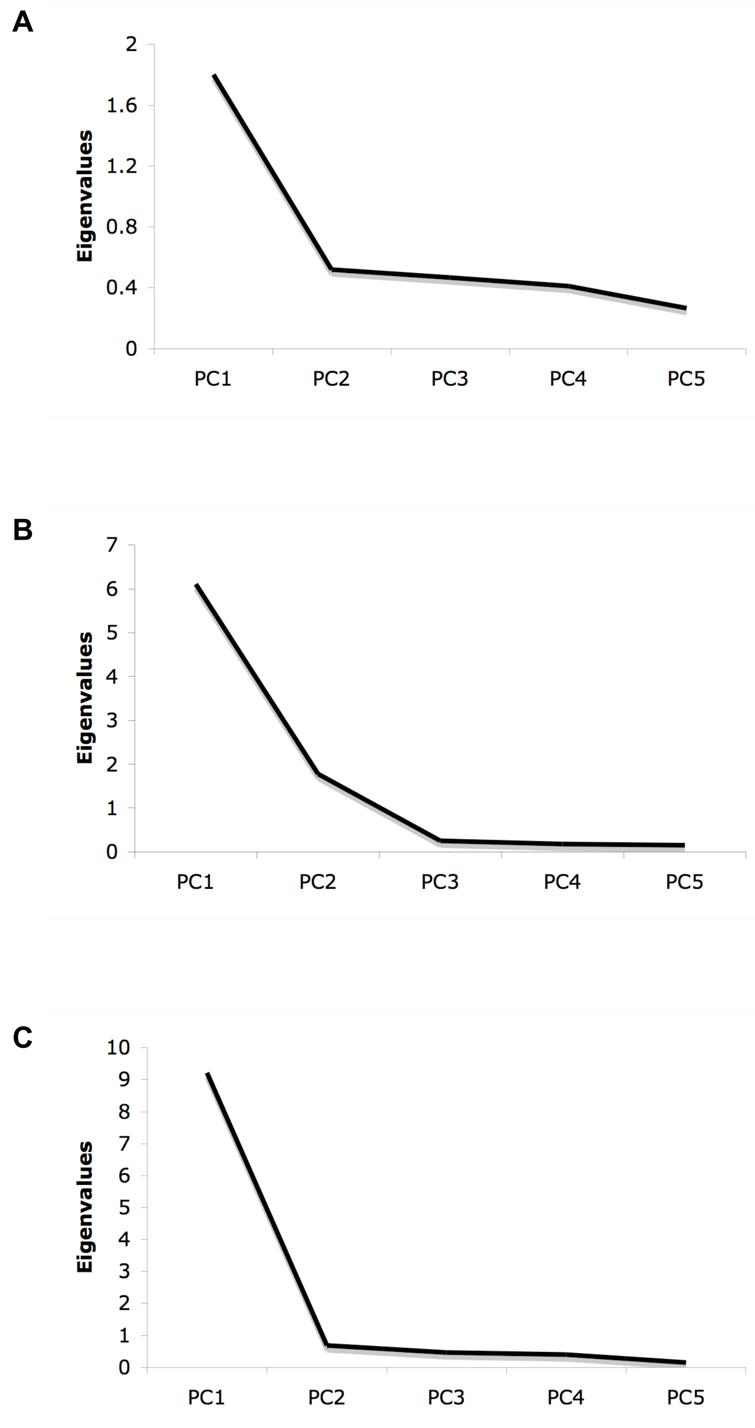
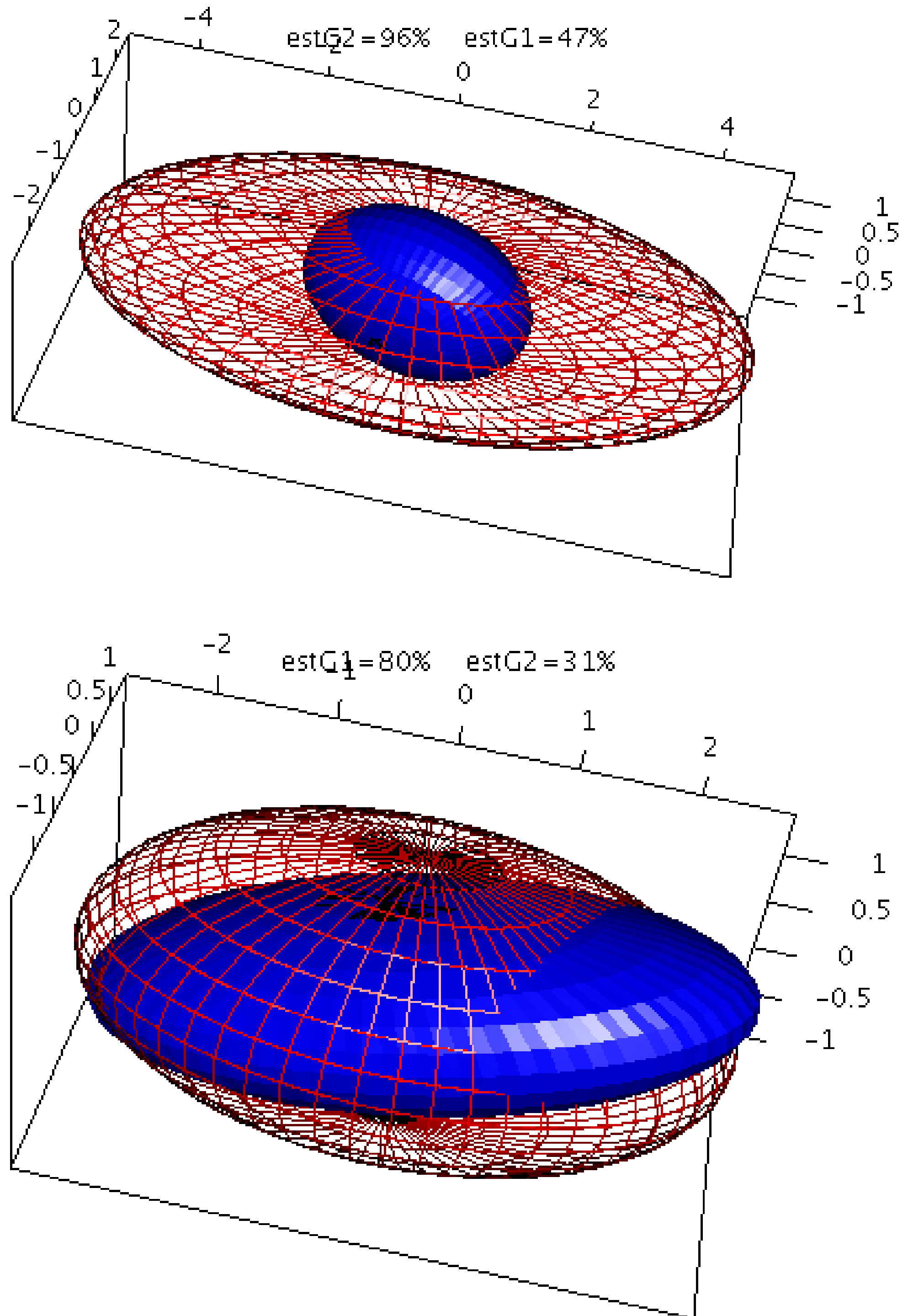


Figure 4.4: The Scree plot displays the eigenvalues for chromosome one (A), two (B) and three (C). After the curve starts to flatten out, the corresponding components may be regarded as explaining little of the total genetic variance. For chromosome two and three, the curve flattens out after the second and first principle component respectively. In contrast, all five principle components seem to be important for explaining the variation in resistance on the first chromosome.

Figure 4.5: A plot of 2 matrices in the subspace defined by the first (red) matrix. The first and second matrices can be identified from the legend, where estG is prefixed by the chromosome. The % signs refer to the amount of variance show in the subplots for the two matrices.



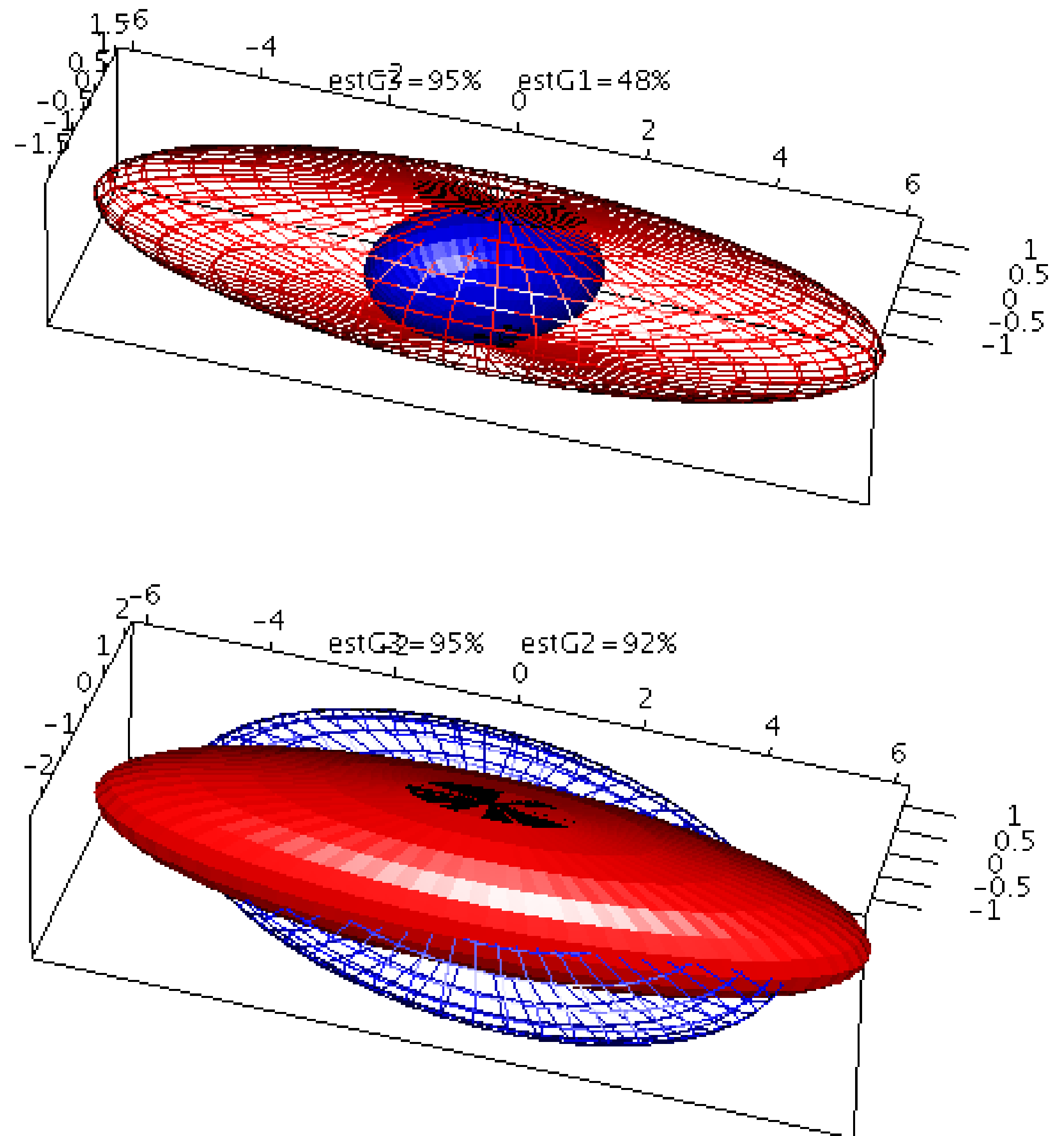


Figure 4.6: A plot of 2 matrices in the subspace defined by the first (red) matrix. The first and second matrices can be identified from the legend, where estG is prefixed by the chromosome. The % signs refer to the amount of variance show in the subplots for the two matrices.

Table 4.4: Response to selection on each of the 5 viruses (S=1.0) using the breeder's equation.

Chromosome one

	Viruses selected on				
	Florida	UK	Spain	France	Greece
Florida	0.297	0.067	-0.012	-0.012	0.095
UK	0.067	0.437	-0.015	-0.015	0.197
Spain	-0.012	-0.015	0.514	0.514	0.126
France	-0.012	0.042	0.015	0.015	0.193
Greece	0.095	0.197	0.126	0.126	1.722

Chromosome two

	Viruses selected on				
	Florida	UK	Spain	France	Greece
Florida	3.590	1.161	0.995	1.221	1.272
UK	1.161	1.877	1.381	0.836	0.945
Spain	0.995	1.381	1.413	0.689	0.816
France	1.221	0.836	0.689	0.719	0.607
Greece	1.272	0.945	0.816	0.607	0.852

Chromosome three

	Viruses selected on				
	Florida	UK	Spain	France	Greece
Florida	3.150	2.352	2.181	1.606	1.470
UK	2.352	2.768	2.113	1.588	1.511
Spain	2.181	2.113	2.339	1.488	1.345
France	1.606	1.588	1.488	1.530	1.123
Greece	1.470	1.511	1.345	1.123	1.127

Table 4.5: Response to selection on all 5 viruses simultaneously (S=0.2) using the breeder's equation.

Virus	Chromosome one	Chromosome two	Chromosome three
Florida	0.087	1.648	2.152
UK	0.145	1.240	2.066
Spain	0.126	1.059	1.893
France	0.146	0.814	1.467
Greece	0.467	0.898	1.315

Table 4.6: Response to selection on all 5 viruses simultaneously (S=1.0) using the breeder's equation

Virus	Chromosome one	Chromosome two	Chromosome three
Florida	0.435	8.239	10.759
UK	0.727	6.200	10.332
Spain	0.628	5.294	9.467
France	0.731	4.071	7.335
Greece	2.333	4.491	6.576

4.4.6 The effect of *ref2p* on paternal transmission

Because the resistant *ref(2)P* allele was associated with a 14% drop in the rate of transmission of the Florida virus, I have examined whether the interaction between *ref(2)P* and the Florida isolate might account for the remaining variation on the second chromosome, which is unexplained by the first eigenvector. To investigate this, paternal transmission rates were re-anaylsed with a model allowing an interaction between *ref(2)P* and the Florida viral isolate and the estimated genetic variation in transmission rates was compared with those estimated from the former model that did not allow this interaction. I found no difference in estimates of the amount of genetic variation in transmission rates between these two models, indicating that this variation is not attributable to the *ref(2)P*-Florida interaction (compare Table A1 in Appendix with Table 4.1).

4.5 Discussion

4.5.1 Transmission rates of viruses

I found that genetic variation affects the transmission of the sigma virus through both eggs and sperm. And that a polymorphism in *ref(2)P*—a gene already well known to affect resistance to sigma—affected the rate of transmission of only one of the viral isolates (Bangham *et al.* 2007, Contamine *et al.* 1989, Dru *et al.* 1993, Fleuriet 1988, Wayne *et al.* 1996). For the Florida viral isolate, maternal transmission was strongly affected by the gene *ref(2)P*, whereas paternal transmission was only weakly affected by the variation in *ref(2)P*. The resistant *ref(2)P* allele was associated with a 87% drop in maternal transmission (consistent with 91% drop in the rate of transmission of the Florida isolate measured in a previous study (Bangham *et al.* 2008*a*)). In this study

the *ref(2)P* polymorphism was associated with a 15% drop in paternal transmission, while previous studies found no affect of *ref(2)P* on paternal transmission.

Aside from the effect of *ref(2)P*, there is substantial genetic variation affecting paternal transmission rates of all five viruses but very little genetic variation affecting maternal transmission. All three chromosomes carried genetic variation affecting transmission, although chromosome one has about half as much as chromosomes two and three, as would be expected for a chromosome that has about half the genes. The considerable genetic variation affecting paternal transmission and the correlation between chromosome size and the amount of genetic variation, suggests that resistance is controlled by many genes spread evenly across the chromosomes.

But why does some genetic variation only affect paternal, and not maternal, transmission of the sigma virus? One possibility might lie in the differences between male and female gametes. Male gametes deliver smaller quantities of virus to offspring than females (Brun & Plus 1998). Therefore, genes that cause variation in viral titre could have a larger effect when the virus is transmitted through males gametes, than through female gametes. Therefore, genes involved in blocking male transmission of the virus are more likely to affect male-specific processes controlling the transmission of the virus into the sperm. It is unlikely that genes blocking male transmission are simply slowing down viral replication in the gametes, as these same genes would be expected to delay the rate at which flies become infected after injection. Indeed, a previous study found low genetic correlation between genes involved in combating infection rate after injections and genes involved in male-transmission, suggesting that genes involved in these two aspects of sigma infection are quite distinct (Bangham *et al.* 2008*a*).

4.5.2 Patterns of genetic variation

I found considerable genetic variation on chromosomes one, two and three of *D. melanogaster* in transmission rates of the sigma virus. These results, along with the results from two previous studies, indicate that resistance to the sigma virus is a trait with high levels of genetic variation (Bangham *et al.* 2007, 2008*b*). Estimates of heritability for male transmission are high for the second and third chromosomes (57%, and 63% respectively) but lower for the first chromosome (25%), which has about half the genes of chromosomes two and three, and is therefore expected to have

lower genetic variance and heritability. What's more, estimates of heritability for male transmission for this study are likely to underestimate true heritability, as transmission rates were measured for each chromosome separately. By adding together the variation associated with each chromosome, the amount of genetic variation would increase (would roughly treble), while the residual variation would remain the same, and so the heritability of male transmission would increase. Although this study measured total (rather than additive) genetic variation, it is clear that resistance to the sigma virus is a trait with high levels of genetic variation. This indicates that these traits have the potential to respond to selection.

Going on to the genetic variation that affects specific viruses: with the exception of *ref(2)P*, the genetic variation that affects particular viral isolates is spread evenly across all three chromosomes. This suggests that many genes underlie specific resistance to the five viral isolates.

Why does the *ref(2)P* polymorphism affect transmission rates of only the Florida virus isolate, and not the others? One possibility is that the Florida viral isolate may be a remnant viral type, which existed before a new viral type spread in response to the increasing frequency of the resistant *ref(2)P* allele in *D. melanogaster* populations. It is thought that sigma populations experience frequent selective sweeps in response to spreading resistant alleles in the fly populations. Indeed, in the last twenty years, European sigma populations have undergone replacement of one viral type, which was sensitive to the resistant *ref(2)P* gene, by an insensitive viral type (Fleuriet 1990, Fleuriet & Sperlich 1992). Viruses from before the sweep would be affected by the *ref(2)P* polymorphism, while viruses that have themselves swept in response to the resistant *ref(2)P* allele will not. In support of this idea, I have shown (Chapter 3) that the Florida viral isolate is phylogenetically distinct from the other viruses used in this study (Carpenter *et al.* 2007).

4.5.3 What is maintaining variation in resistance against the sigma virus?

Sigma viral infections are costly to the fly: sigma virus causes lower female fecundity and reduces over-wintering survival; and its low prevalence, given its rate of transmission, means that it must carry a cost. Therefore resistance to the sigma virus is expected to spread through the population. This raises the question of what maintains genetic variation in resistance in *D. melanogaster* populations? One possibility is

trade-offs exist between resistance against different pathogen genotypes, preventing any one host genotype from going to fixation, thereby maintaining variation.

In this study, I measured the transmission rates of five different viruses to investigate whether trade-offs exist between resistance against these different viruses. I found that most of the genetic variation on chromosome two and three that affects transmission rates is general to all five viruses. This suggests that there is little constraint on selection acting to increase resistance against all viruses, and that trade-offs between resistance against different viruses do not exist. By contrast, I found that some of the genetic variation on chromosome one affecting transmission rates is specific to single viruses. However, the robustness of these specific interactions is unclear, because the smaller genetic variances for chromosome one, make estimating variances and covariances difficult, which may conflate estimates of specificity.

In conclusion, although a small amount of genetic varaition on each chromosome is involved in specific interactions with particular viruses, most the genetic varation in transmision rates is general. Therefore, trade-offs between resistance against the different viral isolates are unlikely to be maintaining variation in resistance in *D. melanogaster* populations.

So what is maintaining the genetic variation affecting resistance? One possibility is that trade-offs between resistances to different viruses will slow the rate at which general resistance can be selected for. Because parasites evolve so rapidly, the target is continually shifting. This shifting target, combined with the slow rate of evolution towards that moving target, increases the number of transient polymorphims in the population, thereby creating variation in resistance. In support of this, I found that although selection for general resistance (to all five viruses) was not impeded, the rate at which selection can increase resistance was slowed by specific interactions with particular viruses.

5 The antiviral role of the Toll pathway against sigma

5.1 Introduction

Vertebrates rely on a sophisticated adaptive immune system to target viral infections, among other pathogens, in a highly specific manner. Invertebrates lack this adaptive immune response, yet are capable of effectively fighting viral infections. One way they do this is by RNA interference (RNAi)—a group of evolutionary conserved mechanisms that use short RNAs to recognize and degrade complementary nucleic acids. The RNAi pathway detects double stranded viral RNA and cuts it into short 21–24 nucleotide fragments called short-interfering (si)RNAs, and uses these siRNAs to target and cleave complementary viral sequence (reviewed by Ding & Voinnet (2007) and summarised in Figure 5.1).

In addition to the antiviral role of the RNAi pathway, the Toll-pathway—an important component of the innate immune system—is thought to be activated by viral infection in invertebrates. However, compared to the RNAi pathway, how the Toll-pathway recognizes and responds to viral infections is not well understood. Most of our understanding of the Toll-pathway has come from the study of its role in immune defences against bacteria and fungi, where recognition of pathogen-associated molecular patterns (PAMPs) by peptidoglycan receptor proteins (PGRPs), initiated a serine protease cascade that cleaves Spätzle. Once cleaved, Spätzle binds to the Toll receptor (a transmembrane receptor), causing the recruitment of three intracellular proteins resulting in the degradation of cytoplasmic Cactus and the release of Dif—a Dorsal-related protein. Dif then translocates into the nucleus, where it activates the transcription of antimicrobial peptides that target bacterial and fungal infections (see Lemaitre & Hoffmann (2007) and Lemaitre (2004) for review and Figure 5.1).

The first evidence that the Toll pathway is also involved in an antiviral response was demonstrated in *Drosophila*. Zambon *et al.* (2005) found that 7 antimicrobial

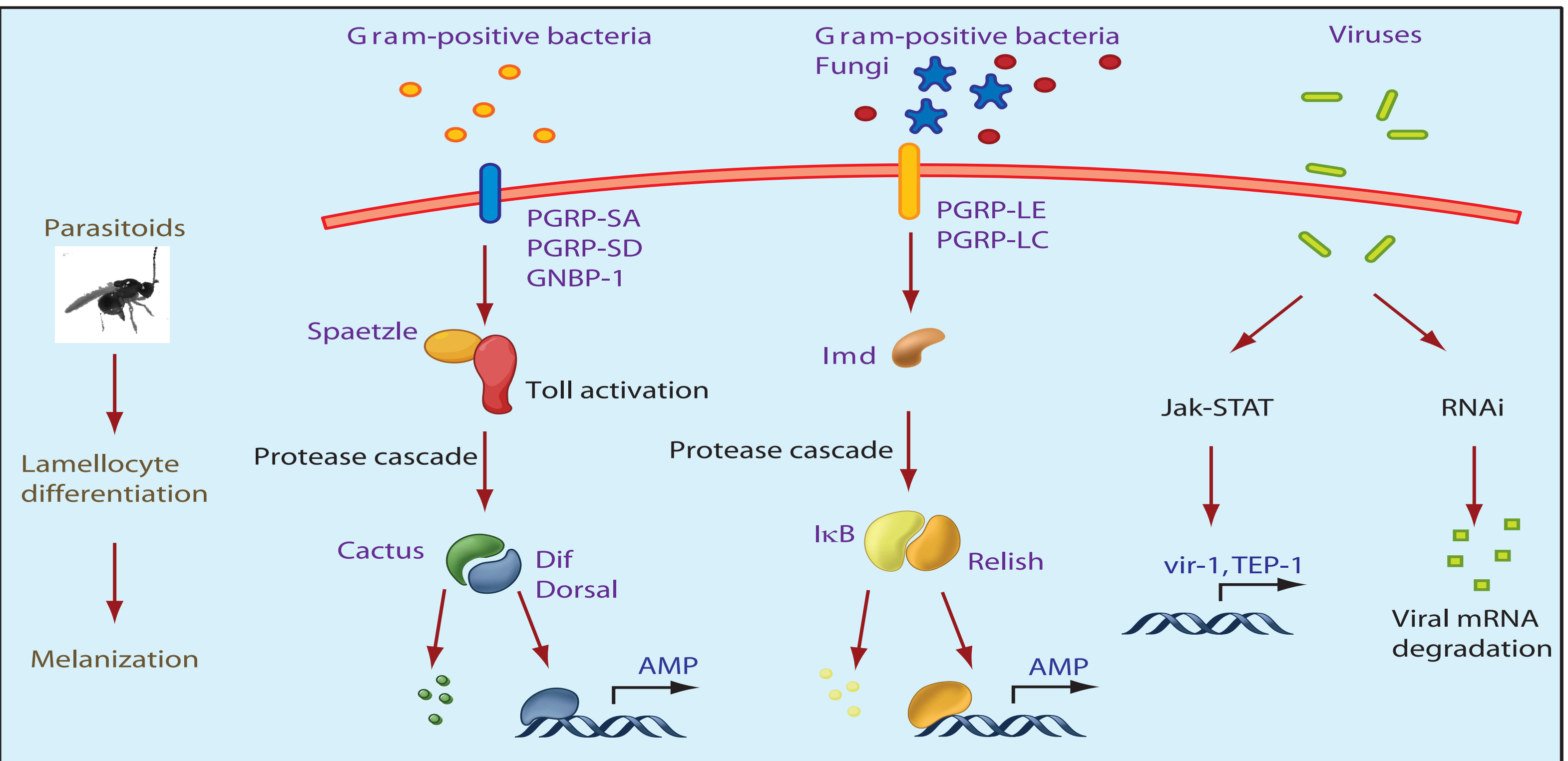


Figure 5.1: Simplified Description of Four of the Immune Responses of Drosophila melanogaster From left to right. Parasitoids lay their eggs inside the larvae or pupae of other insects and, if successful, kill their hosts. In response to such parasitization, lamellocytes differentiate and form several layers around the parasitoid egg, which is melanized to form a hard black capsule. Gram-positive bacteria and fungi trigger the activation of the Toll pathway. Peptidoglycan recognition proteins (PGRPs) and gram-negative bind-ing proteins (GNBPs) recognize the presence of Gram-positive bacteria and fungi and, through Spaetzle and Toll, activate a proteolytic cascade involving serine proteases and serine protease inhibitors. This results in the proteolytic degradation of inhibitor kB (IkB) protein Cactus and ac- tivation of the NF-kB proteins Dif and Dorsal, resulting in the transcription of antimicrobial peptides (AMPs). Gram-negative bacteria trigger the Imd pathway, which also results in a proteolytic cascade. This results in the cleavage of Relishthe C-terminal (IkB-like) part of which is removed and the N-terminal (NF-kB-like) part of which activates AMP transcription. Much less well understood are the antiviral responses of insects. Recent results indicate that viruses trigger the Jak-STAT pathway (involving a Jak kinase called Hopscotch) and the transcription of antiviral genes. RNAi-silencing machinery is also able to target animal viruses. Taken, with permission from Jenny Bangham, from (Bangham *et al.* 2006)

peptides (AMPs)—known to be induced by the Toll pathway—were upregulated after infection with the *Drosophila* X virus (DXV). However, constitutively expressing these AMPs did not result in lower viral titres, as would be expected if they were involved in an antiviral response against DXV. The role of the Toll pathway in fighting viral infections was made more ambiguous by the results of an experiment that showed that Tl^{10b}, a Toll gain-of-function mutant, succumbed to viral infection as rapidly as both wild-type flies and a Toll-knock-out mutant (Dif^{1}), although Tl^{10b} did show a reduced viral titre (Zambon *et al.* 2005).

The role of the Toll pathway is further called into question by the results of laboratory experiments examining the responses of flies following infection with the *Drosophila* C virus (DCV)—a natural pathogen of flies. These studies show that the Toll pathway was not upregulate when DCV was introduced by intrathoracic injection (Dostert *et al.* 2005), and was only weakly upregulated when introduce through the more natural route of feeding (Roxstrom-Lindquist *et al.* 2004). Taken together, the results of these studies suggest if the Toll pathway acts against viral infection, it does so in a manner that is specific to both the type of viral infection and the route of infection.

By examining the role of the Toll pathway in combating infection caused by another *Drosophila* virus—the sigma virus—I hope to show how general a role the Toll pathway plays in *Drosophila*'s antiviral response. One reason to suspect that the Toll-pathway is involved in an antiviral response against the sigma virus is that *ref(2)P*—a gene already known to affect resistance to sigma—encodes a protein that sits within the Toll pathway (Avila *et al.* 2002). A number of studies have shown that *ref(2)P* interacts with both *Drosophila* atypical protein kinase C (aPKC), and *Drosophila* tumor necrosis factor receptor-associated factor 2 (dTRAF2) (Avila *et al.* 2002). Both aPKC and dTRAF2 are involved in the Toll pathway; aPKC is an isozyme that activates the NF-κB complex—a protein complex that is a transcription factor—that when active moves across the nuclear membrane to transcribe antimicrobial peptides and dTRAF2 is a signalling protein just upstream of the NF-κB complex (Sanz *et al.* 1999, Shen *et al.* 2001). *ref(2)P* is most likely fulfilling a similar function to its mammalian homolog—p62—interacting with both aPKCs and TRAF-6, anchoring aPKC and the NF-κB complex together to the intercellular membranes (Sanz *et al.* 1999).

Previous studies have confirmed the role of *ref(2)P* within the *Drosophila* Toll pathway, by demonstrating that over-expressing *ref(2)P* in cell lines activates a promoter protein, just upstream of *Drosomycin*—an antimicrobial peptide gene in the Toll pathway. Moreover, depletion of *ref(2)P* in cell lines leads to a reduction in *Drosomycin* transcription (Avila *et al.* 2002).

In this study, I have carried out several experiments to examine whether the Toll pathway is involved in an antiviral response against the sigma virus. In the first experiment I examine whether the presence of the sigma virus increases the susceptibility of flies to other pathogens. I tested this by examining the susceptibility of flies to a fungal infection (fungal infections activate the Toll pathway) in sigma-infected and -uninfected flies. In the second experiment, I looked to see whether the sigma virus suppresses the Toll-pathway, as might be expected if the Toll-pathway is involved in an anti-viral response against the sigma virus. I tested this by measuring the expression of the *IM1* gene—a reporter gene for the Toll-pathway—in sigma infected and uninfected flies, using bacteria and fungal infections to induce a Toll-mediated immune response. In the final experiment, I directly test the role of Toll as an antiviral mechanism against the sigma virus by assessing the susceptibility of flies with and without a functioning Toll pathway to sigma infection.

5.2 Materials and methods

5.2.1 Infection methods

Virus. For viral infection, flies were injected intra-abdominally using a Narishige glass capillary pulled on a Narishige needle puller. The virus suspension was prepared by homogenizing 120 infected flies in 600 μl of Ringers' solution and centrifuging at 13,000 rpm for 1 minute. The supernatant was then collected and kept on ice and used within 2 hours. Three different viral isolates were used in the experiments; one was collected in Greece (PF136), one in Florida (AP30) and one in France (A3) (Carpenter *et al.* 2007).

Fungus Preparation. To guarantee that the *Beauvaria bassiana* isolate was virulent against *Drosophila*, it was initially passaged through an out-crossed population of *D. yakuba*. Flies were sprayed in a mesh cage with a spore/oil formulation (see below),

cadavers were collected over the following 10 days and placed in humid Petri dishes to promote sporulation. Sporulating cadavers were allowed to dry, homogenized in oil and plated onto potato dextrose agar containing chloramphenicol antibiotic (5 x 10^{-5} g ml^{-1}). Plates were incubated for 10 days (25 degrees centigrade, 24 h dark), then dried at room temperature for 5 days. Sporulating fungal material was scraped from all plates and pooled, then dried on silica gel in a fridge and suspended in oil (87.5% Shellsol T, 12.5% Ondina EL). Spore concentration was 2 x 10^8 spores $ml^{-}1$. The formulation was vortexed and agitated briefly using a probe sonicator prior to use. Tungsten needles were dipped into this formulation for experiments where the fungus was introduced into the body cavity of the fly.

For experiments where fungus was delivered ectopically, a controlled amount of fungus was inserted into each vial. This was done by spraying the fungal/oil suspension onto inkjet transparency films with an airbrush, yielding a mean spore density of 5680 spores/mm^2. After 24 hours, when the lighter Shellsol oil had evaporated, the transparency film was cut into strips, rolled and individual pieces inserted into standard *Drosophila* vials.

Bacteria Preparation. *Micrococcus luteus* (supplied by Christopher French) was autoclaved to kill all live bacteria. The bacteria suspension was then aliquoted into two eppendorfs, one eppendorf was centrifuged for 1 minute at 13,000 rpm and the supernatant was removed to leave a pellet, the second was left as a bacterial suspension.

5.2.2 Measuring susceptibility of sigma-infected flies to fungal infection

5.2.2.1 Fly stocks for fungal infection experiment

To assay for the susceptibility of flies to fungal infection I exposed males and females from an attached-X stock; *X^X*; *bw*; *st* (created by Tony Long and henceforth called *X^X*), that were either infected by the sigma virus or uninfected, to fungal spores and recorded the numbers that died over the course of five days. This stock is isogenic for second and third chromosomes and females of this stock are all identical. Due to a shortage of males, two male genotypes were used (TL44 and TL134), however the two males differ only for their first chromosome.

Because the sigma virus is only transmitted vertically, and because the *X^X* stock

was not infected with the sigma virus at the outset, I injected *XˆX* females inter-abdominally to establish viral infections. I used the viral isolate PF136. From the progeny of injected *XˆX*, I selected a line that was infected with the sigma virus and that had a high rate of vertical transmission. Control lines were set up from females in which the sigma virus infection failed to establish.

5.2.2.2 Experiment

Sigma infected and uninfected *XˆX* virgin females were collected from the standard-density bottle cultures. Pairs of females (aged between 2 and 6 days) were placed in vials containing one yeast pellet with pairs of identical males and allowed to lay for two days. After two days in the vial, the parents were removed from these vials and the female parents were checked for sigma infection; if either female in a sigma-infected cross was uninfected the vial was discarded. The offspring of this cross were collected and the two-day old males and females were separated; placing 15 males and 15 females into vials that did not contain Nipogen. When the flies were four-days-old, they were tipped into fresh vials that contained either fungal-infected acetates or fungal-uninfected acetates, and left for two days before being tipped out into fresh vials. The flies were then tipped every two days onto fresh food and the number of dead flies was recorded. Recording ceased when ~90% of females and 60% of male flies had died, on average, in fungal infected vials.

5.2.3 Measuring the suppression of Toll-pathway in sigma-infected flies

5.2.3.1 Fly stocks for Q-PCR experiment

To assess whether the sigma virus suppresses the Toll-pathway, I measured the expression of the *Immune Induced Molecule 1* (*IM1*) gene—a reporter gene for the Toll-pathway (Wasserman 2004)—in sigma-infected and -uninfected flies. I used bacteria and fungal infections to induce a Toll-mediated immune response.

I measured Toll-gene expression levels in an isogenic stock w^{1118} (*Exelixis*) that was either infected or uninfected with the sigma virus. w^{118} is the background on which *Exelixis* corporation generates deficiencies (Parks *et al.* 2004). To establish sigma infection in the w^{1118} stock I injected w^{1118} females inter-abdominally with

the viral isolate AP30. Control lines were set up from single females that were not injected.

Throughout the experiments, flies were reared at a constant density. To produce the constant density bottle cultures, I washed eggs off apple juice agar-plates that had been left for 12 hours in cages with live yeast. I then pipetted 26μl of eggs into half-pint bottles containing standard *Drosophila* media. The flies were reared at 25°C on a 12-hour light/dark cycle. When setting up crosses throughout the experiment, I achieved approximately standard densities of 40 offspring by keeping virgin females for a few days on food that had been lightly sprinkled with live yeast, and then setting up crosses with 2 females in a vial for 2 days, without additional yeast.

5.2.3.2 Experiment

Eggs were collected from agar plates containing grape juice and a small quantity of live yeast paste, which had been placed in cages containing either sigma-infected or sigma-uninfected adult w^{1118} flies for 6 hours. The agar plates were then stored at 25°C for 24 hours, and the resulting larvae were placed into vials containing standard yeast medium at a standard density of 50 larvae/vial. The eclosing flies were tipped into vials containing standard yeast medium that had received 1 drop of 1% Streptomycin sulfate 24 hours prior to this to prevent bacterial growth. The flies were aged in these vials for 4 days before I collected females, using ice to anaesthetize them, and transferred 10 females to fresh Streptomycin-treated vials for a further day. When the flies were 15 days old, they were anaesthetised on ice and stabbed in the thorax with a fine tungsten needle dipped in either: pelleted-bacteria, bacteria-suspension, fungal-suspension or an undipped needle. A small subset of flies were stabbed with a tungsten needle dipped in oil to control for the effect of the oil used to suspend the fungal spores. Females were aged for 22 hours after the stabbing and killed with liquid nitrogen.

5.2.3.3 Reverse transcription (RT-PCR) and quantitative PCR (Q-PCR)

To measure the upregulation of the Toll pathway, primer sets were designed to measure *Immune Induced Molecule 1* (*IM1*) expression, a Toll pathway-specific gene. *IM1* expression was compared to the expression of *Actin*—a house-keeping gene—as a

control.

RNA was extracted with Trizol®. The PCR primers were used to reverse transcribe the genomic RNA using M-MLV reverse transcriptase, before amplification by PCR using Hot Start Taq. The PCR products were quantified in real time using hybridization probes designed by TIB Molbiol GmbH (Berlin, Germany) using a Roche LightCycler (Basel, Switzerland) calibrated with standard dilutions of control cDNA targets cloned for each gene. The probes were multiplexed.

5.2.4 Measuring susceptibility to sigma infection in flies without a functional toll-pathway

5.2.4.1 Fly stocks for toll-pathway mutant experiment

This experiment was carried out to assess the susceptibility of flies with and without a functioning Toll-pathway to sigma infection. To do this, flies were created that are genetically identical across most of the genome but differ in whether they have a functioning Toll-pathway or not. In addition to this, flies were also given a copy of the resistant *ref(2)P* allele to investigate the whether the Toll-pathway underlies the resistance associated with *ref(2)P*.

To create these flies I used the fly stock Dif^1 (Rutschmann *et al.* 2000), a loss of function mutant in an NF-κB transcription factor in the Toll pathway (supplied by L. Wu) and crossed the Dif^1 males to a transgenic line $w(P[w^+, ref(2)P^{restrictive}]); CyO/ref(2)P^{od3}, Bl$, that is homozygous for the resistant *ref(2)P* allele—$ref(2)P^{restrictive}$—on the first chromosome. This transgene is overexpressed and should cure sigma infection even in the presence of a null mutant *ref(2)P* allele on the 2nd chromosome ($ref(2)P^{od3}$).

Curly-winged male offspring from this cross (henceforth called $P[w^+, ref(2)P^{restrictive}]; Dif^1$) are hemizygous for the *ref(2)P* resistant allele on the first chromosome and retain the loss of function mutation on the second chromosome. These males were crossed to either females of the deficiency stock, $w^{1118}; Df(2L)Exel8036/CyO$, or females of the w^{1118} (*Exelixis*) stock. The deficiency stock is a chromosomal deletion of the region from 36B1–36C9, which includes the *Dif* gene, and the w^{1118} stock has the same genetic background as the deficiency stock. The straight-winged female offspring of the cross to the deficiency stock do

not have a functioning Toll-pathway, while the straight-winged female offspring of the cross to the w^{1118} stock have a functioning Toll-pathway. These females were injected with the viral isolate A3, which is sensitive to the resistant *ref(2)P* allele.

5.2.4.2 Experiment

$P[w^{+}, ref(2)P^{restrictive}]; Dif^{1}/cn, bw$ males were collected from standard-density bottle cultures. After 4 days these males were put into cages with either $w^{1118}; Df(2L)Exel8036/CyO$ virgin females (cross A), or w^{1118} virgin females (cross B), that had been collected from standard-density culture bottles and aged for four days. Eggs were collected from plates left in these cages overnight and pipetted (26μl) into half-pint bottles. From cross A, I collected the virgin female offspring that were heterozygous for the Dif^{1} mutation and deficiency, and from cross B, I collected offspring that were heterozygous for the Dif^{1} mutation and wild-type. The females were aged in yeasted-vials for five days—20 females per vial—before I injected them with a controlled amount of virus and returned them 10 to a vial. I discounted females that died from the injury of the injection. I then assayed for the presence of the virus 4 days later, by exposing these females to CO_2 for 15 minutes at 12°C. By 2 hours post exposure, uninfected flies are awake from this anaesthesia, but flies infected with the sigma virus are dead or paralysed.

5.3 Results

5.3.1 Fungal infection experiment

I determined the factors affecting the number of flies dying from fungal infection using a generalised linear mixed-effects model implemented using R's lmer function (R version 2.6.0). The mortality of flies was measured as a ratio of the number of dead flies to the number of surviving flies over five days as a combined dependent variable and the data follow a quasi-binomial (binomial with over-dispersion) distribution. I analysed non-cumulative death—the proportion of flies that die on a single day out of those that had survived to that day. This ensured that the number of flies dying on any one day did not include flies dying on another day, allowing all days to be analysed together.

I chose the following model, as it permits the effect of sigma-virus infection and fungus-infection on mortality to be measured separately, and test whether their combined effects are additive or non-additive:

Let $v_{i,j,k,l,m,n}$ be the ratio of dead to alive flies for treatment i on day j for sex k, male genotype l and for vial m.

$$v_{i,j,k,l,m,n} = \mu + \beta_i + \alpha_j + \delta_k + \kappa_l + \gamma_m + \epsilon_{i,j,k,l} \tag{5.1}$$

where μ is the mean ratio of dead to alive flies, β_i represents the fixed effect of treatment $i = 1, \ldots, 4$: (1) S + F; (2) NS +F; (3) S + NF and (4) NS + NF, where S is sigma-infection, F is fungus-infection, NS is no-sigma-infection and NF is no-fungus-infection, α_j represents the fixed effect of day $j = 1, \ldots, 5$, δ_k is a fixed effect of sex $k = 1, \ldots, 2$, κ_l is a fixed effect of male genotype $l = 1, \ldots, 2$, γ_m is a random variable representing the deviation for vial m and $\epsilon_{i,j,k,l,m,n}$ is a random variable representing the deviation for observation n from treatment i, day j, sex k, male l, vial m. Markov Chain Monte Carlo methods were used to sample from the posterior distribution of the parameters of the model to test the significance of posthoc contrasts.

Mortality data were collected for a total of 3,150 flies from 210 vials (mean flies per vial at the start of the experiment was 15) over 10 days and measurements were taken every two days. By day 10, when observations ended, mean mortality across the fungal exposed vials was 68%±0.33 (n=64), whereas that of control vials was 7.7%±0.1 (n=96). Using the same model as define above, except analysing each day separately, I found that the mortality trajectories were different for the different sexes; for females, fungal treatment significantly elevated mortality from day 2 onwards (MCMCtest: $P < 0.05$) compared to day 4 onwards for males (MCMCtest: $P < 0.01$), indicating that males take longer to succumb to infection (Figure 5.2). I found that there is a main effect of sex, indicating that differences between males and females persist throughout the experiment (MCMCtest: $P < 0.001$).

I found that significantly more flies died in vials that were treated with fungal spores compared to vials that were not treated with fungal spores—females, on average, infected with fungal spores suffered 23% greater mortality than uninfected-females; fungal-infected males, on average, suffered 9% greater mortality than

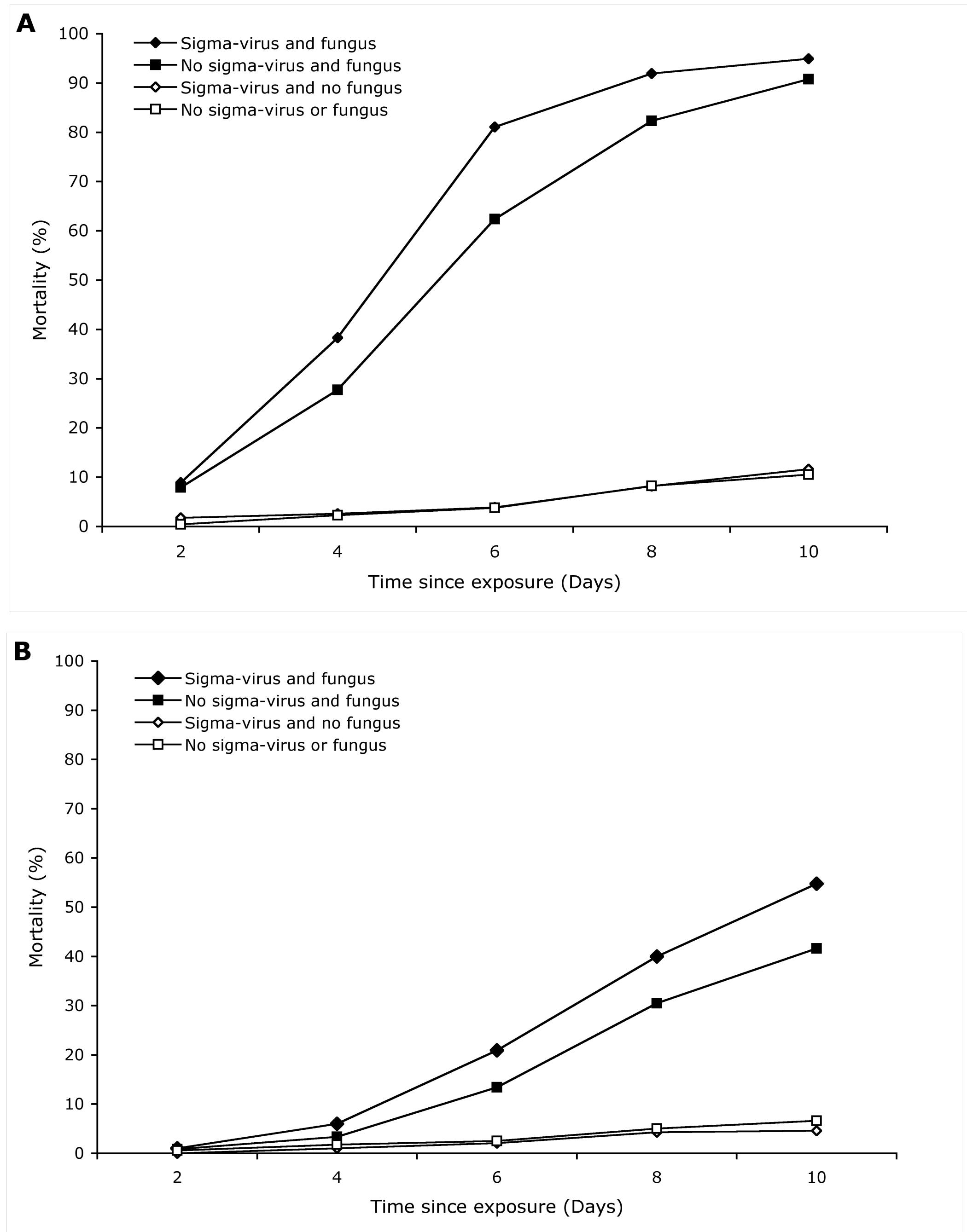


Figure 5.2: Survival trajectories for flies (Graph A: females; Graph B: males) following exposure to fungal spores. Means are unadjusted cumulative mortality.

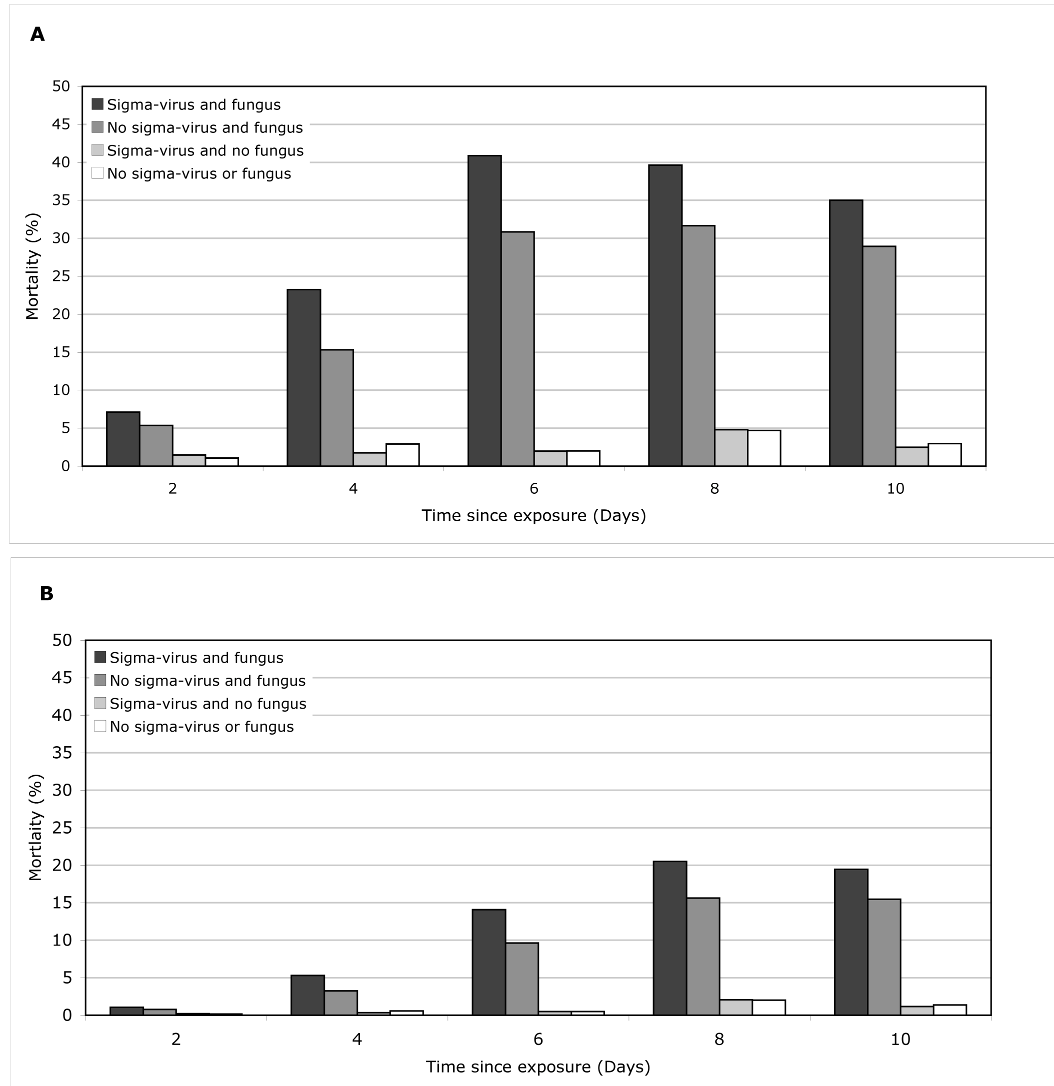


Figure 5.3: Non-cumulative mortality of flies following exposure to fungal infected spores (Graph A: females; Graph B: males). The graph shows back-transformed means from Table 5.1

Table 5.1: Summary of statistics from mixed model with 95% confidence limits calculated from MCMC sampling. Significance (*P*) indicates difference from intercept (S + F + female + day 1) ***' 0.001 **' 0.01 *' 0.05.

	Treatment	Estimate in logits	Std. Error	Lower C.I	Upper C.I	*P*
1	Intercept	-2.66	0.18	-3.133	-2.183	n/a
2	Virus + No Fungus	-3.18	0.23	-3.909	-2.663	**
3	No Virus + Fungus	-0.49	0.14	-0.926	-0.124	*
4	No Virus + No Fungus	-3.13	0.20	-3.756	-2.614	**
5	Day 2	1.29	0.16	0.914	1.721	*
6	Day 3	2.2	0.15	1.790	2.619	*
7	Day 4	2.72	0.16	2.340	3.186	*
8	Day 5	2.72	0.17	2.318	3.178	*
10	Genotype (A)	-0.89	0.16	-1.305	-0.426	**
11	Genotype (B)	-1.78	0.17	-2.240	-1.321	**

uninfected (see Figure 5.3 and Table 5.1 for *P* values).

I was interested in determining whether sigma infection affects the susceptibility of flies to fungal infection. To test this I first looked to see whether sigma infection causes higher mortality in flies in the presence of fungal infection, and second whether sigma infection causes higher morality in flies in the absence of the fungus. I found, in the first test, that significantly more flies died in vials that were infected with both fungal spores and the sigma virus compared to vials only infected with fungal spores (MCMC anaylsis of GLM: $P < 0.001$). By contrast, in the second test, I found no significant difference in the number of flies dying between sigma infected and sigma uninfected flies in the absence of fungal infection (MCMC analysis of GLM: $P = 0.723$). To further test whether the combined effects of sigma infection and fungal infection are additive or non-additive I combined the number of flies dying in vials infected only with fungus, and the number of flies dying in vials infected only with sigma virus (A), and compared this to the number of flies dying in vials infected with both fungal spores and the sigma virus (B). I expect that if the sigma virus increases susceptibility of flies to fungal infection in a non-additive way, then there will be a significant difference between A and B. I found that there is no significant difference between A and B (MCMC analysis of GLM: $P = 0.054$, A: on average, 34% of flies in a vial died compared to B: 41%). However, it is worth noting that this contrast is only marginally not significant, suggesting that there is a strong trend.

5.3.2 Activation of the Toll-pathway experiment

To assess whether the sigma virus suppresses the Toll-pathway, I measured the expression of the *IM1* gene—a reporter gene for the Toll-pathway—in sigma-infected and -uninfected flies that had been stabbed with either bacteria (pellet and suspension) or fungus spore suspension or water. To control for variation in the quality of the RNA extraction between samples I measured the expression of *Actin*, a *Drosophila* gene that is not involved in the Toll-pathway, and express the amount of *IM1* as a proportion of *Actin* per sample. I back-transformed this proportion and analysed it using the following general-linear model implemented using R's (v.2.6.0) lm function:

Let $v_{i,j,k}$ is relative amount of *IM1* for stabbed i, virus-infected j.

$$v_{i,j,k} = \mu + \beta_i + \alpha_j + \epsilon_{i,j,k} \tag{5.2}$$

where μ is the relative amount of *IM1*, β_i represents the fixed effect of stabbing $i = 1, \ldots, 2$, α_j represents the fixed effect of virus infection $j = 1, \ldots, 2$, and $\epsilon_{i,j,k}$ is a random variable representing the deviation for observation k from stabbed i, virus-infected j. The variance components were estimated using maximum likelihood techniques. Each pathogen was analysed separately.

I found that stabbed flies had higher expression of the *IM1* gene than non-stabbed flies for all stabbing treatments (bacteria pellet: $F_{1,18} = 14.359, P < 0.01$; bacteria suspension: $F_{1,25} = 23.898, P < 0.001$; fungus: $F_{1,42} = 6.212, P < 0.05$ and water: $F_{1,24} = 6.279, P < 0.05$). However, sigma virus infected flies had neither higher nor lower expression of the IM1 gene (bacteria pellet: $F_{1,18} = 1.886, P > 0.05$; bacteria suspension: $F_{1,25} = 0.286, P > 0.05$; fungus: $F_{1,42} = 3.131, P > 0.05$ and water: $F_{1,24} = 0.209, P > 0.05$) and I found no interaction between stabbing and viral infection (bacteria pellet: $F_{1,18} = 0.001, P > 0.05$; bacteria suspension: $F_{1,25} = 1.402, P > 0.05$; fungus: $F_{1,42} = 2.905, P > 0.05$ and water: $F_{1,24} = 0.064, P > 0.05$, see Figure 5.4).

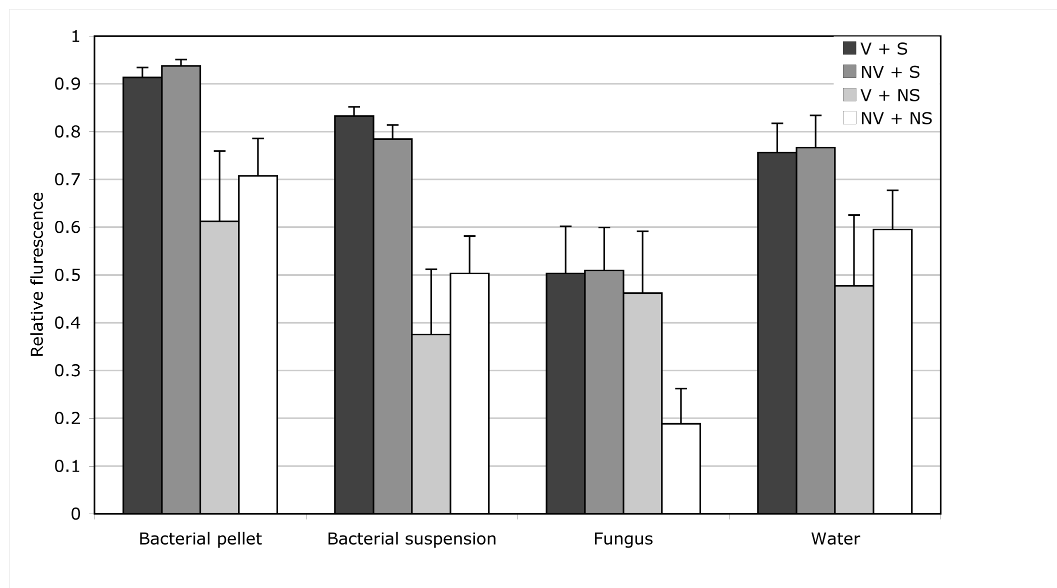


Figure 5.4: Quantification of the *IM1* gene (reporter gene for the Toll pathway) relative to expression of a control gene *Actin*. Sigma-infected (V) and -uninfected flies (NV) were stabbed (S) with tungsten needles that was dipped in either a bacterial pellet, bacterial suspension, fungus or water or left unstabbed (NS). The Y-axis units are on a log-scale and show arbitrary fluorescence units. Error bars show standard errors.

5.3.3 Toll-pathway mutant experiment

I was interested in the role of *ref(2)P* in the Toll pathway. To test this, I created flies that were homozygous for the resistant *ref(2)P* allele—$ref(2)P^{restrictive}$—on the first chromosome. These flies overexpress *ref(2)P* and so should be capable of curing sigma infection even in the presence of another susceptible *ref(2)P* allele. In addition to this, half of these flies were heterozygous for the Dif^1 mutation and deficiency (and therefore did not have a functioning Toll-pathway) and half were heterozygous for the Dif^1 mutation and wild-type (and therefore had a functioning Toll-pathway).

I determined the susceptibility of flies with and without a functioning Toll pathway to sigma infection. The susceptibility of flies to sigma infection was measured as a ratio of the number of dead flies to the number of surviving flies; these data follow a quasi-binomial distribution. I analysed this ratio using the following general-linear model implemented using R's (v.2.6.0) lmer function:

Let $v_{i,j}$ be the ratio of dead to alive flies for genotype i.

$$v_{i,j} = \mu + \beta_i + \epsilon_{i,j} \tag{5.3}$$

where μ is the mean ratio of dead to alive flies, β_i represents the fixed effect of genotype $i = 1, \ldots, 2$, α_j $j = 1, \ldots, 2$: functioning-Toll and no-functioning Toll and $\epsilon_{i,j}$ is a random variable representing the deviation for observation j from genotype i. In total, I injected 1200 flies from 120 vials (mean flies per vial = 10) over 4 days. However, I found no significant effect of the day or the vial (the vial in which the flies were raised) on the fly's susceptibility to sigma infection and so removed these parameters from the model.

I expected that if *ref(2)P* affects a fly's resistance to the sigma virus through a Toll-pathway dependent immune response, then flies that lack a functioning Toll pathway would be more susceptible to viral infection. However, I found that flies that did not have a functioning Toll-pathway were less susceptible to infection by sigma virus when compared to flies with a functioning Toll pathway (with Toll: 57%±0.19, *n*=60; without Toll: 73%±0.18, *n*=59) ($F_{1,117} = 20.982, P = 0.001$) (see Figure 5.5).

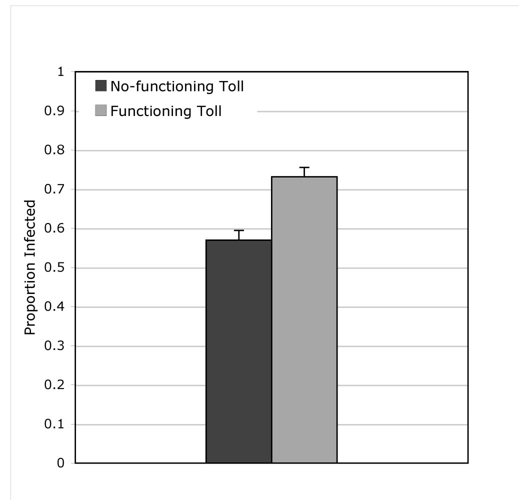


Figure 5.5: Susceptibility of flies with (Dif^1/Df) and without ($Dif^1/+$) a functioning Toll-pathway. The graph shows means from raw data and error bars are standard errors.

5.4 Discussion

In this set of experiments, I was interested in whether the Toll pathway is involved in an antiviral response in *Drosophila*. Previous studies have shown that antimicrobial peptides (AMPs)—known to be induced by the Toll pathway—were upregulated after infection with the *Drosophila* X virus. It was also shown that flies that are deficient for the Toll pathway transcription factor *Dif* are more susceptible to DXV infection (Zambon *et al.* 2005). Furthermore, a gene called *ref(2)P*—which is required by the Toll immune response—has a naturally occurring polymorphism that reduces the rate at which sigma virus replicates within the fly (Avila *et al.* 2002). However, the role of the Toll pathway as an antiviral response remains uncertain because previous studies have shown that Toll pathway genes are not upregulated by DCV infection (Dostert *et al.* 2005), and not all Toll pathway mutants alter the fly's susceptibility to DXV (Zambon *et al.* 2005). More evidence is needed to confirm the role of the Toll pathway in antiviral immunity in flies. By examining the role of the Toll pathway in combating infection caused by another *Drosophila* virus, the sigma virus, I hope to show how general a role the Toll pathway plays in *Drosophila*'s antiviral response.

5.4.1 Sigma-induced susceptibility to fungal infection

In the first experiment, I looked to see whether infection with a fungal pathogen—known to induce a Toll pathway mediated immune response—caused greater mortality in sigma-infected flies compared to sigma-uninfected flies. My results show that the sigma virus increases the susceptibility of flies that are infected with a fungus, but has no affect when flies are clear of fungal infection. Why does the sigma virus increase a fly's susceptibility to secondary infections? One explanation is that the sigma virus is suppressing the Toll pathway. The production of antimicrobial peptides (AMPs) and the cellular response—important for the melanization, digestion and encapsulation of invading microbes—are both under the control of the Toll pathway. Therefore suppression of this pathway would lead to a reduction in the efficiency with which flies clear fungal spores, resulting in a more pronounced infection and greater fungal-induced mortality. Alternatively, the sigma virus may not be suppressing the Toll-pathway, but instead, flies infected with the sigma virus are more susceptible to fungal infection because the increased costs associated with two pathogen infections weaken the immune system, resulting in greater mortality compared to flies infected with only one pathogen.

Greater susceptibility of sigma-infected flies to fungal infection could have profound effects in the wild where flies are constantly exposed to bacteria and fungus during feeding—flies eat rotting plant material. In Chapter 3, I showed that the sigma virus has a prevalence of 0–15% within populations (Carpenter *et al.* 2007). Sigma virus is known to be harmful to flies: wild collected sigma-infected flies suffer a 20% cost compared to uninfected flies (personal communication: Lena Bayer). In laboratory experiments, sigma-infected flies suffer reduced egg viability (Fleuriet 1981), however this reduction in fecundity is unlikely to fully account for the low prevalence of sigma-virus in natural populations of *Drosophila*. The increased susceptibility of sigma-infected flies to other pathogens, might be another factor maintaining the low prevalence of the sigma-virus in *Drosophila* populations.

5.4.2 No activation of the Toll-pathway by sigma

In the second experiment, I was interested in whether sigma suppresses the fly's immune system. Suppression of the immune system allows viruses to replicate freely

without detection. Among the most potent immunosuppressive pathogens of insects are viruses in the family Polydnaviridae (PDV) (Thoetkiattikul *et al.* 2005). PDVs persist as integrated viruses in the genomes of wasps and replicate in the ovaries of females. When females lay an egg into the insect host, they also inject a quantity of virus that infects host immune cells and other tissues, suppressing the host's immune response, allowing the parasitoid to successfully develop (Thoetkiattikul *et al.* 2005). Could the sigma virus have evolved to suppress the fly's immune response? To test this I measured the suppression of the Toll pathway in sigma infected flies quantitatively by measuring the expression of the *IM1* gene—a reporter gene for the Toll pathway. Resistance to the sigma virus may act by preventing the virus from suppressing the immune system, and so to ensure that I could detect immune suppression, I used a fly line known to be susceptible to the viral isolate used in this experiment. Despite successfully up regulating the Toll pathway with both bacterial and fungal challenge, I did not detect suppression of the Toll pathway in sigma virus infected flies.

Why did I fail to detect suppression of the Toll pathway by the sigma virus? One possibility is that the sigma virus does not suppress the Toll pathway and the greater fungal-induced mortality in flies infected with sigma virus, that I observed in the first experiment, is the result of these flies suffer larger costs of infection, and immunity induced by two different pathogens compared to flies infected with only one pathogen. Alternatively, the Toll pathway may bifurcate, with one branch responding to fungal and bacterial infections and the other to viruses. Viruses may therefore not suppress the entire Toll pathway but specifically the antiviral branch. If this is the case, by measuring the expression of a reporter gene for the Toll pathway, such as *IM1*, that is transcribed in response to bacterial and fungal infections, I may not detect suppression of the viral branch of the Toll pathway.

5.4.3 Toll-pathway not involved in an antiviral response against sigma virus

In the third experiment, I was interested in whether *ref(2)P*—a gene already known to affect resistance to sigma—affects a fly's resistance to the sigma virus through a Toll-pathway dependent immune response. One reason to suspect that *ref(2)P* is involved in a Toll-mediated anti-viral response against the sigma virus is that *ref(2)P* encodes a protein that sits within the Toll pathway (Avila *et al.* 2002). A number of studies have shown that *ref(2)P* interacts with both *Drosophila* atypical protein

kinase C (aPKC), and *Drosophila* tumor necrosis factor receptor-associated factor 2 (dTRAF2) (Avila *et al.* 2002). Both aPKC and dTRAF2 are involved in the Toll pathway; aPKC is an isozyme that activates the NF-κB complex—a protein complex that is a transcription factor—that when active moves across the nuclear membrane to transcribe antimicrobial peptides and dTRAF2 is a signalling protein just upstream of NF-κB complex (Sanz *et al.* 1999, Shen *et al.* 2001). *ref(2)P* is most likely fulfilling a similar function to its mammalian homolog—p62—interacting with both aPKCs and TRAF-6 , anchoring aPKC and the NF-κB complex together to the intercellular membranes (Sanz *et al.* 1999).

Previous studies have confirmed the role of *ref(2)P* within the *Drosophila* Toll pathway, by demonstrating that over-expressing *ref(2)P* in cell lines activates a promoter protein, just upstream of *Drosomycin*—an antimicrobial peptide gene in the Toll pathway. Moreover, depletion of *ref(2)P* in cell lines leads to a reduction in Drosomycin transcription (Avila *et al.* 2002). What's more, a previous study has shown that ref(2)P interacts directly with viral proteins, forming complexes with both the N and P viral proteins (Wyers *et al.* 1993). This interaction seems to be specific to sigma virus, as no complex was observed between ref(2)P and vesicular stomatitis virus—a close relative of sigma virus. This study suggests that ref(2)P targets the P protein, which is involved in viral replication, to control viral replication. Interestingly, antibodies used to target ref(2)P also associate with the N protein, suggesting that this nucleocapsid protein has many structural similarities to ref(2)P (Wyers *et al.* 1993).

To test *ref(2)P*'s role in the Toll pathway, I tested the susceptibility of flies that were hemizygous for the resistant *ref(2)P* allele—$ref(2)P^{restrictive}$—and therefore capable of curing sigma infection but that did not have a functioning Toll-pathway. A fly's resistance to the sigma virus is dependent on both which *ref(2)P* allele it has and which viral isolate is infecting it. Therefore in this experiment, it was important to use a resistant allele and a viral isolate that was sensitive to this resistant allele, to ensure that any differences in infection rates could be attributed to the Toll-pathway and not solely to *ref(2)P* in a non-Toll dependent manner.

I expected that if *ref(2)P* affects a fly's resistance to the sigma virus through a Toll-pathway dependent immune response, then flies that lack a functioning Toll pathway would be more susceptible to viral infection. However, I found that flies that

did not have a functioning Toll-pathway (Dif^1/Df) were less susceptible to infection by sigma virus when compared to flies with a functioning Toll pathway (Dif^1/+). This suggests that despite the convincing evidence that *ref(2)P* plays a role in the Toll pathway, the Toll pathway is not involved in an anti-viral response against the sigma virus.

6 ADAR-induced hypermutation in the sigma virus

The work described in this Chapter was carried out by myself, except for the assay of editing rates in *Nicotinic Acetylcholine Receptor α 34E* and *Resistant to dieldrin*, which were carried out by James Brindle and Liam Keegan.

6.1 Introduction

The presence of double stranded (ds)RNA in cells is commonly indicative of a viral infection. Once detected, dsRNA triggers a number of different antiviral immune responses including: the production of interferons—proteins that inhibit viral replication and activate cells to detect and destroy dsRNA (Pestka *et al.* 1987); the activation of the RNAi pathway—that targets and cleaves dsRNA (Bass 2000). dsRNA is also edited by ADARs and recent studies have suggested that this editing may play an antivirial role (Bass 2002).

Adenosine deaminases that act on RNA (ADARs) are RNA-editing enzymes that target regions of dsRNA, converting adenosine (A) to inosine (I) (Bass & Weintraub 1988). Because I base-pairs with cytidine (C) during reverse transcription, these Cs are then base-paired with guanosine (G) during second strand synthesis. Therefore, the conversion of A to I is read by the translation machinery as if it were guanosine and so editing events show up as changes from an A to a G (see Keegan *et al.* (2001) for review). Previous studies have shown that ADARs act efficiently on dsRNA of ~100bp or longer, modifying about 50% of the adenosines (Nishikura *et al.* 1991) and have a strong 5' neighbour preference of $A = U > C > G$, rarely targeting adenosines less than three nucleotides from the 5' terminus, or eight nucleotides from the 3' terminus (Lehmann & Bass 2000, Polson & Bass 1994).

In mammals (Seeburg 1996), *Drosophila* (Palladino *et al.* 2000) and squid (Patton *et al.* 1997) most of the ADAR-edited transcripts are expressed in the central nervous system. Among these edited transcripts are genes involved in ion channels:

the glutamate-gated ion channel receptors (Glu-R) (Seeburg 1996) and serotonin (5-HT2c) receptor in humans (Burns *et al.* 1997), the glutamate-gated chloride-channels (Semenov & Pak 1999), calcium-channels (Palladino *et al.* 2000) and sodium-channels (Smith *et al.* 1998) in *Drosophila*. ADAR-editing in these ion-channel transcripts is thought to generate protein diversity within the nervous system, for example cacophony—a voltage-gated calcium channel—is edited at 10 different sites, generating more than 1000 different isoforms (Smith *et al.* 1998).

Previous studies have shown that *Drosophila* that lack ADAR activity exhibit profound behavioural defects, including slow uncoordinated locomotion, occasional tremors and varying degrees of abnormal body posture. Although flies lacking ADAR are unable to court females, they do not have diminished life span (Palladino *et al.* 2000). In mammals however, failure to edit the transcript for a glutamate receptor (GluR-B) causes mice to die post-natally from neurological dysfunction (Brusa *et al.* 1995).

Aside from their role in the nervous system, ADARs are thought to be involved in hypermutation in some RNA viruses. A to G hypermutations have been found in vesicular stomatitis virus (rhabdoviridae) (Ohara *et al.* 1984), in three retroviruses; human immunodeficiency virus type 1 (HIV-1), rous-associated virus (RAV-1) and avian leucosis virus (ALV) (Felder *et al.* 1994, Hajjar & Linial 1995, Sharmeen *et al.* 1991); two paramyxoviruses; human respiratory syncyial virus (RSV) and measles virus (Cattaneo *et al.* 1988, Rueda *et al.* 1994); the polyoma virus (PV: Polyomaviridae) (Kumar & Carmichael 1997) and the hepatitis delta virus (HDV)(Wong & Lazinski 2002). For most viruses, the A to G mutations occur throughout the viral genomes, converting ~50% of As to Gs, with the exception of HIV-1 and HDV, where mutations occurred at specific sites and are thought to be required for viral packaging.

It is thought that these mutations introduced into viral genomes by ADARs could impact the efficiency of the viral multiplication process by altering either the stability of dsRNA that form during viral replication or by altering viral mRNAs so they no longer encode functional viral proteins (Patterson & Samuel 1995). It has also been proposed that hyper-edited dsRNA by ADARs may form part of an antiviral mechanism whereby editing may tag the dsRNA for subsequent disposal (Scadden & Smith 1997, Scadden 2005). In support of this idea, a previous study showed that an RNAase is responsible for degrading ADAR-edited RNA (Scadden & Smith 1997,

2001*b*). One such RNAase was subsequently identified as Tudor-SN—a component of the RNAi silencing complex (RISC). It was found that Tudor-SN promotes the degradation of RNA that has been edited by ADARs (Scadden 2005).

Are ADAR-edited RNAs degraded by the RNAi pathway? RNAi is conserved across metazoans and processes dsRNA by a class of RNAse III ribonucleases (Dicers) to produce small interfering RNAs (siRNAs). The siRNAs are incorporated into the RNA-induced silencing complex (RISC) that targets and degrades cognate mRNAs (Ding & Voinnet 2007). There are several lines of evidence that suggest that hyper-editing by ADARs might prevent dsRNA from being recognised and efficiently cleaved by the RNAi pathway (Knight & Bass 2002, Scadden & Smith 2001*a*, Tonkin & Bass 2003). However, a very recent study has found that siRNAs show evidence of ADAR-editing (Kawamura *et al.* 2008), suggesting that the role of ADARs in the RNAi pathway is unresolved.

In this Chapter, I describe evidence for ADAR caused hyper-editing in the sigma virus—a naturally occurring pathogen of *D. melanogaster*. I found evidence for A to G hypermutations in a single viral line and so examined multiple other viral isolates from around the world for evidence of hypermutation caused by ADARs. I also investigated the role of ADAR as an antiviral mechanism by looking to see whether the sigma virus has evolved to suppress ADAR-editing by comparing editing rates of a number of genes in virally-infected and -uninfected flies.

6.2 Materials and methods

6.2.1 Detecting hypermutation in the sigma virus

I sequenced 5744bp from two sigma viral lines (A3 and A3E55), supplied by Didier Contamine that had been maintained in the laboratory. These viral lines are a single wild collected isolate which was split between two separate fly lines and maintained at 20°C for between 10 and 20 years. I then sequenced two regions from a large sample of viral isolates: the first encompasses the polymerase-associated protein; the second includes the G gene, encoding the outer-coat protein (590 bp and 637 bp, respectively) to look for hypermutation consistent with ADAR-editing. The sequences come from virus isolated in *D. melanogaster* lines collected in Europe (Greece, UK and Spain)

and North America (Georgia and Florida, USA). These sequences are mostly protein coding but include small intergenic regions. For details of sequencing methods and sequence analysis see Chapter 3 and (Carpenter *et al.* 2007).

Because I found evidence of ADAR-editing in the *PP3* gene in one of the lines supplied by Contamine, I went on to clone using a TOPO TA cloning® kit (Invitrogen, Paisley, UK) and sequence this same section of the *PP3* gene from ~10 viral genomes from 9 different viral isolates. I assume that only a proportion of viral genomes in a fly will be edited and so I hoped to increase our chance to detect ADAR-editing by cloning individual viral genomes rather than direct sequencing the viral RNA from a fly.

6.2.2 Detecting suppression of ADAR-editing by the sigma virus

This experiment was carried out to compare the rates of ADAR editing in viral infected and uninfected flies. I measured the rates of editing in three genes: *Adar* itself, that is edited at one site (Keegan *et al.* 2005); *Nicotinic Acetylcholine Receptor α 34E* (*nAChR*) that is edited at 10 sites and *Resistant to dieldrin* (*Rdl*) that is edited at 6 sites (Hoopengardner *et al.* 2003). I compared editing rates of *Adar* in sigma virus and *Drosophila* X virus (DXV) infected flies, and compared these rates to control lines. For *nAChR* and *Rdl*, only sigma virus-infected lines were compared to controls.

To measure editing rates of *Adar*, *nAChR* and *Rdl* in sigma-infected flies I established the sigma infection by injecting female *w118* (*Exelixis*) flies inter-abdominally with the viral isolate AP30 (collected in Florida (Carpenter *et al.* 2007)) prior to the experiment. Control lines were set up from single females that were not injected. For the experiment, pairs of virgin females (aged between 2 and 6 days) were placed in vials containing yeast with pairs of identical males and allowed to lay for two days. After two days in the vial, the parents were removed from these vials and the female parents were checked for sigma infection; if either female in a sigma-infected cross was uninfected the vial was discarded. When the offspring of this cross were ten days old, they were anaesthetised on ice and homogenized.

To measure editing rates of *Adar* in *Drosophila* X virus (DXV) infected flies compared to control lines, pairs of uninfected females (that had been raised at standard density) were crossed to pairs of males and allowed to lay for two days. After two

days in the vial, the parents were removed from these vials. When the offspring were two-days-old they were injected with either DXV or Ringer's solution intra-abdominally. The injected offspring were then aged for a further four days to allow the viral infection to establish and were then anaesthetised on ice and homogenized.

RNA was extracted from 10 *Drosophila* using Trizol® and cDNA was synthesised with MMLV Reverse Transcriptase (Promega, Madison, WI) and either specific primers (designed to amplify *Adar* (Keegan *et al.* 2005)) or random hexamers (for *nAChR* and *Rdl*) were used to generate cDNA. cDNA for *Adar* was amplified by standard PCR and run through the pyrosequencer to establish editing rates, while editing rates for *nAChR* and *Rdl* were measured by amplifying cDNA by standard PCR, before gel purifying bands using a Qiagen gel extraction kit (Sussex, UK) and sequencing using ABI BigDye (Applied Biosystems, Foster City, CA 94404).

6.2.3 Statistical analysis

To compare editing rates in sigma-infected and -uninfected flies, I analysed the rate of editing by calculating the proportion of As to Gs at the edited site and determined whether infected flies differed in their editing rates compared to controls using a general-linear model implemented using R's (v.2.6.0) lm function, and evaluated this model using maximum likelihood. Each gene was analysed separately. For editing of the *Adar* transcripts, rates were analysed separately in flies infected with DXV and sigma virus using the following linear model:

Let $\nu_{i,j}$ be the editing rate for viral infection i.

$$\nu_{i,j} = \mu + \beta_i + \epsilon_{i,j} \tag{6.1}$$

where μ is the editing rate, β_i represents the fixed effect of viral infection $i = 1, \ldots, 2$ and $\epsilon_{i,j}$ is a random variable representing the deviation for observation j from viral infection i. Editing rates of the *nAChR* and *Rdl* transcripts (that are edited at multiple sites) were analysed separately by fitting the following model to the data:

Let $v_{i,j}$ be the editing rate for viral infection i and site j.

$$v_{i,j} = \mu + \beta_i + \rho_j + \epsilon_{i,j,k} \quad (6.2)$$

where μ is the editing rate, β_i represents the fixed effect of viral infection $i = 1, \ldots, 2$, ρ is a fixed effect of site edited, for *nAChR* it had 10 levels and for *Rdl* it had 6 levels and $\epsilon_{i,j,k}$ is a random variable representing the deviation for observation k from viral infection i and site j.

6.3 Results

6.3.1 Evidence for hypermutation

I found evidence for hypermutation by sequencing 5744bp of the viral genome from two viral lines (A3 and A3 E55) split from each other and maintained at 20°C for 10 to 20 years. I found 25 substitutions between these two viral lines. Using an out-group I assigned these mutations to either the lineage leading to A3 or A3 E55. I found that A3 E55 accumulated significantly more substitutions than A3 ($\chi_2 = 11.58$, *d.f.* = 1, $P < 0.001$), with 21 of the 25 substitutions unique to A3 E55. All of the substitutions occurring in A3 E55 were changes from A to G, if we consider the negatively stranded genome, and all but one were highly clustered within a 565bp region of the third gene encoding the PP3 protein (Figure 6.1).

Adenosine deaminases (ADARs)—RNA editing enzymes that target double-stranded regions of both nuclear-encoded RNA and viral RNA are the only cellular activity described to date that could result in A to G hypermutations in dsRNA. Since ADARs act on specific sequences, tending to prefer adenosines that do not have a 5' guanosine (Lehmann & Bass 2000, Polson & Bass 1994), I examined each of the A to G substitutions in the viral line A3 E55, for a bias in base composition at the 5' position. First I reconstructed the ancestral sequence of A3 and A3 E55 using sequences from natural populations as an out-group. Within the 571bp region, none of the 32 As that were preceded by a G had been mutated. In contrast, 20 of the 109 As at preferred sites had been mutated to a G. This paucity of guanosines preceding

```
Ancestral  CAGTGTCATT TGCAGCATAT CTGTAGTTGA ATCCCTTTTG TGCCTGCTTC GTACCCCATT CTCCTCAAGG GCTCTTAAAA ACTCCCGACC AAACTTCCTT CCGTTGGTCA GATCATACCC
A3         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A3  E55    .......... ...G..G... ....G..... G......... .......... ..G....... .......... ......GG.. G......... .......... .........G ......G...

Ancestral  TGGTGTAATG TCTTCCCTAA GATCTTTAAC TGCTTTGGTG AAGCGAGTCA ATGGGTTAAT ATCCATCACC GAGGTCCTGA TGTTTAGCGT CACTCCATCA GGGCTGAGGG TTGCAGGTGT
A3         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A3  E55    .......... .......... .......G.. .......... .......... .......... .......... .......... .....G.... .........G .......... ..........

Ancestral  GGTTTCAGGA AATATTGGGG CCAACGTGGA GGTGGAAGAG GGAGAGTGCA CTCTTTGGTC TTGGCTACCC GCGGCATCAT AAGCTCCGGC TAGTTTCGAA AGTTGGAAGA AATCAGTGGC
A3         .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .......... .......... ..........
A3  E55    .......... .G........ ...G...... ......G... .......... .......... .......... .......... .......... .......... .......... ..........

Ancestral  TAGATCTGTT GTTGTCATGT GACTACAGAC AAAATTTTCC TCCCGTGTGA GCCGGCCCTC TTGGAAGTGC TGCACATTTA ACTGCTGAGT AATATTCAGC TCTTCTGCAC CCTCCATCAA
A3         .......... .......... .....T.... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A3  E55    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

Ancestral  GATTGTAGTG TTCGATTGTA CTGGCAATAT CCTGTCTGTT ATATTTATTG ATAACGGGAT AGACACAGAT GCAACATGTC TATAATAATA TAGATGGAAA AAGAAAAGGA CTAAGGGTAG
A3         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A3  E55    .......... .......... .......... .......... .......... .......... G.....G... .......... ......G... ........G. .......... ..........
```

Figure 6.1: Nucleotide sequence of the PP3 protein of the sigma virus. The ancestral line was used as an out-group to assigned mutations to either the lineage leading to A3 or A3 E55. A3 E55 shows evidence of clusters of hypermutated A to G changes. A3 E55 and A3 were laboratory maintained lines that were split from each other 15 years ago. All lines were derived from separate RT-PCR.

changed sites is significant when compared to unchanged sites (χ_2 = 6.84, *d.f.* = 1, *P* = 0.008).

I then examined 23 wild viral isolates for evidence of similar hypermutation events. I did this by reconstructing the ancestral sequence for all but one of the wild viral isolates using the most divergent American isolate as an out-group. ADARs could mutate both the viral genome or a replication intermediate and so hypermutations might show up as an excess of either A to G or T to C changes. I counted the number of polarised changes from A to G and T to C versus changes from G to A and C to T along a 1225bp region. I did not find an excess of A to G or T to C changes relative to other transitions (χ_2 = 0.076, *d.f.* = 1, *P* = 0.78), nor did I find any clustering of A to G mutations that would be indicative of the involvement of ADARs.

6.3.2 No evidence for ADAR-editing in cloned virus

Because I found evidence of ADAR-editing in the *PP3* gene of the A3 E55 line, I went on to clone and sequence this same section of the *PP3* gene from ~10 viral genomes from 9 different viral isolates collected from North America and Europe. If I assume that only a proportion of viral genomes in a fly will be edited and so I hoped to increase the chance to detect ADAR-editing by cloning individual viral genomes rather than direct sequencing the viral RNA from a fly.

Across 86 sequences (~600bp in length) I found 14 polymorphisms—all were singletons. I counted the number of changes from A to G and T to C compared to all other changes across all clones within each of the viral isolate. I found that 6 of the 14 changes were from A to G and T to C—the direction of ADAR-editing. I then looked to see if these changes were at sites preferred by ADARs by examining the nature of the neighbouring base 5' of the A to G change and 3' of the T to C change, and compared it to the composition of neighbouring bases 5' and 3'of unchanged As and Ts. If the changes are the result of ADAR-editing, I would expect a paucity of Gs neighbouring A to G changes and a paucity of Ts neighbouring T to C changes. However, I did not find a paucity of guanosines preceding A to G or T to C changes compared to unchanged sites (Fisher's exact test (2-tailed) for bases 5' of As: *P* = 1 (1 of 6 changes was at a preferred site) and 3' of Ts: *P* = 0.621 (3 of 6 changes were at preferred sites).

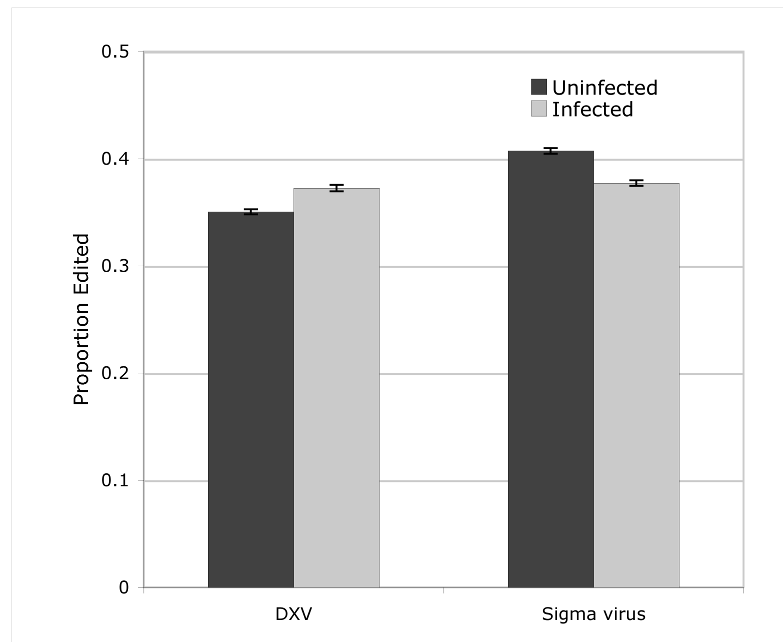


Figure 6.2: Proportion of edited *Adar* transcripts in flies infected with *Drosophila* X virus and the sigma virus compared to uninfected flies. Graph shows means calculated from raw data with standard errors.

6.3.3 No evidence for suppression of ADAR-editing by viruses

I measured the rate at which *Adar* (which is itself edited) was edited in flies that were infected with either the sigma virus or DXV and compared these rates to control flies. If ADARs play an antiviral role in *Drosophila*, then *Drosophila* viruses are likely to evolve to suppress ADAR editing, therefore I measured the rate at which *Adar* was edited in flies that were infected with either the sigma virus or DXV and compared these rates to control flies.

I found that editing rates were significantly lower in sigma-infected flies compared to uninfected flies ($F_{1,52} = 15.886$, $P < 0.001$), however, this only amounted to a 3% reduction. In contrast to this, I found that editing rates were significantly higher in DXV-infected flies compared to control lines ($F_{1,86} = 7.417$, $P < 0.01$), but again this amounted to only a 2% difference (Figure 6.2).

Because the results of the previous experiment suggest that the sigma virus suppresses the rate at which ADAR edits it own transcripts, I was interested in whether the sigma virus suppresses the rate at which ADARs edit other genes in

the fly. It is known that ADAR edits the *Nicotinic Acetylcholine Receptor α 34E* (*nAChR*) transcript at 10 different sites and the *Resistant to dieldrin* (*Rdl*) transcript at 6 different sites (Hoopengardner *et al.* 2003). Both these genes are involved in neurotransmission. However, when I examined editing rates of these two genes, I found no difference between sigma-infected flies and -uninfected at any of the edited sites for either gene (*nAChR*: $F_{1,116}= 0.386$, $P = 0.535$; *Rdl*: $F_{1,55} = 0.001$, $P = 0.969$) (Figure 6.3).

6.4 Discussion

6.4.1 Evidence for ADAR-editing

I have provided the first evidence from outside mammals that viruses can be hypermutated by host ADARs. The clustering of A to G changes observed in the laboratory viral line are typical of mutations that occur as a result of RNA editing by ADARs that convert adenosines to inosines—read by translational machinery as guanosine (Keegan *et al.* 2001). ADARs prefer to edit at sites that don't have a 5' G (Lehmann & Bass 2000, Polson & Bass 1994) and the A to G changes that we see in the sigma virus are all at preferred sites, offering convincing evidence for involvement of RNA editing.

Is ADAR-editing the sigma virus? Further anecdotal evidence comes from the the observation that *Drosophila* that lack ADAR activity exhibit profound behavioural defects, including slow uncoordinated locomotion, occasional tremors and become paralysed when exposed to CO_2 (Palladino *et al.* 2000). These behavioural phenotypes suggest a defect in the nervous control of the fly. Interestingly, flies infected with the sigma virus also exhibit paralysis when exposed to CO_2. It is also worth noting that ADARs are very active in the nervous system—where they are thought to generate protein diversity—editing in both the nucleus and cytoplasm, bringing ADARs in direct contact with the sigma virus, which exists primarily in the cytoplasm of the nervous system. However, it is not known whether *Drosophila* ADARs are active in the cytoplasm. My results suggest that they probably are, or that the sigma virus enters the nucleus.

Evidence of RNA-editing by mammalian ADARs have been found in a number

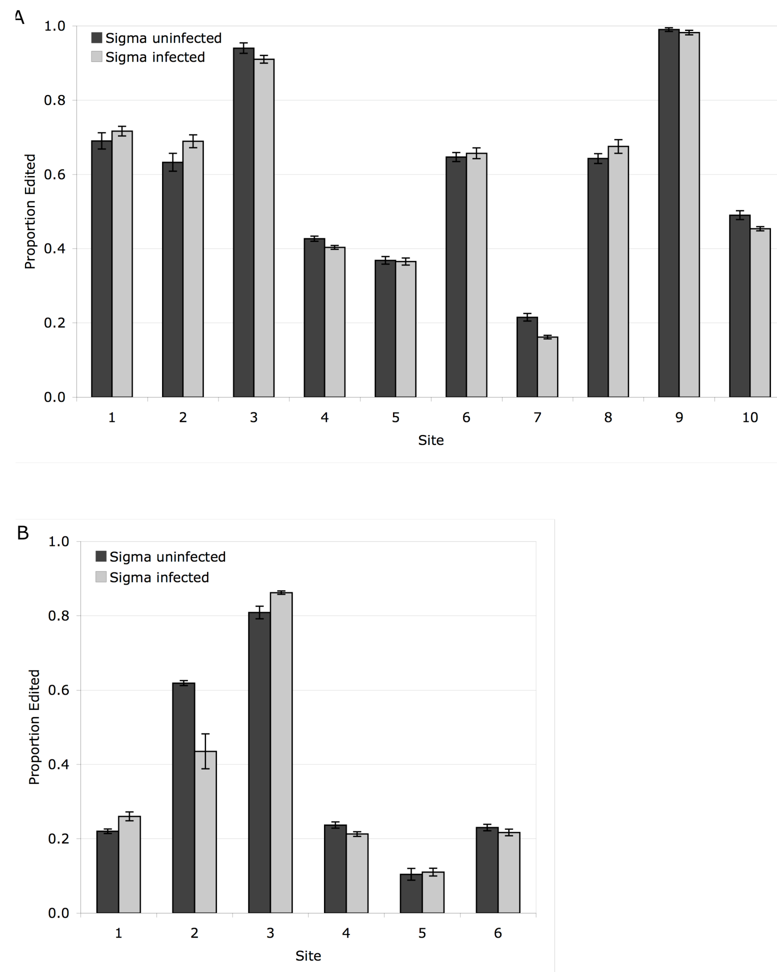


Figure 6.3: Proportion of edited transcripts in two genes (Graph A: *Nicotinic Acetylcholine Receptor* α *34E* (*nAChR*); Graph B: *Resistant to dieldrin* (*Rdl*) in *Drosophila* either infected with the sigma virus or uninfected. *nAChR* is edited at 10 sites, *Rdl* is edited at 6. Graph shows means calculated from raw data with standard errors.

of different viruses. Because many viruses replicate through an intermediate RNA-strand, hypermutations can show up as an excess of either A to G or U to C changes, depending on the strand that is edited. A to G and U to C hypermutations have been found in vesicular stomatitis virus (rhabdoviridae)—one of the closest relatives of the sigma virus—where ~50% of As were changed to Gs in two small regions of the VSV genome (Ohara *et al.* 1984). Similar A to G mutations have been found in three retroviruses; in human immunodeficiency virus type 1 (HIV-1), rous-associated virus (RAV-1) and avian leucosis virus (ALV) (Felder *et al.* 1994, Hajjar & Linial 1995, Sharmeen *et al.* 1991); two paramyxoviruses: human respiratory syncyial virus (RSV) and measles virus (Cattaneo *et al.* 1988, Rueda *et al.* 1994); the polyoma virus (PV: Polyomaviridae) (Kumar & Carmichael 1997) and the hepatitis delta virus (HDV) (Wong & Lazinski 2002).

The function of ADARs in the context of viruses is still unknown. One possibility is that viruses get caught up with ADARs inadvertently and the editing of viral genomes has no biological role. Alternatively, editing events caused by ADARs might be an important source of variation for the viruses. In the measles viruses, viral genomes that accumulate large numbers of mutations due to editing, persist as viral infections in the brain cells of humans (Cattaneo *et al.* 1988). While in Hepatitis delta virus, RNA editing is an essential process in the life-cycle of HDV; ADARs edit a single site that switches production of a protein involved in virus replication to a protein involved in virion assembly (Sato *et al.* 2001).

One further explanation is that editing could be an antiviral defence: introducing highly clustered point mutations into coding regions of viruses to cause deleterious affects. I did not find evidence for hypermutation in wild populations of the sigma virus, however, if these hypermutations are deleterious then it is unlikely that individuals harbouring point mutations introduced by ADARs would persist within the viral populations. If this is the case, RNA editing as an antiviral defence may be common but difficult to observe.

More recently, studies have suggested that hypermutations, instead of introducing deleterious mutations into viral genomes, instead, 'tag' dsRNA for digradation (Scadden & Smith 1997, Scadden 2005). In support of this idea, a previous study has shown that a RNAase—I-RNase—preferentially degrades RNA that has been edited by ADARs over unedited RNA (Scadden & Smith 1997). One such RNAase was

subsequently identified as Tudor staphylococcal nuclease (Tudor-SN)—a component of the RNAi silencing complex (RISC). It was found that Tudor-SN promotes the degradation of RNA that has been edited by ADARs (Scadden 2005). Further evidence for the interaction of ADARs and the RNAi pathway comes from very recent evidence that a large number of the small-interfering (si)RNAs—produced by the RNAi pathway to cleave target RNAs—have evidence of ADAR-editing (Kawamura *et al.* 2008). One possibility is that ADAR-editing determines which branch of the RNAi pathway that dsRNA is processed.

6.4.2 No evidence for suppression of ADAR-editing

If ADARs play an antiviral role in *Drosophila*, then *Drosophila* viruses are likely to evolve to suppress ADAR editing. However, I failed to detect substantial suppression of ADAR editing in sigma or DXV infected flies.

7 Discussion and conclusions

7.1 Summary

The latest mechanistic studies on insect immune systems are radically altering how we think coevolution occurs. New research at the molecular level shows that parasites suppress and evade the hosts immune response and their insect hosts have evolved novel counterstrategies to deal with this. Most of our understanding of the mechanisms of immunity come from studies of a few model organisms, most notably *Drosophila melanogaster* (see Lemaitre & Hoffmann (2007) for review). Many of these studies have focused on the role of the innate immune system to counter bacterial and fungal infections, and, with the exception of RNAi, have largely ignored viruses and anti-viral mechanisms. What's more, few studies have examined the genetics of resistance against bacteria, fungi and particularly viruses in natural populations and still fewer have looked at the genetics of the pathogens that naturally infect *Drosophila* in the wild. In this thesis, I have examined the genetics of *Drosophila*—sigma virus system, using both experimental and phylogenetic approaches with the aim to better understand the coevolutionary dynamics of this system, and how these dynamics impact both the evolution of the sigma virus and it's host—*D. melanogaster*.

Little is known about the biology of the sigma virus in natural populations. For this reason I identified 23 new viral isolates, from 5 different populations of *Drosophila*, from two continents. In Chapter 3, I describe the isolation and characterisation of these new viral isolates and their phylogenetic realtionships to each other. The results indicate that sigma virus has a prevalence of 0–15% in the wild and compared to other RNA viruses, the sigma virus has very low levels of viral genetic diversity across Europe and North America. Based on laboratory measurements of the viral substitution rate, most European and North American viral isolates shared a common ancestor approximately 200 years ago. I suggest two possible explanations for this: the first is that *D. melanogaster* has recently acquired the sigma virus; the second is that a single viral type has recently swept through *D. melanogaster*

populations. Whatever the explanation, the emerging picture suggests that host-parasite associations are very dynamic, with parasites constantly gained and lost from host populations and perhaps jumping between species. A further finding of this study shows that, in contrast to *Drosophila* populations—that are generally panmitic, except for the structure that exists between African and non-African populations—sigma viral populations are highly structured. This is surprising for a vertically transmitted pathogen that has a similar migration rate to its host. The low structure in the viral populations most likely results from the smaller effective population size of the virus compared to the fruit fly. I suggest that vertically transmitted pathogens with their small populations and high mutation rates have the potential to be useful tools in revealing structure and migration in their host populations.

As part of the phylogenetic analysis of these new viral isolates I estimated a substitution rate based on two viral sequences that had been split from each other ~15 years before. One of these viral isolates contained highly clustered point mutations introduced into the coding regions of the sigma virus. These mutations are most likely caused by ADARs—RNA editing enzymes. In Chapter 6, I describe this hyper-mutation and looked for further evidence for ADAR-induced hypermutation in wild viral isolates collected from Europe and North America, although, I failed to find any further evidence of hypermutation. However, it is worth considering that these highly clustered point mutations introduced into coding regions of viruses are most likely deleterious and so individuals harbouring point mutations introduced by ADARs would not persist within the viral populations for long. If this is the case, RNA editing as an antiviral defence may be common but difficult to observe. To further investigate the role of ADARs as an antiviral mechanism, I looked to see whether the sigma virus has evolved to suppress ADAR-editing by comparing editing rates of a number of genes in virally-infected and -uninfected flies, but again, failed to find any.

An insects ability to first evade infection, to recognise and suppress infection, and finally eliminate the infection, is dependent on the genetics of both the pathogen and the host. However, little is known about how specific these interactions are; how these specific interactions are maintained and whether constraints exist that limit pathogens from evolving to be infective to many host genotypes and hosts from evolving to be resist many pathogen genotypes. Understanding these questions, will help us to understand why variation in resistance is so common in natural popualtions. In Chapter 4, I took several of these new viral isolates and tested whether different host

genotypes differ in their susceptibility to viral isolates; whether viral isolates differ in their ability to infect different host genotypes and ultimately whether there are host-viral genotype interactions? To test this, I measured the transmission rates for five different viral isolates collected from around the world in fly lines that differed for their first, second and third chromosome. Although I found considerable genetic variation on chromosomes one, two and three of *D. melanogaster* in transmission rates of the sigma virus, only a small amount of this genetic varaition was involved in specific interactions with particular viruses. This suggests that there are few constraints on flies evolving resistance against all five viruses, although the rate at which selection can increase resistance was slowed by specific interactions with particular viruses.

Finally, in Chapter 5, I showed that sigma-infected flies are more susceptible to fungal infection. This could have profound effects in the wild where flies are constantly exposed to bacteria and fungus during feeding. And the greater susceptibility of sigma-infected flies to secondary infections, could be a factor in maintaining the sigma virus at a low prevalence (as discussed in Chapter 3). One interpretation of this result is that the sigma virus is suppressing the Toll-pathway, which is involved in immune defences against fungal infections. This result, along with existing evidence that the Toll pathway is activated by other *Drosophila* viruses (Zambon *et al.* 2005), led me to investigate whether flies mount a Toll-dependent immune response against the sigma virus. However, in a further experiment, flies without a functioning Toll-pathway were no more susceptible to sigma infection than flies with a functioning Toll-pathway, nor did the sigma virus suppress the Toll-pathway. This result corroborates the result of a recent study (Jennifer Carpenter *et.al.*, unpublished data from a microarray study) that looked at changes in gene expression in *Drosophila* that occur in response to infection with the sigma virus. Although this study detected many changes in gene expression in infected flies, there was no evidence for the activation of the Toll, IMD or Jak-STAT pathways, which control immune responses against other pathogens. This suggests that either *Drosophila* do not mount an immune response against the sigma virus, or that the immune response is controlled by other pathways.

7.2 Conclusions

What can we say about the models of host-parasite coevolution that best describe the interaction between sigma virus and *Drosophila melanogaster*? Models of host-parasite coevolution fall into two classes. The first class, frequency-dependent models, state that selection favours pathogens adapted to the most common host genotypes, and that this in turn confers an advantage to rare host genotypes. Whereas, the second class, arms-race models, propose that new host resistance or parasite virulence mutations arise and sweep to fixation under directional selection. Under frequency-dependent models, resistance polymorphisms are maintained as long-term balanced polymorphism, whereas, under arm-race models, resistance polymorphisms are transient, existing only during the sweep.

In Chapter 3, I investigated patterns of genetic variation in the sigma virus collected worldwide. I found very low levels of genetic diversity across European and North American viral isolates. One explanation for this is that a single viral type has recently swept through *D. melanogaster* populations, resulting in low genetic variation among the viral isolates. Why might a single viral type sweep through *Drosophila* populations? Arms-race models predict a polymorphism affecting infectibility is likely to sweep through pathogen populations in response to the spread of a resistant polymorphism in the host population.

Is there any evidence for a sweeping resistant mutation in *Drosophila* population? Bangham *et al.* (2007) examined a large sample of second chromosomes (where *ref(2)P* occurs) and found significantly less variation among the resistant haplotypes compared to the susceptible haplotypes than expected by chance. This result suggests that the resistant mutation has slowly increased in frequency as part of a selective sweep. Bangham *et al.* (2007) also showed that the resistance mutation is several thousand years old. It is clear that the resistance mutation although not ancient, long predates the spread of the infective virus described in Chapter 3.

Is it possible to reconcile the findings of this study with the data presented in Chapter 3? First, it is worth remembering that the populations of flies described in Chapter 3 came mostly from Europe, whereas Bangham *et al.* (2007) examined a population of flies from the USA, and so the populations on the two continents might have been affected by different selective pressures. Second, it is possible that an arms

race is going on between ref(2)P and the sigma virus, but that the ref(2)P resistance gene has only recently become frequent enough to select for viral countermeasures. Bangham *et al.* (2007) have evidence to show that the resistant mutation is recessive, which means it could take thousands of years to reach the current frequency. In support of this, the frequency of the resistant mutation in samples collected across three continents has never exceeded 23%. This means that homozygous flies—that are resistant to sigma infection—are rare (~5%).

Whether these studies have documented two successive sweeps—the first in the virus and the second in *Drosophila*—or not, is unclear, and although I cannot reject the hypothesis that negative frequency-dependent selection is acting here, the simplest explanation is that both viral and fly populations have seen host resistance or parasite virulence mutations swept to fixation under directional selection.

Further evidence that an arms-race model, rather than a frequency-dependency model, best describes the dynamics of this system comes from Chapter 4. One of the predictions of frequency-dependent models of coevolution is that trade-offs exist, either between resistance against different pathogen genotypes or with other components of fitness. Despite finding considerable genetic variation in transmission rates in Chapter 4, only a small amount of this genetic varaition was involved in specific interactions with particular viruses. This suggests that there are few constraints on flies evolving resistance against all five viruses. And ultimately, trade-offs between resistance against the different viral isolates are unlikely to be maintaining variation in resistance in *D. melanogaster* populations and therefore frequency-dependent models are unlikely to apply to this system.

Additionally, the approach taken in Chapter 4 demonstrates that specificity (as observed by Carius *et al.* (2001)) should not be assumed to lead to constraints that ultimately maintain variation in resistance. In this respect, this work is an extension beyond Carius and demonstrates that measuring the strength of the trade-offs is important, because, if the trade-offs between resistance against different pathogen genotypes are small enough - even if they result in specific interactions between host and parasite - they are unlikely to prevent the evolution of general resistance, and so are unlikely to maintain variation that we see in natural populations.

7.3 Future directions

This thesis paves the way for many potential avenues for future research. The conclusions of Chapter 3, in which I examined the phylogenetic relationship between wild collected sigma-viral isolates, suggest that sigma virus may have jumped into *D. melonogaster* relatively recently. An obvious next question is where did the sigma virus come from and how long ago did it cross a species boundary? Work has already begun to isolate the sigma virus from other *Drosophila* species. If successful, it will be interesting to determine how different these sigma viruses are from sigma viruses isolated in *D. melanogaster*. Such a study would give a good indication for how quickly and in what ways viruses adapt to their new host. Moreover, this study would allow a number of broad evolutionary questions to be asked, such as, how often do pathogens jump between species, and what is the impact of such a jump on the fly's genome?

In Chapter 4, I examined the genotype-by-genotype interactions between *D. melanogaster* and the sigma virus. In this chapter, I showed that a polymorphism in *ref(2)P*—a gene already well known to affect resistance to sigma—affected the rate of transmission of only one of the viral isolates. It would be interesting to investigate the molecular mechanisms underlying the specific interaction between a polymorphism in the *ref(2)P* gene and Florida viral isolate. In Chapter 3, I suggest that the Florida viral isolate may be a remnant viral type, which existed before a new viral type spread in response to the increasing frequency of the resistant *ref(2)P* allele in *D. melanogaster* populations. Viruses from before the sweep would be affected by the *ref(2)P* polymorphism, while viruses that have themselves swept in response to the resistant *ref(2)P* allele will not. It would be interesting to compare the Florida viral isolate to one of the other viruses from Chapter 4, to identify which gene (or genes) in the virus is interacting with *ref(2)P* gene in the fly.

In Chapter 6, I describe a cluster of hyper-mutations that are most likely caused by ADARs. Despite failing to find further evidence of hypermutation in other wild virus isolates, there is a growing body of anecdotal evidence that suggests that viruses are being edited by ADARs (Cattaneo *et al.* 1988, Felder *et al.* 1994, Hajjar & Linial 1995, Kawamura *et al.* 2008, Kumar & Carmichael 1997, Ohara *et al.* 1984, Sharmeen *et al.* 1991, Wong & Lazinski 2002). Furthermore, very recent evidence that a large number of the small-interfering (si)RNAs—produced by the RNAi pathway to cleave

target RNAs—have evidence of ADAR-editing suggest that ADARs may be involved in the RNAi pathway (Kawamura *et al.* 2008). However, I failed to detect siRNAs generated from sigma viral sequence, suggesting that the sigma virus is not being processed by the RNAi pathway (see Appendix B).

Despite this result, the role of ADARs in the RNAi pathway deserves further research. It is well known that RNAi is an important defense against viruses and transposable elements (TEs), and viruses and TEs are processed by different branches of the RNAi-pathways. However, it is not understood how these different pathways recognise the different dsRNA targets. One possibility is that ADARs play an important role in 'tagging' viral sequences so they are processed by the relevant RNAi pathway. One possible approach to test this idea, is to examine a wide range of siRNAs generated in cell culture to investigate whether ADAR-editing is seen in siRNA generated from both TEs and viral sequence or just viral sequence.

Bibliography

Agrawal, A. & Lively, C. M. (2002). Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research* **4**, 79–90.

Avila, A., Silverman, N., Diaz-Meco, M. T. & Moscat, J. (2002). The Drosophila Atypical Protein Kinase C-Ref(2)P Complex Constitutes a Conserved Module for Signaling in the Toll Pathway. *Molecular and Cellular Biology* **22**, 8787–8795.

Bangham, J., Jiggins, F. M. & Lemaitre, B. (2006). Insect Immunity: The Post-Genomic Era. *Immunity* **25**, 1–5.

Bangham, J., Obbard, D., Kim, K., Haddrill, P. & Jiggins, F. (2007). The age and evolution of an antiviral resistance mutation in Drosophila melanogaster. *Proceedings of the Royal Society B: Biological Sciences* **274**, 2027–2034.

Bangham, J., Kim, K., Webster, C. L. & Jiggins, F. M. (2008*a*). Genetic Variation Affecting Host-Parasite Interactions: Different Genes Affect Different Aspects of Sigma Virus Replication and Transmission in Drosophila melanogaster. *Genetics* **178**, 2191–2199.

Bangham, J., Knott, S., Kim, K., Young, R. & Jiggins, F. (2008*b*). Genetic variation affecting host-parasite interactions: major-effect quantitative trait loci affect the transmission of sigma virus in Drosophila melanogaster. *Molecular Ecology* **17**, 3800–3807.

Bass, B. L. (2000). Double-stranded RNA as a template for gene silencing. *Cell* **101**, 235–238.

Bass, B. L. (2002). RNA editing by adenosine deaminases that act on RNA. *Annual Review of Biochemistry* **71**, 817–846.

Bass, B. L. & Weintraub, H. (1988). An Unwinding Activity That Covalently Modifies Its Double-Stranded-Rna Substrate. *Cell* **55**, 1089–1098.

Bell, G. & Smith, J. (1987). Short-term selection for recombination among mutually antagonistic species. *Nature* **328**, 66–68.

Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. C. (2001). Evolutionary dynamics of plant R-genes. *Science* **292**, 2281–2285.

Berkalof, A., Breglian, J. & Ohanessi, A. (1965). Mise En Evidence De Virions Dans Des Drosophiles Infectees Par Le Virus Hereditaire Sigma. *Comptes Rendus Hebdomadaires Des Seances De L'Academie Des Sciences* **260**, 5956–5959.

Boyd, K. R. & Whitakerdowling, P. (1988). Flanders Virus-Replication and Protein-Synthesis. *Virology* **163**, 349–358.

Bras, F., Teninges, D. & Dezelee, S. (1994). Sequences of the N-Gene and M-Gene of the Sigma-Virus of Drosophila and Evolutionary Comparison. *Virology* **200**, 189–199.

Brun, P. & Plus, N. (1998). The Viruses of Drosophila, 2d. In Ashburner, M. & Wright, T., eds., *The Genetics and Biology of Drosophila*, pp. 625–702. Academic Press, London.

Brusa, R., Zimmermann, F., Koh, D. S., Feldmeyer, D., Gass, P., Seeburg, P. H. & Sprengel, R. (1995). Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science* **270**, 1677–1680.

Burdon, J. J. (1987). Phenotypic And Genetic Patterns Of Resistance To The Pathogen Phakopsora-Pachyrhizi In Populations Of Glycine-Canescens. *Oecologia* **73**, 257–267.

Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., SandersBush, E. & Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303–308.

Bussereau, F. (1973). Vesicular Stomatitis-Virus Induced Co2 Sensitivity in Drosophila melanogaster .3. Study of Different Serotypes of Virus. *Annales De Microbiologie* **A124**, 535–554.

Bussereau, F. (1975). Co2 Sensitivity Induced by 2 Rhabdoviruses, Piry and Chandipura, in Drosophila melanogaster. *Annales De Microbiologie* **B126**, 389–403.

Caracristi, G. & Schlotterer, C. (2003). Genetic differentiation between American and European Drosophila melanogaster populations could be attributed to admixture of African alleles. *Molecular Biology and Evolution* **20**, 792–799.

Carius, H., Little, T. & Ebert, D. (2001). Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution* **55**, 1136–1145.

Carpenter, J. A., Obbard, D., Maside, X. & Jiggins, F. M. (2007). The recent spread of a vertically transmitted virus through populations of Drosophila melanogaster. *Molecular Ecology* **16**, 3947–3954.

Carre-Mlouka, A., Gaumer, S., Gay, P., Petitjean, A. M., Coulondre, C., Dru, P., Bras, F., Dezelee, S. & Contamine, D. (2007). Control of Sigma Virus Multiplication by the ref(2)P Gene of Drosophila melanogaster: An in Vivo Study of the PB1 Domain of Ref(2)P. *Genetics* **176**, 409–419.

Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., Termeulen, V. & Billeter, M. A. (1988). Biased Hypermutation and Other Genetic Changes in Defective Measles Viruses in Human-Brain Infections. *Cell* **55**, 255–265.

Chaboudez, P. & Burdon, J. J. (1995). Frequency-Dependent Selection In A Wild Plant-Pathogen System. *Oecologia* **102**, 490–493.

Cherry, S. & Silverman, N. (2006). Host-pathogen interactions in drosophila: new tricks from an old friend. *Nature Immunology* **7**, 911–917.

Christian, P. (1987). *Studies of Drosophila C and A viruses in Australian populations of Drosophila melanogaster*. Ph.D. thesis, Australian National University, Canberra.

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**, 525–533.

Davis, P. L., Holmes, E. C., Larrous, F., Van der Poel, W. H. M., Tjornehoj, K., Alonso, W. J. & Bourhy, H. (2005). Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. *Journal of Virology* **79**, 10487–10497.

Dezelee, S., Bras, F., Contamine, D., Lopezferber, M., Segretain, D. & Teninges, D. (1989). Molecular Analysis of Ref(2)P, a Drosophila Gene Implicated in Sigma Rhabdovirus Multiplication and Necessary for Male-Fertility. *Embo Journal* **8**, 3437–3446.

Ding, S.-W. & Voinnet, O. (2007). Antiviral immunity directed by small RNAs. *Cell* **130**, 413–426.

Dinoor, A. (1977). Oat Crown Rust Resistance In Israel. *Annals Of The New York Academy Of Sciences* **287**, 357–366.

Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J. A. & Imler, J.-L. (2005). The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila. *Nature Immunology* **6**, 946–953.

Dru, P., Bras, F., Dezelee, S., Gay, P., Petitjean, A. M., Pierre-Deneubourg, A., Teninges, D. & Contamine, D. (1993). Unusual Variability of the Drosophila

melanogaster ref(2)P Protein Which Controls the Multiplication of Sigma Rhabdovirus. *Genetics* **133**, 943–954.

Dybdahl, M. F. & Lively, C. M. (1998). Host-parasite coevolution: Evidence for rare advantage and time-lagged selection in a natural population. *Evolution* **52**, 1057–1066.

Felder, M. P., Laugier, D., Yatsula, B., Dezelee, P., Calothy, G. & Marx, M. (1994). Functional and Biological Properties of an Avian Variant Long Terminal Repeat Containing Multiple a-Conversion to G-Conversion in the U3 Sequence. *Journal of Virology* **68**, 4759–4767.

Felix, R., Guzman, J. & de G. Arellano, A. (1971). Distribution of CO2 sensitivity (sigma virus) in an urban population of D. melanogaster from Mexico City. II. Low dispersal, a factor which explains differences among locations. *Drosophila Information Service* **47**, 105–109.

Ferrari, J., Muller, C. B., Kraaijeveld, A. R. & Godfray, H. C. J. (2001). Clonal variation and covariation in aphid resistance to parasitoids and a pathogen. *Evolution* **55**, 1805–1814.

Fleuriet, A. (1976*a*). Presence Of Hereditary Rhabdovirus Sigma And Polymorphism For A Gene For Resistance To This Virus In Natural-Populations Of Drosophila melanogaster. *Evolution* **30**, 735–739.

Fleuriet, A. (1976*b*). Presence of the hereditary rhabdovirus sigma and polymorphism for a gene for resistance to this virus in natural populations of Drosophila melanogaster. *Evolution* **30**, 735–739.

Fleuriet, A. (1978). Reproduction In Cage Populations Of A Polymorphism Regularly Observed In Natural-Populations Of Drosophila melanogaster In France. *Genetics* **88**, 755–759.

Fleuriet, A. (1981). Effect Of Overwintering On The Frequency Of Flies Infected By The Rhabdovirus Sigma In Experimental Populations Of Drosophila melanogaster. *Archives Of Virology* **69**, 253–260.

Fleuriet, A. (1986). Perpetuation Of The Hereditary Sigma-Virus In Populations Of Its Host, Drosophila melanogaster - Geographical Analysis Of Correlated Polymorphisms. *Genetica* **70**, 167–177.

Fleuriet, A. (1988). Maintenance Of A Hereditary Virus - The Sigma-Virus In Populations Of Its Host, Drosophila melanogaster. *Evolutionary Biology* **23**, 1–30.

Fleuriet, A. (1990). Evolution Of Natural-Populations In The Drosophila melanogaster Sigma System .2. Northern And Central France. *Genetica* **81**, 33–41.

Fleuriet, A. (1991). Polymorphism Of The Hereditary Rhabdovirus Sigma In Wild Populations Of Its Host, Drosophila melanogaster. *Genetics Selection Evolution* **23**, 25–37.

Fleuriet, A. & Sperlich, D. (1992). Evolution Of The Drosophila melanogaster Sigma Virus System In A Natural-Population From Tubingen. *Theoretical And Applied Genetics* **85**, 186–189.

Fleuriet, A., Periquet, G. & Anxolabehere, D. (1990). Evolution Of Natural-Populations In The Drosophila melanogaster Sigma Virus System 1. Languedoc (Southern France). *Genetica* **81**, 21–31.

Flor, H. (1955). Host-parasite interaction in Flax Rust - its genetics and other implications. *Phytopathology* **45**, 680–685.

Frank, S. (1996). Problems inferring the specificity of plantpathogen genetics. *Evolutionary Ecology* **10**, 323–325.

Frank, S. A. (1994). Recognition And Polymorphism In Host-Parasite Genetics. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences* **346**, 283–293.

Garcia-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology* **39**, 157–186.

Garcia-Cortes, L. & Sorensen, D. (2001). Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. *Genetics Selection Evolution* **33**, 443–452.

Gay, P. (1968*a*). Adaptation of Viral Population to Reproducing in Refractory Host. *Annales De Genetique* **11**, 98.

Gay, P. (1968*b*). Hereditary Transmission of Sigma Virus by Drosophila Bearing Refractaire Gene. *Annales de L'institut Pasteur* **115**, 321.

Gay, P. (1978). Drosophila Genes Which Intervene In Multiplication Of Sigma Virus. *Molecular and General Genetics* **159**, 269–283.

Gooding, R. (1996). Genetic variation in arthropod vectors of disease-causing organisms: Obstacles and opportunities. *Clinical Microbiology Reviews* **9**, 301–306. ISSN 0893-8512.

Habayeb, M. S., Ekengren, S. K. & Hultmark, D. (2006). Nora virus, a persistent virus in Drosophila, defines a new picorna-like virus family. *Journal of General Virology* **87**, 3045–3051.

Habayeb, M. S., Ekengren, S. K. & Hultmark, D. (2007). Nora virus, a persistent virus in Drosophila, defines a new picorna-like virus family. *Journal of General Virology* **88**, 3493–3495.

Hajjar, A. M. & Linial, M. L. (1995). Modification of Retroviral Rna by Double-Stranded-Rna Adenosine-Deaminase. *Journal of Virology* **69**, 5878–5882.

Haldane, J. (1949). Disease and evolution. *La Ricerca Scientifica* **19**, 68–76.

Haldane, J. (1954). The statics of evolution. In Huxley, J., Hardy, A. & Ford, E., eds., *Evolution as a process*, pp. 109–121. Allen and Unwin.

Hasegawa, M., Kishino, H. & Yano, T. A. (1985). Dating of the Human Ape Splitting by a Molecular Clock of Mitochondrial-DNA. *Journal of Molecular Evolution* **22**, 160–174.

Heaton, L. A., Hillman, B. I., Hunter, B. G., Zuidema, D. & Jackson, A. O. (1989). Physical Map of the Genome of Sonchus Yellow Net Virus, a Plant Rhabdovirus with 6 Genes and Conserved Gene Junction Sequences. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 8665–8668.

Hedrick, P. W. (1999). Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313–318.

Henter, H. (1995). The potential for coevolution in a host-parasitoid system 2. Genetic-variation within a population of wasps in the ability to parasitize an aphid host. *Evolution* **49**, 439–445.

Henter, H. & Via, S. (1995). The potential for coevolution in a host-parasitoid system 1. Genetic-variation within an aphid population in susceptibility to a parasitic wasp. *Evolution* **49**, 427–438.

Hirschhorn, J. N. & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95–108.

Hogenhout, S. A., Redinbaugh, M. G. & Ammar, E.-D. (2003). Plant and animal rhabdovirus host range: a bug's view. *Trends in Microbiology* **11**, 264–271.

Holub, E. B. (2001). The arms race is ancient history in Arabidopsis, the wildflower. *Nature Reviews Genetics* **2**, 1471–0056.

Hoopengardner, B., Bhalla, T., Staber, C. & Reenan, R. (2003). Nervous system targets of RNA editing identified by comparative genomics. *Science* **301**, 832–836.

Houck, M. A., Clark, J. B., Peterson, K. R. & Kidwell, M. G. (1991). Possible Horizontal Transfer of Drosophila Genes by the Mite Proctolaelaps-Regalis. *Science* **253**, 1125–1129.

Howard, R. & Lively, C. (1994). Parasitism, mutation accumulation and the maintaiance of sex. *Nature* **367**, 554–557.

Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014.

Hudson, R. R., Boos, D. D. & Kaplan, N. L. (1992). A Statistical Test for Detecting Geographic Subdivision. *Molecular Biology and Evolution* **9**, 138–151.

Hughes, A. L. & Yeager, M. (1997). Coordinated amino acid changes in the evolution of mammalian defensins. *Journal of Molecular Evolution* **44**, 675–682. ISSN 0022-2844.

Hughes, G. J., Orciari, L. A. & Rupprecht, C. E. (2005). Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene. *Journal of General Virology* **86**, 1467–1474. Part 5.

Jakobsen, I. B. & Easteal, S. (1996). A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *Computer Applications in the Biosciences* **12**, 291–295.

Jiggins, F. & Hurst, G. (2003). Purifying Selection on Drosophila melanogaster Peptidoglycan Recognition Proteins. *Journal of Molecular Evolution* **57**, 598–605.

Jiggins, F. M. & Kim, K.-W. (2005). The Evolution of Antifungal Peptides in Drosophila. *Genetics* **171**, 1847–1859.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T. N., Siomi, M. C. & Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute-2 in somatic cells. *Nature* p. Advanced online.

Keegan, L. P., Gallo, A. & O'Connell, M. A. (2001). The many roles of an RNA editor. *Nature Reviews Genetics* **2**, 869–878.

Keegan, L. P., Brindle, J., Gallo, A., Leroy, A., Reenan, R. A. & O'Connell, M. A. (2005). Tuning of RNA editing by ADAR is required in Drosophila. *Embo Journal* **24**, 2183–2193.

Kidwell, M. G., Kidwell, J. F. & Sved, J. A. (1977). Hybrid Dysgenesis in Drosophila melanogaster - Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics* **86**, 813–833.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Prudhomme, N. & Bucheton, A. (1994). Retroviruses in Invertebrates - the Gypsy Retrotransposon Is Apparently an Infectious Retrovirus of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1285–1289.

Knight, S. W. & Bass, B. L. (2002). The role of RNA editing by ADARs in RNAi. *Molecular Cell* **10**, 809–817.

Kraaijeveld, A. & van Alphen, J. (1995). Geographic variation in encapsulation ability of Drosophila melanogaster larvae and evidence for parasitoid-specific components. *Evolutionary Ecology* **9**, 10–17.

Kraaijeveld, A. R. & Godfray, H. C. J. (1997). Trade-off between parasitoid resistance and larval competitive ability in Drosophila melanogaster. *Nature* **389**, 278–280.

Kraaijeveld, A. R., Van Alphen, J. J. M. & Godfray, H. C. J. (1998). The coevolution of host resistance and parasitoid virulence. *Parasitology* **116**, S29–S45.

Kumar, M. & Carmichael, G. G. (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3542–3547.

Kurath, G., Ahern, K. G., Pearson, G. D. & Leong, J. C. (1985). Molecular-Cloning of the 6 Messenger-Rna Species of Infectious Hematopoietic Necrosis Virus, a Fish Rhabdovirus, and Gene Order Determination by R-Loop Mapping. *Journal of Virology* **53**, 469–476.

Lachaise, D. & Silvain, J. F. (2004). How two Afrotropical endemics made two cosmopolitan human commensals: the D. melanogaster-D. simulans palaeogeographic riddle. *Genetica* **120**, 17–39.

Lambrechts, L., Halbert, J., Durand, P., Gouagna, L. C. & Koella, J. C. (2005). Host genotype by parasite genotype interactions underlying the resistance of anopheline mosquitoes to Plasmodium falciparum. *Malaria Journal* **4**, 1475–2875.

Landesdevauchelle, C., Bras, F., Dezelee, S. & Teninges, D. (1995). Gene 2 of the Sigma Rhabdovirus Genome Encodes the P-Protein, and Gene-3 Encodes a Protein Related to the Reverse-Transcriptase of Retroelements. *Virology* **213**, 300–312.

Lazzaro, B. P., Sceurman, B. K. & Clark, A. G. (2004). Genetic basis of natural variation in D melanogaster antibacterial immunity. *Science* **303**, 1873–1876.

Lehmann, K. A. & Bass, B. L. (2000). Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875–12884.

Lemaitre, B. (2004). The road to Toll. *Nature Reviews Immunology* **4**, 521–527.

Lemaitre, B. & Hoffmann, J. (2007). The Host Defense of Drosophila melanogaster. *Annual Review of Immunology* **25**, 697–743.

L'Heritier, P. (1948). Sensitivity to CO2 in Drosophila; a review. *Heredity* **2**, 325–48.

L'Heritier, P. & Teissier, G. (1937). Une anomolie physiologique hereditaire chez la Drosophile. *C.R. Acad. Sci. Paris* **205**, 1099–1101.

Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563.

McVey, D. (1990). Reaction of 578 spring spelt wheat accessions to 35 races of wheat-stem rust. *Crop Science* **30**, 1001–1005.

Nakamura, N. (1978). Dosage Effects of Non Permissive Allele of Drosophila Ref(2)P Gene on Sensitive Strains of Sigma Virus. *Molecular and General Genetics* **159**, 285–292.

Nei, M. & Gojobori, T. (1986). Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Nishikura, K., Yoo, C., Kim, U., Murray, J. M., Estes, P. A., Cash, F. E. & Liebhaber, S. A. (1991). Substrate-Specificity of the Dsrna Unwinding Modifying Activity. *Embo Journal* **10**, 3523–3532.

Obbard, D. J., Jiggins, F. M., Halligan, D. L. & Little, T. J. (2006). Natural Selection Drives Extremely Rapid Evolution in Antiviral RNAi Genes. *Current Biology* **16**, 580–585.

Ohanessian-Guillemain, A. (1963). Study of genetic factors controlling relations between the sigma virus and Drosophila, its host. *Ann Genet* **5**, 1–64.

Ohara, P. J., Nichol, S. T., Horodyski, F. M. & Holland, J. J. (1984). Vesicular Stomatitis-Virus Defective Interfering Particles Can Contain Extensive Genomic Sequence Rearrangements and Base Substitutions. *Cell* **36**, 915–924.

Palladino, M. J., Keegan, L. P., O'Connell, M. A. & Reenan, R. A. (2000). A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. *Cell* **102**, 437–449.

Parks, A. L., Cook, K. R., Belvin, M., Dompe, N. A., Fawcett, R., Huppert, K., Tan, L. R., Winter, C. G., Bogart, K. P., Deal, J. E. *et al.* (2004). Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. *Nature Genetics* **36**, 288–292.

Patterson, J. B. & Samuel, C. E. (1995). Expression and Regulation by Interferon of a Double-Stranded-Rna-Specific Adenosine-Deaminase from Human-Cells - Evidence for 2 Forms of the Deaminase. *Molecular and Cellular Biology* **15**, 5376–5388.

Patton, D. E., Silva, T. & Bezanilla, F. (1997). RNA editing generates a diverse array of transcripts encoding squid Kv2 K+ channels with altered functional properties. *Neuron* **19**, 711–722.

Pestka, S., Langer, J. A., Zoon, K. C. & Samuel, C. E. (1987). Interferons and Their Actions. *Annual Review of Biochemistry* **56**, 727–777.

Plus, N. & Duthoit, J. (1969). Un nouveau virus de Drosophila melanogaster, le virus. *Proceedings of the National Academy of Sciences of the United States of America* **268**, 2313–2315.

Plus, N., Croizier, G., Jousset, F. X. & David, J. (1975). Picornaviruses Of Laboratory And Wild Drosophila melanogaster - Geographical Distribution And Serotypic Composition. *Annales De Microbiologie* **A126**, 107.

Plus, N., Croizier, G., Veyrunes, J. C. & David, J. (1976). A comparison of bouyant density and polypeptides of Drosophila P, C and A viruses. *Intervirology* **7**, 346–350.

Polson, A. G. & Bass, B. L. (1994). Preferential Selection of Adenosines for Modification by Double-Stranded-RNA Adenosine-Deaminase. *Embo Journal* **13**, 5701–5711.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13757–13762.

Richardmolard, C., Blondel, D., Wyers, F. & Dezelee, S. (1984). Sigma Virus - Growth in Drosophila melanogaster Cell-Culture - Purification - Protein-Composition and Localization. *Journal of General Virology* **65**, 91–99.

Riehle, M. M., Markianos, K., Niare, O., Xu, J., Li, J., Toure, A. M., Podiougou, B., Oduol, F., Diawara, S., Diallo, M. *et al.* (2006). Natural Malaria Infection in Anopheles gambiae Is Regulated by a Single Genomic Control Region. *Science* **312**, 577–579.

Roxstrom-Lindquist, K., Terenius, O. & Faye, I. (2004). Parasite-specific immune response in adult Drosophila melanogaster: a genomic study. *EMBO Reports* **5**, 207–212.

Rozen, S. & Skaletsky, H. (2000). *Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology.*. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ.

Rueda, P., Garciabarreno, B. & Melero, J. A. (1994). Loss of Conserved Cysteine Residues in the Attachment (G) Glycoprotein of 2 Human Respiratory Syncytial Virus Escape Mutants That Contain Multiple a-G Substitutions (Hypermutations). *Virology* **198**, 653–662.

Rutschmann, S., Jung, A. C., Hetru, C., Reichhart, J.-M., Hoffmann, J. A. & Ferrandon, D. (2000). The Rel Protein DIF Mediates the Antifungal but Not the Antibacterial Host Defense in Drosophila. *Immunity* **12**, 569–580.

Salvaudon, L., Héraudet, V. & Shykoff, J. A. (2007). Genotype-specific interactions and the trade-off between host and parasite fitness. *BMC Evolutionary Biology* **7**, 1471–2148.

Sanz, L., Sanchez, P., Lallena, M. J., Diaz-Meco, M. T. & Moscat, J. (1999). The interaction of p62 with RIP links the atypical PKCs to NF-κB activation. *Embo Journal* **18**, 3044–3053.

Sato, S., Wong, S. K. & Lazinski, D. W. (2001). Hepatitis delta virus minimal substrates competent for editing by ADAR1 and ADAR2. *Journal of Virology* **75**, 8547–8555.

Scadden, A. D. J. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nature Structural and Molecular Biology* **12**, 489–496.

Scadden, A. D. J. & Smith, C. W. J. (1997). A ribonuclease specific for inosine-containing RNA: A potential role in antiviral defence? *Embo Journal* **16**, 2140–2149.

Scadden, A. D. J. & Smith, C. W. J. (2001*a*). RNAi is antagonized by A-G hyper-editing. *EMBO Reports* **21**, 1107–1111.

Scadden, A. D. J. & Smith, C. W. J. (2001*b*). Specific cleavage of hyper-edited dsRNAs. *Embo Journal* **20**, 4243–4252.

Schlenke, T. A. & Begun, D. J. (2003). Natural selection drives drosophila immune system evolution. *Genetics* **164**, 1471–1480.

Seeburg, P. H. (1996). The role of RNA editing in controlling glutamate receptor channel properties. *Journal of Neurochemistry* **66**, 1–5.

Semenov, E. P. & Pak, W. L. (1999). Diversification of Drosophila chloride channel gene by multiple posttranscriptional mRNA modifications. *Journal of Neurochemistry* **72**, 66–72.

Sharmeen, L., Bass, B., Sonenberg, N., Weintraub, H. & Groudine, M. (1991). Tat-Dependent Adenosine-to-Inosine Modification of Wild-Type Transactivation Response Rna. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 8096–8100.

Shen, B. H., Liu, H., Skolnik, E. Y. & Manley, J. L. (2001). Physical and functional interactions between Drosophila TRAF2 and Pelle kinase contribute to Dorsal activation. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8596–8601.

Shimodaira, H. & Hasegawa, M. (1999*a*). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution* **16**, 1114–1116.

Shimodaira, H. & Hasegawa, M. (1999*b*). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution* **16**, 1114–1116.

Siva-Jothy, M. T., Tsubaki, Y., Hooper, R. E. & Plaistow, S. J. (2001). Investment in immune function under chronic and acute immune challenge in an insect. *Physiological Entomology* **26**, 1–5.

Smith, L., Peixoto, A. & Hall, J. (1998). RNA editing in the Drosophila DMCA1A calcium-channel alpha 1 subunit transcript. *Journal of Neurogenetics* **12**, 227–240.

Srinivasappa, J., Saegusa, J., Prabhakar, B. S., Gentry, M. K., Buchmeier, M. J., Wiktor, T. J., Koprowski, H., Oldstone, M. B. & Notkins, A. L. (1986). Molecular mimicry: frequency of reactivity of monoclonal antiviral antibodies with normal tissues. *The Journal of Virology* **57**, 397–401.

Swofford, D. L. (2002). Phylogenetic analysis using parsimony. Version 4.0b10 for Microsoft Windows. Sinauer Associates, Sunderland, Mass.

Teninges, D. & Brasherreng, F. (1987). Rhabdovirus Sigma, the Hereditary Co2 Sensitivity Agent of Drosophila - Nucleotide-Sequence of a Cdna Clone Encoding the Glycoprotein. *Journal of General Virology* **68**, 2625–2638. Part 10.

Teninges, D. & Plus, N. (1972). P virus of Drosophila melanogaster, as a new picornavirus. *Journal of General Virology* **16**, 103–109.

Thoetkiattikul, H., Beck, M. H. & Strand, M. R. (2005). Inhibitor NF κ B-like proteins from a polydnavirus inhibit NF κ B activation and suppress the insect immune response. *Proceedings of the National Academy of Sciences* **102**, 11426–11431.

Tinsley, M., Blanford, S. & Jiggins, F. (2006). Genetic variation in the susceptibility of Drosphila melanogaster to biopesticide pathogens. *Parasitology* **132**, 767–773.

Tonkin, L. A. & Bass, B. L. (2003). Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* **302**, 1725–1725.

Turelli, M. & Hoffmann, A. A. (1991). Rapid Spread of an Inherited Incompatibility Factor in California Drosophila. *Nature* **353**, 440–442.

Wasserman, S. (2004). Nature's fortress against infection. *Nature Immunology* **5**, 474–475.

Wayne, M., 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**, 191–199.

Webster, R., Saghaimaroof, M. & Allard, R. (1986). Evolutionary response of barley composite cross to Rhynchosporium-secalis analysed by pathogenic complexity and by gene-by-race relationships. *Phytopathology* **76**, 661–668.

Williamson, D. (1961). Carbon Dioxide Sensitivity in Drosophila affinis and Drosophila athabasca. *Genetics* **46**, 1053–1060.

Wong, S. K. & Lazinski, D. W. (2002). Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proceedings of the National Academy of Sciences* **99**, 15118–15123.

Wyers, F., Dru, P., Simonet, B. & Contamine, D. (1993). Immunological Cross-Reactions and Interactions between the Drosophila melanogaster Ref(2)P-Protein and Sigma Rhabdovirus Proteins. *Journal of Virology* **67**, 3208–3216.

Zambon, R. A., Nandakumar, M., Vakharia, V. N. & Wu, L. P. (2005). The Toll pathway is important for an antiviral response in Drosophila. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **102**, 7257–7262.

A Appendix

Tables calculated from paternal transmission rates with a model allowing an interaction between *ref(2)P* and the Florida viral isolate to allow comparisons with those estimated from the former model that did not allow this interaction.

Table A.1: G-matrices for chromosome two. 95% confidence limits in brackets.

Chromosome two with	[USA]	[UK]	[SPA]	[FRA]	[GRE]
[USA]	**3.594** (1.913, 6.455)				
[UK]	1.165	**1.848** (0.805, 3.253)			
[SPN]	0.956	1.366	**1.380** (0.708, 2.512)		
[FRA]	1.163	0.801	0.646	**0.679** (0.215, 1.77)	
[GRE]	1.273	0.922	0.812	0.598	**0.864** (0.249, 2.564)

Table A.2: Eigenvectors

Chromosome two

	[,PC1]	[,PC2]	[,PC3]	[,PC4]	[,PC5]
[USA,]	0.6793872	0.67112081	-0.04416171	0.28486660	-0.0702191
[UK,]	0.4433416	-0.55764435	-0.64136597	0.09041373	-0.2701037
[SPA,]	0.3712945	-0.48307848	0.62815872	0.41238361	0.2532428
[FRA,]	0.2974227	-0.01853073	-0.18266712	-0.49843678	0.7933410
[GRE,]	0.3399432	-0.07015299	0.39843291	-0.70155322	-0.4781124

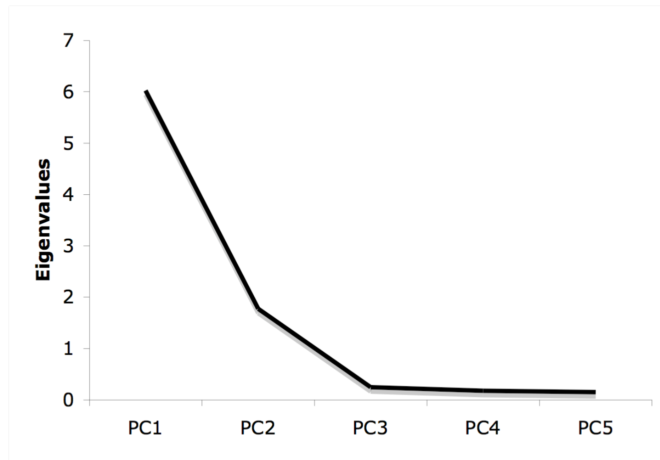


Figure A.1: The Scree plot displays the eigenvalues for chromosome two.

Table A.3: Response to selection on each of the 5 viruses (S=1.0) using the breeder's equation.

Chromosome two

	Viruses selected on				
	Florida	UK	Spain	France	Greece
Florida	3.594	1.166	0.956	1.164	1.273
UK	1.166	1.848	1.367	0.801	0.923
Spain	0.956	1.367	1.380	0.647	0.813
France	1.164	0.801	0.647	0.679	0.599
Greece	1.273	0.923	0.813	0.599	0.865

B Appendix

B.1 Is the sigma virus processed by RNAi?

The work described in this Appendix was carried out by myself, except for the siRNA assay, which was carried out by Elizabeth Bayne.

B.2 Introduction

RNA interference (RNAi) is a group of evolutionary conserved mechanisms that use short RNAs (ca. 20-30 nucleotides) to recognize and degrade complementary nucleic acids. One arm of the RNAi pathways processes viruses by recognising double stranded (ds)RNA in cells. dsRNA is relatively unusual, and typically alerts the host to the presence of a viral infection. Viruses, once detected, are cut by Dicer into short 21-24 nucleotide fragments called short-interfering RNAs (siRNAs, also known as viRNAs when they are derived from viruses: Ding & Voinnet (2007). These are then loaded into an Argonaute-containing effector complex (RISC; the RNA-Induced Silencing Complex), and one strand of the viRNA is cleaved and degraded. The active Argonaute complex then cleaves viral RNA with the complimentary sequence to the viRNA.

In this Appendix, I investigate whether the sigma virus is recognised and processed by the RNAi pathway. To do this I extracted viral RNA from two *Drosphila* lines infected with the sigma virus and one uninfected line. Viral RNA from these lines was run out on a gel and probed with a sigma-virus specific probe. I expect that if the sigma virus is processed by the RNAi pathway, the sigma-specific probe will bind to siRNA that have been generated by cleaving the viral sequence into 21-24 nucleotide fragments.

B.3 Methods and materials

B.3.1 RNA extraction

RNA was extracted from 20 flies with Trizol (see Chapter 2 for methods) from two sigma-infected line and one -uninfected line. Two different viral isolates were used: A3 (supplied by D. Contamine) and E24 (collected in Essex, see Carpenter *et al.* (2007)).

B.3.2 siRNA analysis

RNA was mixed with an equal volume of FDE sample buffer (formamide containing 10mM EDTA, 1mg/ml xylene cyanol, 1mg/ml bromophenol blue), denatured by incubating at 65°C for 15 minutes, and then placed on ice until loading. As a positive control a sample containing an oligo (1μM) with complementarity to the Sigma probe was also prepared. An 8% polyacrylamide gel was prepared using the SequaGel system (National Diagnostics), and pre-run for 30 minutes at 150V prior to loading. The loaded gel was run at 300V for 2 hours in 0.5x TBE in Hoefer SE600 Ruby apparatus. After running the gel was equilibrated by soaking in 10mM sodium phosphate pH7 for 10 minutes then in 20xSSC for 10 minutes. The RNA was transferred from the gel to Hybond-NX membrane (GE Healthcare) by capillary blotting in 20x SSC for at least 16 hours. The blot was then UV-cross-linked in a Stratalinker (Stratagene, 2400mJ), and pre-hybridised in hybridisation buffer (0.5M Na phosphate pH7.2, 10mM EDTA, 7% SDS) at 42°C for 2 hours prior to addition of the probe.

For the Sigma probe, 1μl of each of three Sigma PCR products (10-20ng/ul) were mixed with 8μl of water and denatured by incubating at 95°C for 10 minutes and chilled on ice. To this 4μl of High Prime mix (Roche) and 5μl (50uCi) of [α32P]dCTP (50uCi) were added, and the reaction was incubated at 37°C for 1 hour. Unincorporated nucleotides were removed using a G25 sephadex spin column (GE healthcare) and the probe was denatured at 95°C for 10 minutes before adding to the hybridisation buffer. Hybridisation proceeded at 42°C overnight. The blot was then washed 2x 15 minutes in wash buffer (2xSSC, 0.2%SDS) at 50°C and then visualised on a Storm PhosphorImager (Molecular Dynamics) following overnight exposure to a

phosphoscreen. The blot was stripped prior to re-probing by washing 4x 5 minutes in boiling 0.2% SDS.

For the miR-8 probe, 1μl of oligo (100μM) complementary to miR-8 was mixed with 9.5μl water, 1.5μl PNK buffer, 1μl PNK (Promega) and 2μl (20μCi) of ATP and incubated at 37°C for 1 hour. The probe was then added to the blot in hybridisation buffer.

B.4 Results

First the samples were run out on a gel and probed with a miRNA probe. The miRNA is a positive control to ensure that short RNAs of ~22nt were successfully extracted (and can be resolved as a band on the gel). Next, the samples were probed with a sigma-virus specific probe. If the sigma virus is processed by the RNAi pathway then I expect to see a similar sized band in lanes marked A3 and E24, as in lane M, marked ”control oligo”. No band was seen in lanes containing either the sigma-infected or sigma-uninfected RNA. I therefore conclude that no siRNAs were generated from sigma viral sequence. This suggests that the sigma virus is not being processed by the RNAi pathway. However, low copy number of virally generate siRNAs and mismatches between them and the probe cannot be totally ruled out and further experiments are needed.

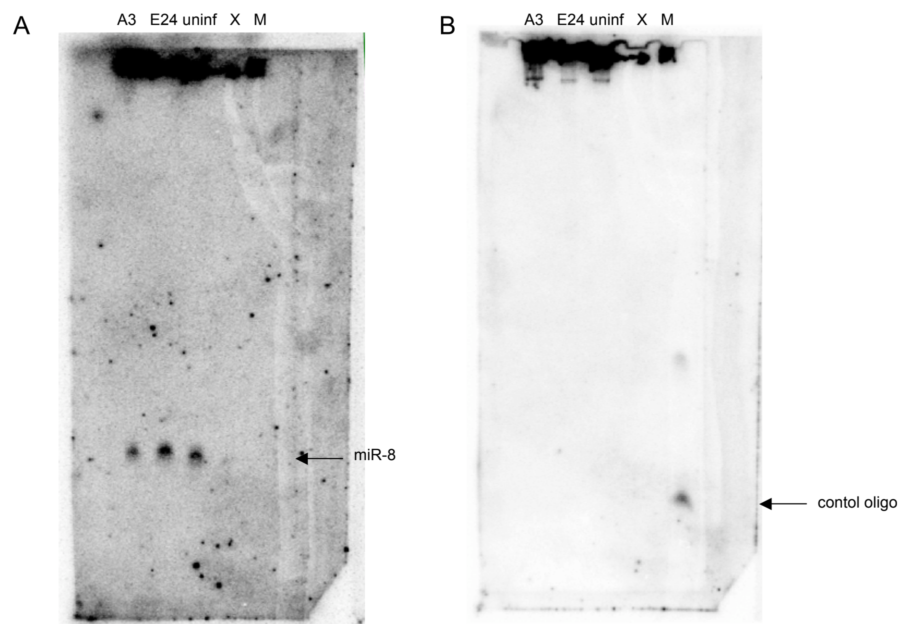


Figure B.1: A: Blot probed with miRNA-probes. B: Blot probed with sigma-virus specific probe). Samples: A3 (sigma-infected), E24 (sigma-infected), uninf (sigma-uninfected), X (empty lane) and M (PCR primers; acting as a positive control for the sigma probe)

C Publications