DROSOPHILA AND THE RHABDOVIRUS SIGMA: A MODEL SYSTEM FOR THE EVOLUTION OF VIRULENCE FOLLOWING INTERSPECIFIC HOST SHIFTS

By

LUIS FERNANDO MATOS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

To my family

ACKNOWLEDGMENTS

I thank the following people and entities who have had a positive impact on me during my doctoral program. Dr. James Maruniak, provided me with the opportunity of continuing my graduate studies and for that I am forever indebted to him. I thank my other committee members: Dr. Howard Frank and Dr. Barry Alto for valuable comments to the research proposal and this document. Dr. Marta Wayne has been simply amazing. I have benefited from her mentorship and friendship more than I can tell. She was and is an excellent example of how to be a good scientist and a good person. Finally, I thank my amazing wife because without her sustaining me through the trials and triumphs of these past few years this dissertation would not have been possible. Thank you!

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DMELSV	The wild type *Drosophila melanogaster* sigma rhabdovirus purified from the offspring of wild-caught flies
DMELSV-A	Ancestral sigma virus evolved in *Drosophila melanogaster* that was injected into naïve native and novel host
DSIMSV-E	Sigma rhabdovirus that was evolved in *D. simulans*
ADAR	Adenosine deaminases acting on RNA
PCR	Polymerase chain reaction

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DROSOPHILA AND THE RHABDOVIRUS SIGMA: A MODEL SYSTEM FOR THE EVOLUTION OF VIRULENCE FOLLOWING INTERSPECIFIC HOST-SHIFTS

By

Luis Matos

December 2012

Chair: James Maruniak
Major: Entomology and Nematology

Understanding the evolution of virulence continues to be an important area of research as we strive to prevent and cure diseases of humans and livestock. Of particular interest are the events that occur after a pathogen invades a novel host (host shift). Most of the currently accepted theory on the evolution of virulence following a host shift is based on data developed using single celled hosts in serial passage experiments. Based on these data, it is expected that after a host shift a pathogen will be highly virulent to its novel host. Studying these parameters is difficult because natural host shifts happen in the absence of observation and because the standing genetic variation present in the pathogen population immediately prior to the shift is unknown. Additionally, the host shifts of interest involve complex multicellular hosts. To address these issues I developed a tractable model system using *Drosophila* spp. and the sigma rhabdovirus which, respectively, are good models for dipteran disease vectors and pathogenic RNA viruses. Using this system, the genetic variation in wild sigma virus was determined using next generation SOLiD™ sequencing, a shift onto a novel host was induced and tracked in the laboratory and the resulting evolved virus was returned to its native host. The sigma rhabdovirus has lower genetic variation than

is typically expected of RNA viruses and a part of the variation is due to ADAR activity. The sigma was more infective on the novel host than on the native host and, as predicted, it was more virulent on the novel host. Contrary to prediction, I found no evidence of attenuation when the evolved virus was returned to its native host; the opposite was true early in the infection. These results indicate that the evolution of virulence following a host shift is more complex than expected and indicate that this new model system can be used successfully to learn about the evolution of virulence.

CHAPTER 1
INTRODUCTION

Background

Virulence, simply defined, is a pathogen's ability to cause disease in its host (KNIPE and HOWLEY 2001; LITTLE *et al.* 2010). Avirulent or mildly virulent organisms cause no or little disease, respectively, and can persist in the population through many host generations with zero or minimal adverse affects to the host (BEACH *et al.* 2009). Conversely, highly virulent organisms usually damage their hosts severely, thus inflicting fitness losses and even death of the host (KNIPE and HOWLEY 2001). The latter situation can occur when a pathogen host-shifts from its native host onto a novel host species. The human immunodeficiency virus (HIV) (RAMBAUT *et al.* 2004) and severe acute respiratory syndrome (SARS) (MULLER and MCGEER 2007) are just two examples of pathogens shifting hosts from non-human species into humans. In recent years several other pathogens have shifted into humans hosts, leading to severe disease in some cases (FRASER *et al.* 2009; MULLER and MCGEER 2007; TSUKAMOTO *et al.* 2007; VAN DER MEULEN *et al.* 2005). Unfortunately, predictions of how severely these pathogens would affect human populations were not always correct. The recent case of swine flu (H1N1 influenza) is a prime example of how inaccurate these predictions can be. The predicted deadly effect of the emerging pathogen on the human population was, thankfully, grossly overestimated as the numbers of influenza cases were similar to previous years (COLLIGNON 2010; ISAACS 2010; KELLY 2010). This error led to increased anxiety among the world's population and produced serious financial losses to industry and governments worldwide as they prepared to combat an epidemic that materialized far below the predicted case levels (FLYNN 2010). Because of the poor

predictions, vaccines were overproduced and thus wasted. Holland alone was left with nearly 20 million unused doses (DOLGIN 2010). To prevent these issues we need a better understanding of the evolution of virulence that occurs following an interspecific host shift. This knowledge will assist the scientific community as it predicts the epidemiology of emerging pathogens. Here I will review the literature on 1) the evolution of virulence, 2) introduce the model system, 3) discuss the background and significance of the project, 4) and describe the experimental design that will be used to test hypotheses about the evolution of virulence after an interspecific host shift.

Evolution of Virulence

Virulence is an old concept that originated from Latin "virulentus" and the first recorded use of "virulent" occurred in the 1400s A.D to describe something bad or poisonous (Council of Sci. Eds., 2011). This definition persisted for a long time and was even used by Charles Darwin ("The virulence of this poison…") to describe the potency of a spider's venom (DARWIN 1909-14). More recently the term virulence has been more directly associated with the level of disease that a pathogen is capable of inflicting on its host (KNIPE and HOWLEY 2001; LITTLE *et al.* 2010). Thus, modern studies on the evolution of virulence seek to understand what factors contribute to changes (increases or decreases) in a pathogen's virulence level. To this end, three major hypotheses have been formulated to explain the evolution of virulence. These hypotheses argue that the evolution of virulence is either coincidental (LEVIN and EDEN 1990), the result of short-sightedness on the part of the pathogen (LEVIN and BULL 1994) or a tradeoff (ANDERSON and MAY 1982; EWALD 1983).

Coincidental Hypothesis

The coincidental hypothesis of virulence evolution postulates that a factor's contribution to a pathogen's level of virulence within a particular host evolved for some reason other than effecting virulence in that host (LEVIN 1996; LEVIN and EDEN 1990). Levin (1996) uses Gould and Lewontin's (1979) beautiful analogy of the spandrels of San Marco to illustrate this point. That is that the spandrels were not constructed to frame the frescoes, though they do frame them wonderfully. Instead, the spandrels are a necessary structural feature without which the church would not stand (GOULD and LEWONTIN 1979). Therefore, the lack of the spandrels would negate the successful construction of the church and would eliminate the need for frescoes. In the same way, this hypothesis argues that virulence is a byproduct of some other necessary process that occurs while the pathogen colonizes the host. For example, the neurotoxins produced by *Clostridium tetani* and *C. botulinum*, both soil bacteria, did not evolve for the purpose of being highly toxic in humans, yet they are.

Short-Sighted Hypothesis

The short-sighted hypothesis for the evolution of virulence proposes that changes in pathogen virulence at the population level are a result of random mutation followed by selection occurring at the level of the individual (LEVIN and BULL 1994). To support their hypothesis Levin and Bull examine three diseases [bacterial meningitis, poliomyelitis, and the acquired immune deficiency syndrome (AIDS)] and use facts from the ecology of the respective disease-causing agents to support this hypothesis. The authors argue that the evolutionary strategy for these three pathogens is "short-sighted" because the within-host dynamics do not improve each pathogen's ability to infect and successfully exploit the next host (LEVIN and BULL 1994). In particular, the authors cite

the case of polio where the pathogen infects the neurons of the host despite the fact that this does not increase transmission. Thus, the pathogen's evolution is "short-sighted" by favoring short term intra-host success at the cost of long term inter-host transmission. In fact, this short-sightedness may yield a pathogen population that is well adapted within a particular individual host that is potentially maladapted to invade another host when or before the present host expires. Interestingly, the only empirical evidence for the short-sighted hypothesis was reported in honeybee colonies infected with a social parasite (MORITZ *et al.* 2008). Moritz et al. (2008) showed that, in fact, the parasitic bee species became extremely well adapted within a single colony (host) at the cost of being able to infect other colonies (hosts).

Trade-Off Hypothesis

The trade-off hypothesis was developed in the early 1980s (ANDERSON and MAY 1982; EWALD 1983) and holds that for a parasite to increase its ability to persist on a host it must give up something, i.e. a reduction in virulence. However, for the parasite to succeed under the assumptions of this hypothesis, the parasite must retain some minimal level of virulence and transmissibility (ANDERSON and MAY 1982; EWALD 1983). Additionally, this theory argues that the host's natural death rate, its death rate due to infection, and the rate of recovery from infection are inextricably linked. Because of this connection, the tradeoff for a decrease in virulence is a concomitant decrease in transmission such that as virulence approaches zero so too does transmission (ALIZON *et al.* 2009). Therefore, under this hypothesis the parasite cannot ever become avirulent or it will cease to be transmitted to new hosts and will go extinct. This flies in the face of the previously (pre-Anderson and May, 1983) commonly accepted avirulence hypothesis, which stated as a parasite and host co-evolve the parasite will become

progressively less virulent until it is completely avirulent (ALIZON *et al.* 2009). Therefore, the development of the trade-off hypothesis was a major turning point in the study of the evolution of virulence because it cemented the idea that the interaction between the host and the parasite is a fluid one that is affected not only by the parasite but also by the host parameters.

This hypothesis is appealing from an epidemiology management standpoint because it suggests that some measure of control might be had over parasite virulence by managing the parameters that contribute to parasite virulence (ALIZON *et al.* 2009). The idea here is that if a host/pathogen system is managed correctly, one might be able to manipulate a given parameter such that the parasite would trade-off high virulence for some other parameter, say transmissibility. In a human disease outbreak, for example, this would lead to many people (a large percentage of the population) being infected with a weekly virulent pathogen (few or no deaths) rather than fewer people (a small percentage of the population) infected with a highly virulent pathogen (many deaths). The trade-off hypothesis proposes that a delicate balance exists between virulence, transmission, and host recovery (ALIZON 2008; ANDERSON and MAY 1982). Exactly how the trade-offs in each parameter work to regulate the evolution of virulence is not yet fully understood and more research is needed to produce more data on tradeoff curves (ALIZON *et al.* 2009; BOLKER *et al.* 2010).

Alizon (2008) argues that when discussing trade-offs and parasite evolution the effects of host recovery-transmission tradeoff must be taken into account in addition to or even above virulence-transmission trade-offs. This is because the dynamics of the virus population within the host will affect host recovery and, in turn, produce a

paralleling change in the rate at which the parasite is transmitted from the recovering host to a naïve host (ALIZON 2008). This change in transmission rate is of particular importance in human diseases where either vaccination or antimicrobial drugs alter within-host virus dynamics and typically increase both the rapidity with which hosts recover and the total number of hosts recovering. Additionally, it is possible that host recovery that is aided by vaccines and antimicrobials could actually drive the evolution of more virulent strains capable of overcoming these countermeasures (ANDRE *et al.* 2006; GENTON *et al.* 2002; MACKINNON *et al.* 2008). In the case of malaria, for example, a trial vaccine selected for alternative alleles that were not targeted by the vaccine (GENTON *et al.* 2002). Whether theses alternate alleles were more virulent is not known; however, children who were previously vaccinated against one allele were more likely to become infected with the alternate allele (GENTON *et al.* 2002). This indicates that although vaccination may reduce recovery time by altering the immune status of the host, this gain in recovery is traded-off with a concomitant increase in transmission of the alternate allele in the population. Therefore, although the example with malaria supports Alizon's (2008) argument that recovery is an important parameter, it also clearly shows that unlike his model, recovery is only another piece in the puzzle that is the trade-off hypothesis of virulence evolution. To a certain extent, the three hypotheses that have been proposed to explain the evolution of virulence are not mutually exclusive. Additionally, no one hypothesis encompasses all of the diverse biological systems that have been studied to date. However, most of the cases where we see the evolution of virulence seem to follow the tradeoff hypothesis.

Experimental organisms

Understanding the evolution of virulence often requires the use of tractable model systems. Currently, one model system has been developed for the study of host-shifts by pathogens. The fungal causative agent of anther-smut disease (*Microbotryum violaceum*) was shown to host-shift from its native host *Silene alba* (white campion) to *Silene vulgaris* (bladder campion) and the ecology and genetics of this shift have been examined (ANTONOVICS *et al.* 2002). This plant-fungus system, however, is not of immediate relevance to human diseases as host-shifts from a plant to an animal vector likely are extremely rare. Here we propose to develop *Drosophila melanogaster*, *D. simulans* and the rhabdovirus sigma (DMelSV, endemic in *D. melanogaster*) as a model system for the evolution of virulence and host-shifting.

Drosophila Spp.

Insect study systems, particularly dipteran systems, are prime candidates for use as model systems in both the evolution of virulence and the mechanisms of host-shifting. There are several reasons for this: 1) many emerging diseases have dipteran vectors and 2) insects offer the possibility of using large numbers of individuals per experiment without requiring the cost and specialized facilities associated with maintaining vertebrates (SCULLY and BIDOCHKA 2006). In addition to these main reasons, using *Drosophila* spp. as hosts offers several unique features that make it a good system in which to study virus host-shifts.

Drosophila is a dipteran and dipterans such as mosquitoes, sand flies, and blackflies are all vectors of disease-causing RNA viruses (HOGENHOUT *et al.* 2003). Using the *Drosophila* as the host in a model system to study the evolution of host-shifts eliminates the need to work and special facilities necessary to keep pathogen-carrying

insects. However, dipteran vectors are extremely similar to *Drosophila* spp. in morphology, physiology, and genetics. The vectors for which *Drosophila* is a good model include mosquitoes, which transmit many viral pathogens (Dengue fever, all of the encephalitis agents, Yellow fever, and many others) and sand, tse tse or black flies, which transmit disease-causing protozoans and worms (Leishmaniasis, sleeping sickness, and river blindness (SCHNEIDER 2000). Further, although many genetic tools have been developed in non-model organism like mosquitoes (deletions and non-recombining chromosomes), these are still no match for the tools available in *Drosophila* (SCHNEIDER 2000). Establishing *Drosophila* as the host in this model system for the evolution of virulence likely will, as before (ROBERTS 2006), spur on and facilitate work in other organisms. Additionally, it is a tractable and widely used system that will allow current users, should they choose, to expand their research into this exciting field with minimal effort. This could allow the evolution of virulence to be studied by a larger number of people to produce more new data to inform theory on this subject.

Finding a good model system for host-shifts (one where the two hosts are closely related, but where the host-shift has not actually happened yet) is challenging; the *D. melanogaster*/*D. simulans*/DMelSV system provides such an opportunity *Drosophila simulans* can be reared easily in the lab under the same conditions as *D. melanogaster*, and has the same short generation time and worldwide distribution as its better known sibling. Additionally, genomic resources are now available for *D. simulans*, with 6+ genomes completed (www.dpgp.org), and tiling and expression microarrays are under development (McIntyre, Wayne, and Nuzhdin per. comm.), opening the door to host genomics as well as viral genomics in the future. Another benefit of the *Drosophila* host

is that *D. melanogaster* and *D. simulans* do not hybridize successfully despite their close phylogenetic relationship. Hybrids have only been possible in the lab and, even then, they are sterile (STURTEVANT 1920) which explains why we don't find hybrids in the wild. Finally, *D. melanogaster* is infected by a rhabdovirus (DMelSV). DMelSV is endemic to wild *D. melanogaster* populations and is not found in *D. simulans*, the sibling species to *D. melanogaster*, even though the two species are sympatric over most of their range (FLEURIET 1982; FLEURIET 1988). Therefore, this system is ideal because it allows us to artificially shift the virus at will.

Sigma Rhabdovirus

The sigma rhabdovirus is an excellent pathogen to be used in a system modeling the evolution of virulence because it infects *D. melanogaster*, thus affording us all of the benefits of that model organism. Sigma is a member of the Mononegavirales order. This virus order contains the causative agents of human diseases such as rabies, hemorrhagic septicemia, hematopoietic necrosis and several other economically important diseases of livestock (HOGENHOUT *et al.* 2003). There are many other RNA viruses of human interest with dipteran vectors, some of which already infect humans and others which do not yet infect us: La Crosse encephalitis, Rift Valley fever virus (Bunyaviridae), Chikungunya virus, O'nyong'nyong virus, Eastern, Western, and Venezuelan equine encephalitis, Ross River virus, Semliki Forest virus (Togaviridae), Yellow fever virus, and Japanese encephalitis (Flaviviridae) (FAUQUET CM 2005). We will use the *D. melanogaster*/*D. simulans*/Sigma system to model the evolution of virulence as the RNA virus host-shifts between dipteran hosts.

This virus has a negative single-stranded RNA genome and belongs to the family Rhabdoviridae. As with many members of this family, the *DMelSV* genome includes six

genes in the order 3'-N-P PP3-M-G- -L-5' (CARPENTER 2008; TENINGES *et al.* 1993). These genes code for the nucleocapsid protein (N), the polymerase-associated protein (P), the matrix protein (M), the glycoprotein (G), the polymerase (L) and a reverse transcriptase (PP3) (CARPENTER 2008; TENINGES *et al.* 1993). Interestingly, the PP3 gene is found in only a few other Rhabdovirus species and shares greater homology with retroviruses and retrotransposons, suggesting that *DMelSV* acquired this gene through recombination with one of these sources (CARPENTER 2008).

Further examination of the viral genome revealed that virus populations are genetically similar. Carpenter et al. (2007) reported very low genetic diversity for virus isolates from Europe and North America. The authors suggest that this is the case because either *Drosophila melanogaster* acquired the virus within the last 200 years or a single virus variant swept through the host populations (CARPENTER *et al.* 2007). The latter is likely as this is a vertically transmitted and the inheritance frequency is higher than would be predicted by strict Mendelian inheritance (VODOVAR *et al.* 2004). Following the sweep the virus has subsequently accumulated mutations that account for the minor differences among the widely distributed populations.

Some of these new mutations in the DMelSV genome likely arose by adenosine deaminases that act on RNA (ADAR). ADARs are enzymes that target dsRNA and change adenine (A) bases to Inosine (I). This change affects translation when I is read as guanine (G) and thus paired with anti-codons that possess a cytosine (C) instead of a thymine (T) at that particular position (MOERDYK-SCHAUWECKER *et al.* 2009). Additionally, the I bases can pair with C bases during RNA dependent RNA replication. Subsequently, the C bases pair with G bases resulting in the introduction of a mutation

(A to G) in the genomes of RNA viruses (MOERDYK-SCHAUWECKER *et al.* 2009). Carpenter et al (2009) sequenced ~6kb from viral genomes from wild flies and flies kept in the lab for 10-20 years. The number of singleton mutations was determined and an analysis of the distances between the singletons was done. In DMelSV, more candidate bases (A without upstream G) were mutated to Gs than would be expected by random change alone (CARPENTER *et al.* 2009). Additionally, these mutations occur clustered in specific regions of the genome (MOERDYK-SCHAUWECKER *et al.* 2009). Clustering of A-G mutations is another indication of ADAR activity. These data confirm that ADAR activity is occurring in the *DMelSV* genome and is, at least in part, responsible for the mutations accumulated after the recent sweep (CARPENTER *et al.* 2009; CARPENTER *et al.* 2007).

The sigma rhabdovirus is vertically transmitted transovarially from mother to offspring and produces CO_2 sensitivity (exposed flies are permanently paralyzed) in infected individuals (FLEURIET 1988; L'HERITIER 1958). Infections occurring early in the insect's development lead to infection of the germ cells and consequently a stable Sigma infection in *D. melanogaster* individuals of both sexes (BRUN and PLUS 1980). Whereas stabilized males produce non-stabilized offspring at modest frequencies, nearly 100% of offspring from stabilized females and over 80% of offspring from non-stabilized females are infected (FLEURIET 1988). Sigma is widespread in the fly's body, without any apparent tissue preference by the virus in both naturally and artificially infected flies (BRUN and PLUS 1980). Flies can be artificially infected by injection with infected hemolymph in saline to produce a stable infection and CO_2 sensitivity (L'HERITIER and HUGON DE SCOEUX 1947). The mean incubation time necessary for expression of CO_2 sensitivity is inversely related to the number of infectious units (I.U.)

delivered by injection and ranges from 5-20 days (1000 to <10 I.U.) (BRUN and PLUS 1980). The ease of artificial infection (by injection) coupled with stable vertical transmission post-injection and Sigma's natural virulence to *D. melanogaster* are additional traits that favor using this system for host-shift studies.

Finally, DMelSV is virulent in *D. melanogaster*. A vertically transmitted pathogen can only survive in the host population when its virulence is low and vertical transmission is high; typically horizontal transmission is also high (LIPSITCH *et al.* 1995). However, it is unlikely that Sigma is acquired horizontally. There is no evidence that the CO_2 sensitivity imparted to infected adult flies by Sigma is relevant in natural populations because adult flies likely never experience high concentrations of the gas in the field. Infected flies show decreased egg to adult survival (FLEURIET 1981a), reduced female fertility, and decreased overwintering ability in both sexes (FLEURIET 1981b). Moreover, the virulence of DMelSV was empirically demonstrated by an increase in infection frequency under relaxed selection pressure followed by a decrease in infection frequency after selection pressure was restored (YAMPOLSKY *et al.* 1999). These data follow the ecological theory which predicts that incidence of infection will decrease in a system where the pathogen is uniparentally transmitted and decreases the fitness of its host (FINE 1975). This being the case, natural selection against the infected host will cause the parasite's frequency to decline (FINE 1975). DMelSV fits these predictions because it is only effectively transmitted vertically by the female [stabilized males cannot sire stabilized offspring on un-stabilized females (FLEURIET 1988) and according to Yampolsky et al. (1999) DMelSV infection declines to zero when the flies are under strong selection pressure. Therefore, DMelSV is virulent in *D. melanogaster* and we

expect that DMelSV will be virulent in *D. simulans* because the two species are genetically and, perhaps more important, morphologically and physiologically very similar. This and the other characteristics mentioned above make the *D. melanogaster*/*D. simulans*/DMelSV system a good model in which to study host-shift evolution. The *D. melanogaster*/*D. simulans*/DMelSV model system will allow us to study several of the important traits of virus host-shifts which can facilitate the emergence of diseases, namely: how do virus fitness and virulence change between the native and novel host, and what role do virus and host evolution play in successful colonization of a novel host.

Significance

The genetics of the evolution of virulence remain an active area of investigation, particularly with respect to viral pathogens. Here we examine the evolution of virulence of the *Drosophila melanogaster* rhabdovirus virus (DMelSV) after it was artificially shifted onto a novel host, *Drosophila simulans*. A stable host-shift is at least a two stage process: 1) an initial stage in which the novel host receives the parasite, and 2) a subsequent stage during which the parasite adapts to the new host so that it can be maintained in the new host's population. This latter stage is relatively understudied (WOLFE *et al.* 2007), yet it is key to understanding the emergence of those most devastating diseases such as HIV that are passed efficiently from human to human (MAY *et al.* 2001). These diseases are able to persist on novel host because they adapt.

We define an adaptation as an individual nucleotide substitution that increases fitness. Adaptations are expected to be rare events for two reasons: first, because most mutations are unconditionally deleterious; and second, because most organisms are thought to be at equilibrium with their environment, *i.e.* already well adapted, thus

limiting the opportunity for further adaptation. Open questions about adaptation include: 1) the frequency with which adaptations arise; 2) whether adaptations tend to come from extant genetic variation in the population, or from new mutations; 3) the effect size of adaptive substitutions, and whether this decreases over the course of adaptation to a new environment; and 4) how frequently organisms tend to use the same solution to a given environmental challenge, *i.e.* convergent and/or parallel adaptation, and whether or not this "evolutionary repeatability" is dependent on the complexity of the genome (ORR 2005).

The genetics of adaptation are especially intriguing in host-parasite systems. Abundant empirical evidence demonstrates that most new mutations are deleterious, rather than adaptive. However, due to the compactness of the genome of RNA viruses like DMelSV (~12kb), their large population sizes, and their extraordinarily high mutation rate, it is possible that a relatively high fraction of mutations which begin as deleterious experience fitness reversals during their lifespan in the population due to compensatory mutation at other sites, such that they eventually fix as adaptations (COWPERTHWAITE *et al.* 2006). This casts into question whether most adaptations in RNA viruses are conferred by existing genetic variation, or by new mutations.

Much of what we know about adaptation in response to host-shifts comes from serial passage experiments (SPEs), where infection is caused by the experimenter (*i.e.* by injection), and typically no initial genetic variation is present. In such cases, clearly adaptation is conferred by new mutations. However, it is unclear whether this would be the case were genetic variation present. Accordingly, our system will evaluate the

relative contributions of standing variation *versus* new variation to adaptation by including multiple, naturally occurring genetic variants in the original inoculum of virus.

One of the major conclusions from SPE work is that virulence tends to increase with successive passages. This has important implications for emerging disease, and is often interpreted to mean that virulence will increase in a new host. This leads to two questions. First, will adaptation in a new host behave similarly to SPE, such that virulence actually will increase? And second, does adaptation to a new host actually require an increase in virulence?

One possibility is that due to co-infection by multiple genetic variants, competition for growth rate occurs in the new host, thus selecting for adaptations conferring rapid replication and thus rapid consumption of host resources. This theory implicitly assumes that growth rate and virulence are, if not interchangeable, tightly correlated. However, virulence may arise from host responses such as stress instead of, or in addition to, consumption of resources by replicating viruses. The pattern of change in virulence and its relationship to viral load has rarely been studied outside of the SPE context. Additionally, virulence with SPEs may increase because it is not costly: viral survival is guaranteed by the experimenter; therefore it is selectively irrelevant if viral precedence comes at such a cost to the host as to preclude transmissibility. Thus, increase in virulence may be an artifact of the SPE protocol (EBERT 1998).

A second major conclusion from SPE studies of the evolution of virulence is that increasing virulence in the new host is accompanied by attenuation to the original host. Attenuation is defined as both a decrease in virulence and a decreased ability to replicate and/or infect the ancestral host. Why should the new adaptations come at the

cost of the old ones? While these data have been interpreted in light of the generalist/specialist hypothesis, a more general answer to this question is loss of selective constraint: when the virus encountered the ancestral host repeatedly, fitness recovery on the ancestor was quick, more consistent with increases in frequency of alleles still present in the population than with back-mutation. Though budget constrains us from exploring this particular wrinkle in virulence evolution, we will assess the generality of attenuation by introducing the *D. simulans* evolved virus back into *D. melanogaster* after 16 fly generations (>> 16 viral generations).

Another question of great interest is whether there are multiple adaptive solutions to the same evolutionary problem. This has most famously been phrased as whether or not the outcome of evolution would be the same if the "tape of life" were played over again (GOULD 1990). Some recent work has suggested that even in complex eukaryotes, replicate "tapes" might have surprisingly similar outcomes (VERMEIJ 2006). Convergent and/or parallel evolution (evolution arriving at similar solutions to the same challenge) has been widely found in experimental evolution systems involving microbes (MOYA *et al.* 2004). Does this conclusion depend on whether adaptations arise from new mutations, or from standing variation? Bull et al. (1997) demonstrated a remarkable rate of convergence among RNA viruses in the context of host-pathogen coevolution due solely to new mutation (BULL *et al.* 1997). Similarly, convergence and parallelism were demonstrated in *E. coli* descended from a single genetic variant, but those experiments were in the context of adaptation to a non-evolving, chemostat environment (WOODS *et al.* 2006). However, while mutational identity has been demonstrated in RNA viruses (WICHMAN *et al.* 1999b), this result was not demonstrated

in *E. coli*; rather, different mutations occurred in the same genes at rates greater than in a set of randomly chosen comparison genes (WOODS *et al.* 2006).

The *D. melanogaster*/*D. simulans*/DMelSV system models two key features of emerging diseases: DMelSV is an RNA virus, thus likely to evolve rapidly, and the host-shift is between closely related species (HOLMES and RAMBAUT 2004). Additionally, controlled experiments in a model system will allow us to quantify the virulence of the virus in the native and the novel hosts before, during, and after multiple, identical host-shifts. Therefore, this system is an excellent model in which to test several hypotheses on the evolution of virulence: 1) standing genetic variation is the source material for the evolution of virulence; 2) the virus is more infective and more virulent on the novel host; 3) virus evolution on the novel host adversely affects re-infection and replication on the native host.

CHAPTER 2
LOW GENETIC VARIATION BOTH WITHIN AND BETWEEN HOSTS FOR SIGMA VIRUS (RHABDOVIRIDAE) IN *DROSOPHILA*

Background

The sigma virus is a member of the Mononegavirales, the virus order that contains the causative agents of diseases such as rabies, hemorrhagic septicemia, hematopoietic necrosis and other economically important diseases of humans and livestock (HOGENHOUT *et al.* 2003). Additionally, there are many other RNA viruses with dipteran vectors that are also human pathogens (FAUQUET CM 2005). DMelSV is vertically transmitted in the dipteran *Drosophila melanogaster*, and causes a characteristic CO_2 sensitivity such that when infected flies are exposed to the gas, they die or are paralyzed. We use the DMelSV-*D. melanogaster* model to examine the standing genetic variation that is present in wild virus populations.

Rhabdoviruses have negative single-stranded RNA genomes. DMelSV encodes six genes (3'-N-P-PP3-M-G-L-5') (CARPENTER 2008; TENINGES *et al.* 1993). These genes code for the nucleocapsid protein (N), the polymerase-associated protein (P), the matrix protein (M), the glycoprotein (G), the polymerase (L) and a reverse transcriptase (PP3) (CARPENTER 2008; TENINGES *et al.* 1993). Interestingly, the PP3 gene is found in only a few other rhabdovirus species, and shares greater homology with retroviruses and retrotransposons, suggesting that these viruses acquired this gene through recombination with one or both of these sources (CARPENTER 2008).

Viruses with RNA genomes are notorious for their high mutation rates (DOMINGO and HOLLAND 1997). However, it is well known that negative strand RNA viruses may have lower mutation rates than positive strand viruses (DOMINGO and HOLLAND 1997).

Although mutation rate has not been directly estimated in DMelSV, substitution rate has (4.6 x 105 substitutions/site/year, CARPENTER *et al.* 2007). For neutral sites, substitution rate is equal to mutation rate (KIMURA 1968). The substitution rate estimated for DMelSV is likely to be close to the neutral mutation rate, but if so, is lower than average even for a negative strand virus. Additional evidence that the mutation rate for DMelSV may be low comes from Brusini et al. 2012 (*American Naturalist,* in review), where DMelSV-infected *D. melanogaster* lines were selected for either high or low virus titers for 12 host generations. Although a significant response to selection was found, very few mutations were fixed in the virus; indeed, in some cases, no mutations were fixed between pairs of lines with large divergence in titers. The low mutation rate may be due to DMelSV's lifestyle as a vertically transmitted parasite: some vertically transmitted, apparently symbiotic viruses in plants also have extremely low mutation rates (ROOSSINCK 2010). However, some genetic variation does exist in the virus, on decadal time scales and across space; and in previous studies of DMelSV variation, a subset of the genomic variation in DMelSV was attributed to host-encoded adenosine deaminases acting on RNA (ADAR) (CARPENTER *et al.* 2009), rather than mutation during viral replication.

Adenosine deaminases action on RNA are enzymes that target dsRNA. The presence of dsRNA is often a sign of viral infection (for example, replication intermediates for single-stranded RNA viruses are double stranded), and it is thought that ADARs may have evolved as an antiviral response (KEEGAN *et al.* 2001). Specifically, ADARs change adenosine (A) bases to inosine (I). This change affects translation, because I then pairs with a cytidine (C) instead of a thymidine (T) at that

particular position on the anticodon (KEEGAN et al. 2001). Additionally, the I bases can pair with C bases during RNA-dependent RNA replication of the + strand genomic intermediate that is generated during rhabdovirus replication (ROSE and WHITT 2001). Subsequently, the C pairs with G, resulting in the introduction of a mutation (A to G) in the viral genomes (KEEGAN et al. 2001). Two enzymes (ADAR1 and ADAR2) are capable of this type of mutation (KIRCHER and KELSO 2010). ADAR1 catalysis the A-to-I deamination in double stranded RNA targets, while ADAR 2 catalyses the A-to-I mutations in pre-mRNA (GEORGE and SAMUEL 2011). The likelihood that a given A will be deaminated is conditioned on its 3' or 5' neighbors. Deamination is more likely to happen when the 5' neighbor is either an A or U (27 and 32% deamination, respectively, U ≈ A > C = G; Table 2-2-1 and Table 2-2) (MUELLER *et al.* 2006). This preference spans many host taxa [Rhabditid nematodes to humans (BASS 2002)], including *D. melanogaster* (LANDER and WATERMAN 1988).

In *Drosophila* there is a single ADAR gene, dADAR. dADAR deletion mutants exhibited locomotor issues and tremors that worsened with age, indicating that dADAR activity also affects nervous system function and stability, as well as potentially having antiviral activity (PETSCHEK *et al.* 1996). Although dADAR was not specifically implicated, ADAR activity also has been found in the DMelSV infecting *D. melanogaster (*CARPENTER et al. 2009*)*. Carpenter et al (2009) sequenced ~6kb from viral genomes from wild flies and flies kept in the lab for 10-20 years. In the *D. melanogaster* sigma virus, more candidate bases (A without upstream G) were mutated to Gs than would be expected by chance alone (CARPENTER *et al.* 2009). Additionally, these A-to-G mutations were clustered in specific regions of the genome, as is typical of ADAR

activity (CARPENTER *et al.* 2009; MOERDYK-SCHAUWECKER *et al.* 2009). These data confirm that ADAR activity is occurring in the DMelSV genome and is, at least in part, responsible for the mutations accumulated after the recent sweep (CARPENTER et al. 2009; CARPENTER et al. 2007). Host-mediated dADAR activity can lead to hypermutation in some rhabdoviruses (CARPENTER *et al.* 2009; OHARA *et al.* 1984). Hypermutation, in turn, can contribute to decreases in pathogen virulence (MEYERS *et al.* 2003). Thus, the upregulation of ADAR activity in DMelSV-infected *D. melanogaster* could be an antivirulence mechanism.

Here we use SOLiD™ deep-sequencing from multiple Drosophila individuals within a single host population to comprehensively evaluate the patterns of sigma sequence variation within a single population and within hosts, and to determine how much, if any, of the observed variation is attributable to ADAR. The SOLiD™ sequencing platform is ideal for this project because the short nature of the DMelSV genome will produce extremely deep coverage at every site. Therefore, once the standard quality controls are implemented, accurate, high quality, deep coverage of viruses within a single host can be obtained (HARISMENDY *et al.* 2009). Given high coverage, we can expect the probability of false negatives (failing to identify variants) to be vanishingly low, while the probability of false positives will only approach 0.1, making the probability of either error type extremely rare (BENAGLIO and RIVOLTA 2010; KUIKEN *et al.* 2006; LUSTIG *et al.* 2000; ONDOV *et al.* 2008). We will first identify mutated sites, then consider the surrounding sequence to identify the role of ADAR mutation within and between flies (Table 2-1 and Table 2-2) (LEHMANN and BASS 2000).

Materials and Methods

Fly Collection and Propagation

Flies were collected from a single location in Athens, GA in August 2007 using banana baits. Flies were allowed to oviposit on prepared Drosophila food for 24 hrs, after which DMelSV infection was determined by exposing the flies to CO_2 because flies infected with this virus become paralyzed upon exposure (BRUN and PLUS 1980; L'HERITIER and HUGON DE SCOEUX 1947). The offspring of each infected fly was kept under standard reading conditions (24°C and 16:8 light: dark). The result was six isofemale lines infected with DMelSV.

Ribonucleic Acid (RNA) Extraction

Ribonucleic acid was extracted from each of the six infected females using TRIzol (www.invitrogen.com) according to the standard manufacturer's protocol and stored at -80°C. Briefly, the fly was submerged in TRIzol in a 1.5ml Eppendorf tube and homogenized with a Kontes Teflon homogenizer. Chloroform was added to the homogenate and the tube was vortexed and incubated at room temperature for five min and then centrifuged at 12,000g and 2°C for 10 min. The upper aqueous phase that formed was removed and the RNA was precipitated by the addition of isopropanol followed by incubation at -20°C for 30 min. The precipitated RNA was centrifuged again with the conditions listed above. The resulting pellet was washing twice in cold 70% ethanol, dried briefly at room temperature and suspended in TE (10mM Tris, 1mM EDTA, pH 8.0). Purified RNA was quantified using a NanoDrop and the samples were stored at -80°C. For each fly, 500ng of RNA were used per reaction.

Reverse Transcription and Polymerase Chain Reaction

A portion of the viral genome (6370 bp; partial N, G, M, X, P and partial L genes) was amplified from each fly in overlapping pieces using the Superscript III one-step RT-PCR system with platinum Taq high fidelity and following the manufacturer's standard protocol (www.invitrogen.com). Briefly, for each primer pair, 500ng of fly RNA were mixed with the reaction mix, sense and anti-sense primers, the superscript enzyme mix, 1 µl of RNAse inhibitor (www.promega.com) and water. The reactions were incubated at 45°C for 30 min and then 94°C for 2 min. This was followed by 20 cycles of 94°C for 15 seconds, 55°C for 30 seconds and 68°C for 1 min. A final extension step for 10 min at 68°C was run. Each of the 4 segments, which ranged from 1073 to 2544 bases, was amplified twice independently for 20 cycles to reduce propagation of PCR error, while producing enough product to sequence (thus a total of 4 genomic regions x 2 replicates = 8 fragments per sample).

The resulting PCR products were electrophoresed, the bands were cut out of the agarose gel, and the DNA was purified from the resulting plugs using Qiagen QiaQuick Gel Extraction gel purification kits following manufacturer's protocols (www.qiagen.com). The resulting purified fragments were quantified after extraction using a NanoDrop and only products of 4ng/ µl or greater were used for sequencing. The eight purified fragments representing the virus from each fly were combined into individual samples to yield a minimum of 2 µg of DNA per sample.

Libraries were made from each sample, following the standard manufacturer's protocol for SOLiD™ sequencing. Briefly, the 2µg of DNA for each sample was sheared and size-selected. The sequencing adaptors were ligated and amplified. The dsDNA library was mixed with the bead emulsion according to the manufacturer's instructions.

The beads were deposited and 48bp reads were obtained. All 8 samples (six flies, with two flies repeated as internal controls) were bar-coded, pooled, and run on a single region of a SOLiD™ 5500xl (Applied Biosystems, Foster City, CA) plate. Two of the runs (two replicates of the same fly) failed; therefore, only 6 samples (representing five individual flies) were included in subsequent analyses.

SOLiD™ Reads Mapping and Alignment

Read from SOLiD™ sequencing were mapped to the reference genome of the sigma AP30 (GenBank accession number NC_013135) using default settings of MAQ 0.6.6 (http://maq.sourceforge.net/index.shtml). Once reads were mapped, the total number of reads per site for each sample and overall were determined. Out of 6570 sequenced sites, there were 137 sites that were covered by 100 reads or less in at least one of the samples, and thus, these sites were excluded from further consideration (of these, only 45 sites were excluded because of under-representation of a single sample), for a total of 6433 sites used in further analysis.

The parameters of the reads density per site distributions for each sample are given in Table 2-1. The median values ranged from 9391 (in BC1 sample) to 99574 (in BC7 sample). There were a total of 6084 and 349 sites located either in protein-coding (also referred to as coding) or non-coding regions, respectively.

Adenosine Deaminases Acting on RNA and Polymorphism Analysis

Sites with ADAR activity were determined using the consensus nucleotides in each individual sample, taking into account 5' adjacent nucleotides (Table 2-2). The sites have A-G SNPs were classified into "strong" and "weak" ADAR sites based on their 5' neighbors. Out of 6084 coding sites, there were 1779 ADAR sites, and 4305

non-ADAR sites. Among ADAR sites, there were 493 sites in the ‘weak” category, and 1286 in the “strong” category.

Results

Whole-Genome Sequencing

One of the unique samples and one of the replicate sequencing samples failed, leaving us with five unique sequence samples (BC1, BC2, BC3, BC5 and BC8). BC7 was the technical replicate of BC2. Although the absolute number of reads per site was significantly larger in BC7 than in BC2 (average of 137235 ± 1756 versus 65846 ± 962; paired *t*-test, $P < 0.001$), there was a strong correlation between the number of reads per site at individual sites in these two replicas ($R = 0.921$; Figure 2-1), supporting consistency in the results derived from these technical replicates.

Table 2-1 shows the summary statistics of the number of reads per site among all samples. The mean and the median numbers of reads per site varied from 22,283 (± 377) and 9391 in BC1 sample (which also had the smallest total number of reads) to 165,650 (± 2971) and 86496 in the BC8 sample (which had the largest number of total reads). Overall coverage (computed as per the Lander/Waterman equation, where coverage = L * (N / G), where G is the haploid genome length, L is the read length, N is the number of reads) varied from slightly over 1,000,000X in BC1 to over 7,700,000X in BC8 (Table 2-3, LANDER and WATERMAN 1988). However, their coverage ranged from over 2,200X to almost 18,000X among the samples when only the reads with non-consensus variants were considered (Table 2-1).

Our initial analysis examined the overall sequencing coverage across the genome. The nucleotide composition varied between protein-coding and non-coding regions, with the latter category having more A/T sites than the former. However,

because the majority of ADAR sites were located in the coding regions, we focused on the subset of 6084 coding sites and characterized different classes of observed mutation events (Figure 2-2). The results of this analysis are presented in Tables 2-4 through 2-10. Subsequently we characterized the number of read that fell within the coding regions and were either in non-A sites, weak or strong ADAR sites for every codon position (Tables 2-6, 2-7 and 2-8).

Adenosine Deaminases Acting on RNA (ADAR)-Driven Substitutions

Strong ADAR sites are an adenosine (A) with preferred 5' and 3' neighbors (Table 2-1, LEHMANN and BASS 2000): the combination of neighboring bases affects the propensity for ADAR activity at a particular A base (Table 2-2, LEHMANN and BASS 2000). There were 6433 total sites with sufficient read coverage to be included in our analysis (*i.e.*, that passed quality control tests), of which 6084 sites fell within the sequenced, protein-coding regions of the portion of the DMelSV genome. A further subset of 1779 sites within the coding region harbored the consensus A nucleotide and were thus identified as potential ADAR sites. We then considered the respective 5' nucleotide neighbor and assigned the A-harboring sites into strong and weak ADAR-preferences groups (see Table 2-4 and Figure 2-3 for details). There were 1286 and 493 sites with strong and weak ADAR preference, respectively. Notably, among the 1779 sites that harbored a consensus A nucleotide, the vast majority (1747 sites, 98.2%) of these sites harbored non-consensus sequence variants at relatively low frequencies (<5%, See Figure 2-4); while the remaining fraction (32 sites, 1.8%) of sites harbored minor variants with frequencies as high as 48.5% (in BC7 sample). Thus, we separated the sites into two sub-categories – those with high-frequency SNPs (single-nucleotide polymorphisms) present in at least 5% of all reads (total of 32 sites), and

sites without high-frequency SNPs (1747 sites). Table 2-4 depicts the summary statistics (means and medians) for the observed reads harboring the minority nucleotide at 1747 sites by individual samples. The direction of the mutation is assumed starting from the consensus nucleotide (A in all cases at these sites; for example, A-to-T, A-to-C, etc.) and the 5' ADAR preference is also taken into account. As evident from Table 2-4 and Figure 2-5, the number of A-to-G reads exceeds the number of A-to-T and A-to-C reads in all samples and in all ADAR site categories (Table 2-4). It should be noted that although the absolute number of reads per site (Figure 2-5) cannot be directly compared across samples (because of differences in overall coverage between samples), the trends in A-to-G, A-to-T and A-to-C changes are essentially the same among samples. Further, minor differences can be observed between samples, with BC1 and BC5 having similar number of A-to-G changes in both weak and strong site categories, while BC3 and both technical replicates (BC2 and BC7) show slightly more A-to-G changes at weak ADAR sites. The opposite trend is observed in BC8. Interestingly, slightly higher numbers of both A-to-T and A-to-C changes are observed in all samples at sites characterized as strong ADAR preference sites. The reason for this is not clear; it might be biologically relevant, or attributable to stochastic differences in the number of reads due to the 3-fold difference in the number of sites in each category (485 strong versus 1262 weak sites).

The SNP sites harboring non-A-to-G changes (in other words, those with A-to-T or A-to-C changes) were combined into a single mutation category in 2x2 contingency tables to test whether the sites with strong ADAR preference harbor more A-to-G mutations than the other two possible changes. A Fisher's exact test confirms that three

of the samples, BC8, BC2 and BC7, have significantly more mutations in ADAR-preferred than in non-ADAR-preferred site categories (P = 0.0019, P = 0.0473 and P = 0.0184, respectively; Table 2-5). In BC1, BC3 and BC5 the two preferred and non-preferred categories are statistically indistinguishable (P = 0.3583, P = 0.0575 and P = 0.1990, respectively). It is reassuring that the technical replicates, BC2 and BC7, have similar mutational trends, with smaller number of A-to-G mutations harbored at strong ADAR sites compared to weak sites. On the other hand, BC8 sample shows more A-to-G hypermutations at the strong ADAR sites than at the weak ADAR sites. Therefore, in these rare SNP sites there was between-host heterogeneity with regard to the mutational pattern.

This pattern of ADAR-driven mutations is much more prominent at a subset of 32 sites that harbor high-frequency SNPs (within-fly variants at 5% or more). For these sites, the strong preference ADAR category sites harbor significantly more A-to-G hypermutation changes than the weak category sites in all but one sample (Fisher's test: BC1, P = 0.1613, all other $P < 0.0005$; Table 2-7). As with the larger subset of sites, these differences cannot be solely attributed to the differences in the absolute number of reads per site, as these values do not differ significantly between weak and strong site categories (ANOVA, $P > 0.05$ for all samples). Additionally, when the locations of the SNP-harboring A-consensus sites were considered, we noted a cluster of 7 ADAR sites (out of total 15 SNP sites) in the glycoprotein (G) gene (Table 2-8), as might be expected if the ADAR mechanism were responsible for these mutational changes (CARPENTER *et al.* 2009; LEHMANN and BASS 2000).

The functional consequences of high-frequency SNPs across all 94 SNP-harboring sites are explored in Table 2-8. On average, half of all SNPs in each sample lead to non-synonymous (amino acid changing) substitution, with a similar fraction of nonsynonymous changes occurring at ADAR sites (Figure 2-6). The overall numbers of sites harboring high-frequency SNPs are quite similar between samples (ranging from 54 to 60 between all but BC8 samples). Interestingly, the majority of sites with non-synonymous substitutions (31 out of 42 sites) are shared between samples, with only 11 sites harboring different mutations in different samples, suggesting that either some sites are evolving under rather strong selective pressures leading to (convergent) amino acid changes, and/or flies may have been co-infected by multiple viral haplotypes that, in turn, had already been selected to reasonably high frequency in the viral population prior to infection.

Discussion

We set out to elucidate the role of ADAR on DMelSV genetics and role of the genetic diversity of DMelSV populations within individual wild-caught *D. melanogaster* females using SOLiD™ deep-sequencing. As expected, the SOLiD™ platform combined with the small nature of the viral genome resulted in extremely deep sequencing coverage of the roughly 6500 bp region of the genome we explored. Although we found significant genetic variation, this variation was not on the scale that might be expected from most RNA viruses (DOMINGO and HOLLAND 1997). We also found significant ADAR activity, as well as differences in ADAR intensity between hosts, which suggests that, at least in wild virus (within wild flies), ADARs play an important role in shaping the genome of the DMelSV rhabdovirus, potentially at least as important as replication-driven mutation.

Error is always one of the main concerns with any deep-sequencing project, particularly one such as ours, where the sequencing platform providing extremely high coverage is combined with a very short template (6.5kb). To help correct for possible error, we included an internal control where the same sample was sequenced twice during the same sequencing run. Although the numbers of reads were different for the two runs, the numbers of reads per site per individual were highly correlated. This indicates that, on average, every position along the target length was proportionally represented in our two technical replicates. Thus, the differences in read number are likely resulting from slight differences in sample processing after the original sample was halved that led to an effectively larger sample in the BC7 replicate. Additionally, the two technical replicates had similar mutation trends; with the smaller number of A-to-G mutations harbored at strong ADAR sites compared to weak ADAR sites. Therefore, the technical replicates were very similar despite the different initial read coverages.

Among sequenced sites, we identified 1779 A-harboring sites on the consensus sequence that were likely candidates for ADAR activity. Of these, 1747 (98.2%) were identified as sites harboring rare variants where A bases were mutated. Although many of these variants were A-to-G mutations, strong ADAR sites were not significantly preferred over weak sites in all samples, which suggest that (i) ADAR activity was not responsible for the production of some A-to-G mutations in these minor variants; and that (ii) biological heterogeneity in ADAR-driven activity exists among hosts.

Table 2-1. Descriptive statistics parameters characterizing the number of reads per site at 6433 sites for each virus sample.

Sample	Mean	SE Mean	Minimum	Median	Maximum	Total Reads	Minority-variants	Coverage
BC1	22,283	377	105	9,391	257,565	143,349,074	313,668	1,047,299
BC2	65,846	962	208	36,078	590,978	423,588,427	947,782	3,094,710
BC3	96,756	1,848	213	33,455	1,353,053	622,430,429	1,355,807	4,547,437
BC5	82,762	1,180	242	41,362	776,760	532,408,152	1,427,714	3,889,740
BC7	137,235	1,756	332	99,573	1,283,161	882,832,544	1,959,110	6,449,918
BC8	165,650	2,971	641	86,496	2,389,086	1,065,627,070	2,451,521	7,785,403

Shown are mean (with standard error, SE), median, minimum and maximum. BC2 and BC7 are technical replicates of the same fly. 137 sites which had at least one sample with the total number of reads below 100 were excluded (see Methods).

Table 2-2. Independent nearest neighbor preferences for adenosine deaminases acting on RNA (ADAR). Based on Lehmann and Bass (2000)*.

Site classification	5' neighbor	Total	% deaminated
ADAR-preferred	A	42	27
ADAR-preferred	U	32	32
ADAR-non-preferred	G	35	9
ADAR-non-preferred	C	36	13
	3' neighbor	Total	% deaminated
ADAR-non-preferred	A	42	11
ADAR-preferred	U	34	27
ADAR-preferred	G	37	29
ADAR-non-preferred	C	32	13

*Lehmann KA, Bass BL 2000. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39: 12875-12884

Table 2-3. Adenosine deaminases acting on RNA (ADAR) triplet preferences. Based on Lehmann and Bass (2000).

Triplet	% deaminated
U<u>A</u>U	47
A<u>A</u>G	44
U<u>A</u>G	42
A<u>A</u>U	30
C<u>A</u>G	23
A<u>A</u>C	23
U<u>A</u>C	21
U<u>A</u>A	17
C<u>A</u>U	17
G<u>A</u>U	14
G<u>A</u>G	13
A<u>A</u>A	11
C<u>A</u>A	9
G<u>A</u>A	6
G<u>A</u>C	5
C<u>A</u>C	4

*Lehmann KA, Bass BL 2000. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39: 12875-12884

Table 2-4. Mutations at 1747 sites harboring consensus A nucleotide within protein-coding regions, excluding the 32 sites that harbor high-frequency single nucleotide polymorphisms.

Sample	Mutation	5' ADAR preference	N	Mean	SE Mean	Median
BC1	A-G	Weak	485	27.2	3.1	11.0
		Strong	1,262	27.2	1.8	10.0
BC1	A-T	Weak	485	5.5	0.8	1.0
		Strong	1,262	9.2	0.6	3.0
BC1	A-C	Weak	485	5.6	0.5	2.0
		Strong	1,262	8.7	0.9	2.0
BC2	A-G	Weak	485	89.7	8.8	37.0
		Strong	1,262	86.0	5.7	31.0
BC2	A-T	Weak	485	17.7	1.9	5.0
		Strong	1,262	25.2	1.9	7.0
BC2	A-C	Weak	485	14.7	1.3	6.0
		Strong	1,262	28.3	2.9	7.0
BC3	A-G	Weak	485	133.0	15.2	41.0
		Strong	1,262	112.8	7.7	33.0
BC3	A-T	Weak	485	23.9	3.1	6.0
		Strong	1,262	32.3	2.4	9.0
BC3	A-C	Weak	485	27.5	2.5	9.0
		Strong	1,262	34.4	3.8	8.0
BC5	A-G	Weak	485	113.3	10.5	50.0
		Strong	1,262	113.3	7.6	46.0
BC5	A-T	Weak	485	32.2	3.6	11.0
		Strong	1,262	39.0	3.3	14.0
BC5	A-C	Weak	485	28.9	2.3	13.0
		Strong	1,262	43.2	4.0	13.0
BC7	A-G	Weak	485	190.6	17.4	100.0
		Strong	1,262	159.8	9.0	76.0
BC7	A-T	Weak	485	39.5	4.5	14.0
		Strong	1,262	51.9	3.8	17.0
BC7	A-C	Weak	485	37.6	3.4	15.0
		Strong	1,262	55.1	5.5	16.0

Table 2-4 Continued.

Sample	Mutation	5' ADAR preference	N	Mean	SE Mean	Median
BC8	A-G	Weak	485	185.9	19.7	85.0
		Strong	1,262	214.1	22.5	72.5
BC8	A-T	Weak	485	33.0	3.0	12.0
		Strong	1,262	73.7	6.4	18.0
BC8	A-C	Weak	485	36.7	3.4	13.0
		Strong	1,262	67.5	9.7	14.0
BC1	A-TC	Weak	485	11.1	1.1	1.0
		Strong	1,262	17.9	1.2	6.0
BC2	A-TC	Weak	485	32.4	2.8	13.0
		Strong	1,262	53.5	3.8	17.5
BC3	A-TC	Weak	485	51.4	5.1	17.0
		Strong	1,262	66.7	5.0	18.0
BC5	A-TC	Weak	485	61.0	5.1	26.0
		Strong	1,262	82.2	5.7	32.0
BC7	A-TC	Weak	485	77.1	6.9	30.0
		Strong	1,262	107.0	7.7	37.0
BC8	A-TC	Weak	485	69.8	5.5	28.0
		Strong	1,262	141.2	13.2	38.0

Single nucleotide polymorphisms (SNPs) mean numbers (with standard error, SE) of different categories of mutations (A to G, A to T, A to C and cumulative A to T or C) at (over 5% of alternative nucleotide). Sites are categorized as having either weak or strong adenosine deaminases acting on RNA (ADAR) preferences.

Table 2-5. Adenosine deaminases acting on ribonucleic acid (ADAR) preferentially changed strong ADAR sties.

	Mean # of A-to-G		Mean # of A-to-T-or A-to-C		
Line	Weak	Strong	Weak	Strong	Fisher's test (p)
BC1	27	27	11	18	NS (0.3583)
BC3	133	113	51	67	NS (0.0575)
BC5	113	113	61	82	NS (0.1990)
BC8	186	214	70	141	*P* = 0.0019
BC2	90	86	32	53	*P* = 0.0473
BC7	191	170	77	107	*P* = 0.0184

P-values from Fisher's exact test and mean numbers of different classes of mutations per site (ADAR-driven A to G hypermutations and non-ADAR-driven A to T or A to C mutations) at strong and weak ADAR sites. Based on 1747 sites harboring consensus A nucleotide within protein-coding regions [excluding 32 sites that harbor high-frequency SNPs (over 5% of alternative nucleotide)].

Table 2-6. Descriptive statistics parameters characterizing the number of reads per site at protein-coding sites in the 1st codon position for each sample, separated by whether the site is A-harboring [and further subdivided into weak or strong adenosine deaminases acting on RNA (ADAR preference] or not.

Sample	ADAR preference	N	Mean	SE Mean	Minimum	Median	Maximum
	Non-A	1,412	22,949	830	105	9,432	257,565
BC1	Weak	184	23,984	2,474	240	9,447	201,206
	Strong	432	22,645	1,350	146	9,818	191,636
	Non-A	1,412	100,571	4,091	213	33,881	1,353,053
BC3	Weak	184	102,739	12,067	594	34,172	1,039,318
	Strong	432	94,419	6,475	426	33,858	951,589
	Non-A	1,412	85,557	2,584	727	42,149	776,760
BC5	Weak	184	87,501	7,625	783	46,341	618,836
	Strong	432	84,788	4,278	242	45,353	575,819
	Non-A	1,412	171,050	6,615	949	89,231	2,389,086
BC8	Weak	184	183,497	19,560	2,084	91,067	1,931,032
	Strong	432	163,909	10,174	2,165	92,499	1,774,885
	Non-A	1,412	68,128	2,110	216	36,476	590,978
BC2	Weak	184	67,039	5,865	480	36,524	454,967
	Strong	432	67,565	3,628	233	37,213	423,946
	Non-A	1,412	142,330	3,898	332	102,740	1,283,161
BC7	Weak	184	141,191	11,098	1,589	98,422	953,685
	Strong	432	138,212	6,227	396	100,121	917,676

Shown are mean (with standard error, SE), median, minimum and maximum. BC2 and BC7 are technical replicates of the same fly.

Table 2-7. Descriptive statistics parameters characterizing the number of reads per site at protein-coding sites in the 2nd codon position for each sample, separated by whether the site is A-harboring [and further subdivided into weak or strong adenosine deaminases acting on RNA (ADAR) preference] or not.

Sample	ADAR preference	N	Mean	SE Mean	Minimum	Median	Maximum
BC1	Non-A	1,382	22,809	816	121	9,495	254,707
	Weak	233	25,557	2,255	246	9,705	191,446
	Strong	411	21,868	1,453	115	9,335	192,412
BC3	Non-A	1,382	99,675	4,053	372	33,976	1,335,779
	Weak	233	109,571	10,596	590	35,597	1,076,402
	Strong	411	91,977	7,032	242	32,324	927,653
BC5	Non-A	1,382	85,337	2,559	692	44,627	770,288
	Weak	233	92,355	6,812	1,706	41,464	600,593
	Strong	411	81,956	4,575	309	41,058	581,909
BC8	Non-A	1,382	168,626	6,504	979	89,840	2,360,591
	Weak	233	194,946	17,214	2,269	99,048	1,880,237
	Strong	411	162,423	11,358	1,095	87,804	1,733,237
BC2	Non-A	1,382	67,727	2,103	269	36,502	582,717
	Weak	233	74,719	5,523	559	38,341	482,167
	Strong	411	64,101	3,658	219	35,615	432,324
BC7	Non-A	1,382	140,730	3,854	540	100,526	1,262,557
	Weak	233	156,616	10,104	2,260	106,374	1,068,402
	Strong	411	132,285	6,613	376	98,319	939,369

Shown are mean (with standard error, SE), median, minimum and maximum. BC2 and BC7 are technical replicates of the same fly.

Table 2-8. Descriptive statistics parameters characterizing the number of reads per site at protein-coding sites in the 3rd codon position for each sample, separated by whether the site is A-harboring (and further subdivided into weak or strong ADAR preference) or not.

Sample	ADAR preference	N	Mean	SE Mean	Minimum	Median	Maximum
BC1	Non-A	1,511	22,582	774	114	9,419	237,843
	Weak	76	23,110	3,682	113	10,898	220,251
	Strong	443	24,021	1,537	275	9,823	252,351
BC3	Non-A	1,511	97,257	3,820	317	33,428	1,239,851
	Weak	76	104,221	18,736	435	39,166	1,143,098
	Strong	443	104,927	7,338	840	34,823	1,321,982
BC5	Non-A	1,511	84,472	2,445	623	43,215	727,102
	Weak	76	87,874	12,070	243	43,643	676,046
	Strong	443	87,974	4,608	2,451	42,835	765,518
BC8	Non-A	1,511	166,210	6,033	641	88,317	2,227,523
	Weak	76	176,578	32,233	953	103,082	2,099,507
	Strong	443	182,970	12,279	2,740	95,069	2,330,211
BC2	Non-A	1,511	66,691	1,999	208	36,062	541,349
	Weak	76	72,447	10,088	219	37,928	500,600
	Strong	443	70,373	3,701	748	38,070	575,834
BC7	Non-A	1,511	138,177	3,580	470	99,490	1,084,533
	Weak	76	149,555	18,917	427	105,172	1,059,282
	Strong	443	147,188	6,955	3,171	105,992	1,242,504

Shown are mean (with standard error, SE), median, minimum and maximum. BC2 and BC7 are technical replicates of the same fly.

Table 2-9. P-values from Fisher's exact test and mean numbers of different classes of mutations per site [adenosine deaminases acting on RNA (ADAR)-driven A to G hypermutations and non-ADAR-driven A to T or A to C mutations] at strong and weak ADAR sites.

Line	Mean # of A-to-G		Mean # of A-to-T-or-C		Fisher's test (p)
	Strong	Weak	Strong	Weak	
BC1	24	222	25	151	NS (0.1683)
BC3	51	781	155	332	$P < 0.0001$
BC5	157	836	108	335	$P = 0.0002$
BC8	159	1204	235	505	$P < 0.0001$
BC2	48	614	58	220	$P < 0.0001$
BC7	99	923	198	458	$P < 0.0001$

Based on reads at 32 sites that harbor high-frequency SNPs (over 5% of alternative nucleotide).

Table 2-10. Summary of non-synonymous substitutions (different from the consensus) that occur in each sample, organized by gene, codon number and base position within each codon (codon position).

Gene	Codon #	Codon position	Consens. base	Consens. codon	Consens. AA	BC1	BC2	BC3	BC5	BC7	BC8	ADAR
N	104	2	T	GTA	V	E	E	E	E	E	G	
N	114	1	T	TCG	S	P	P	P	P	P	A	
N	114	2	C	TCG	S	Stop	Stop	Stop	Stop	Stop	W	
N	407	2	G	CGC	R	H	H	H	H	H		
N	412	2	C	GCT	A		V					
P	689	2	T	ATA	I	T	T	T	T		R	
P	734	3	C	ACC	T						T	
P	825	1	T	TAG	STOP	E	E		E			
X	855	1	T	TTA	L						V	
X	901	1	A	AAT	N	D	D	D	D	D	D	N/D
X	903	1	G	GTC	V	L	L	L		L		
X	1027	2	T	ATG	M	K	K	K	K	K	R	
X	1120	2	C	ACC	T	N	N	N	N	N	S	
M	1262	1	T	TGG	W	R	R	R	R	R	G	
M	1262	2	G	TGG	W	L	L	L	L	L		
M	1262	3	G	TGG	W	C	C		C			
M	1263	1	C	CAT	H		Y	Y	Y	Y		
M	1263	2	A	CAT	H		R					H/R
M	1264	1	G	GCC	A		P		P			
M	1264	2	C	GCC	A						G	A/G
G	1414	3	C	GAC	D	E	E	E	E			
G	1578	2	A	GAG	E	G	G	G	G	G	G	E/G
G	1579	1	A	AGC	S	G	G		G		G	
G	1580	2	G	GGG	G	E	A	A	A	A		S/G
G	1588	2	A	CAT	H	P	P	P	P	P	R	H/R
G	1604	2	A	GAT	D	V	V	V	V	V	G	D/G
G	1679	1	A	AAA	K	E	E		E		E	K/E
G	1757	2	T	CTC	L	P	P	P	P	P		
G	1764	1	A	ATT	I	F	F	F	F	F	V	I/V

Table 2-10. Continued.

Gene	Codon #	Codon position	Consens. base	Consens. codon	Consens. AA	BC1	BC2	BC3	BC5	BC7	BC8	ADAR
G	1765	1	G	GTT	V		F	F	F	F		
G	1827	2	T	CTG	L	Q	Q	Q	Q	Q	R	
G	1835	2	T	TTC	F	S						
G	1883	2	C	TCC	S	F	F	F	F	F	C	
G	1901	1	T	TGA	STOP	G	G	G	G		G	
G	1901	3	A	TGA	STOP		C	C	C			
L	1970	1	C	CAT	H	Y		Y	Y	Y	D	
L	2005	3	G	TGG	W	C	C	C	C	C		
L	2009	3	T	TTT	F						L	
L	2058	3	C	ATC	I						M	
L	2059	1	A	ACC	T	P	P	P	P	P	A	T/A
L	2059	2	C	ACC	T	S	S	S	S	S	S	
L	2060	1	T	TTG	L						V	
					Total Mutations	60	60	54	59	56	45	
					Non-Synonymous	32	34	28	32	24	26	
					Total ADAR	7	8	6	7	5	12	
					Non-syn ADAR	3	4	2	3	2	9	

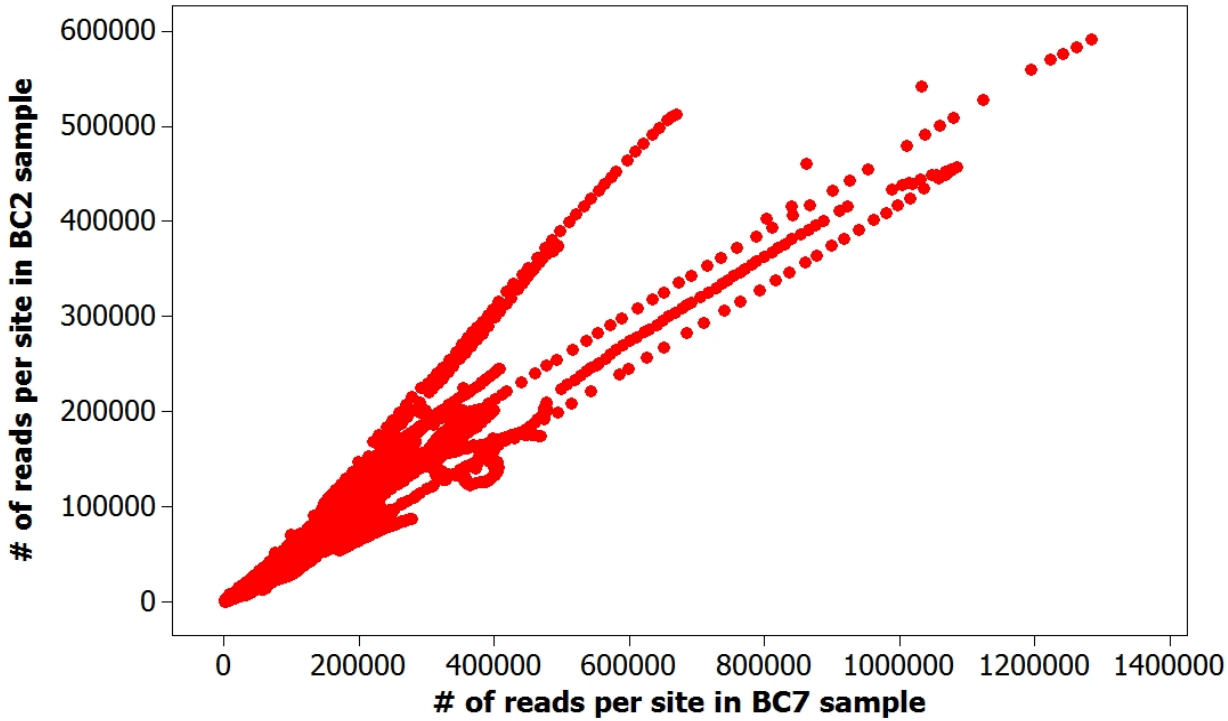

Figure 2-1. Correlation of read number for our technical replicate indicates that the two samples behaved similarly in the sequencing process ($P < 0.001$ and $R = 0.921$).

	A	T	G	C
A		β	α	β
T	β		β	α
G	α	β		β
C	β	α	β	

Figure 2-2. Shown are two possible transitions (α) and transversions (β) mutation events (that together form a substitution matrix).

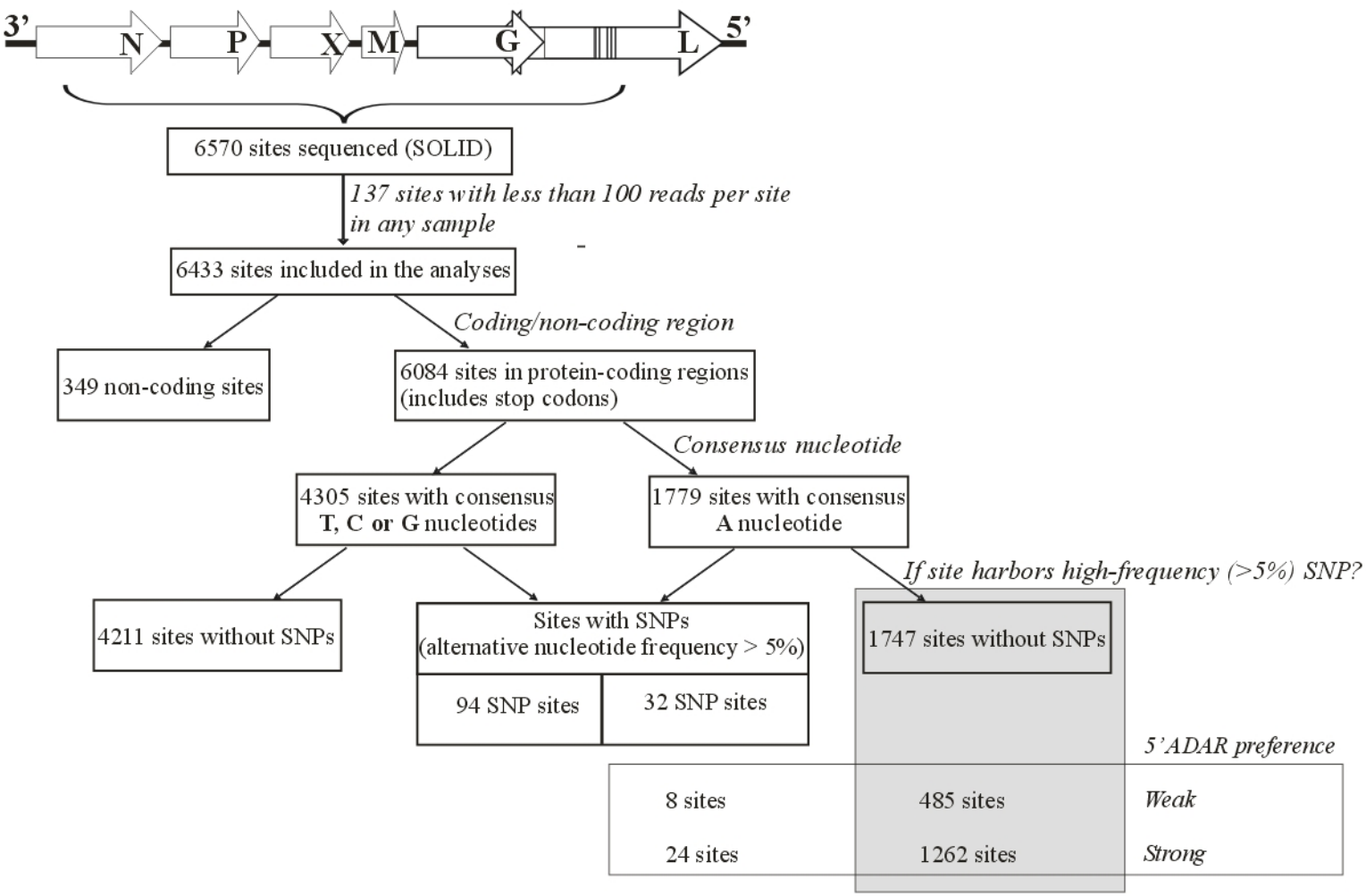

Figure 2-3. Flow chart describing how the 1747 candidate sites for adenosine deaminases acting on RNA (ADAR) activity were identified.

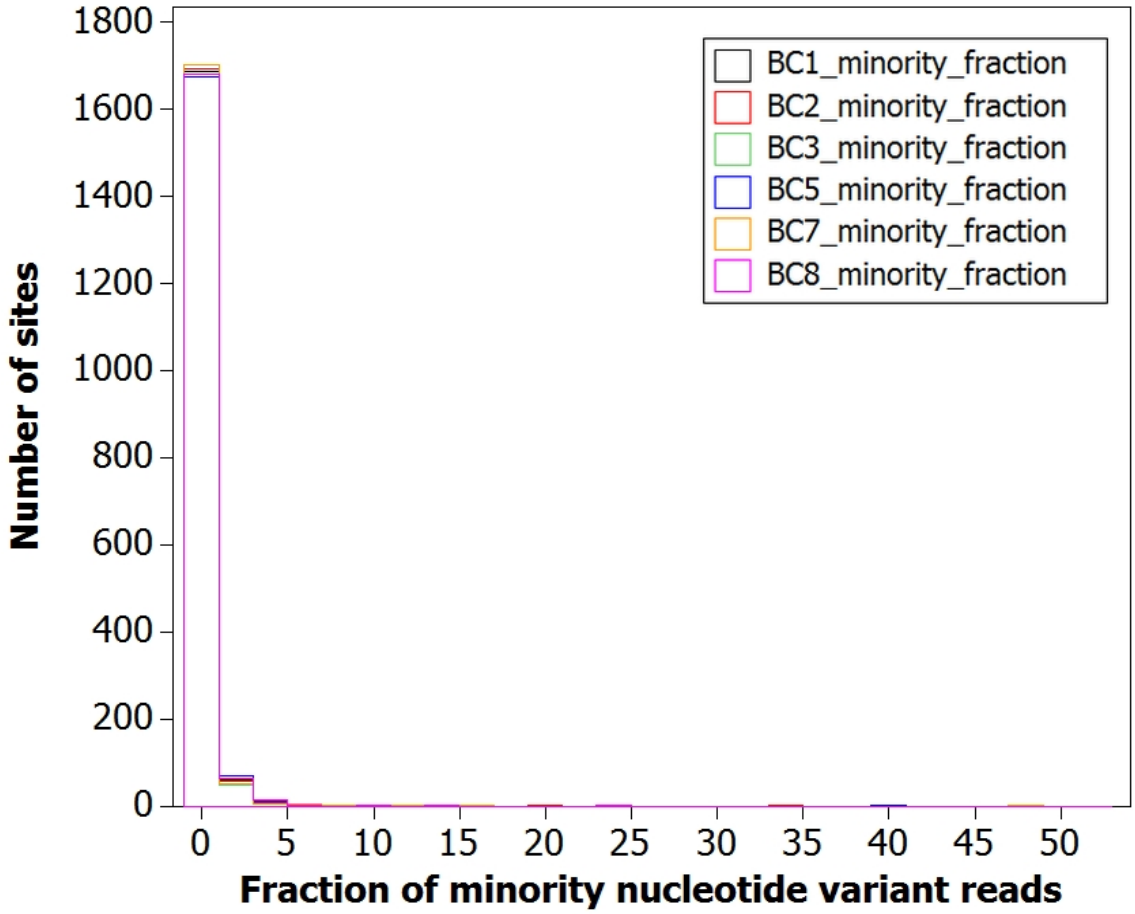

Figure 2-4. Minority reads were a minor component of our total read count. The number of minority nucleotide reads accounted for < 5% of the reads on any one base position along the genome.

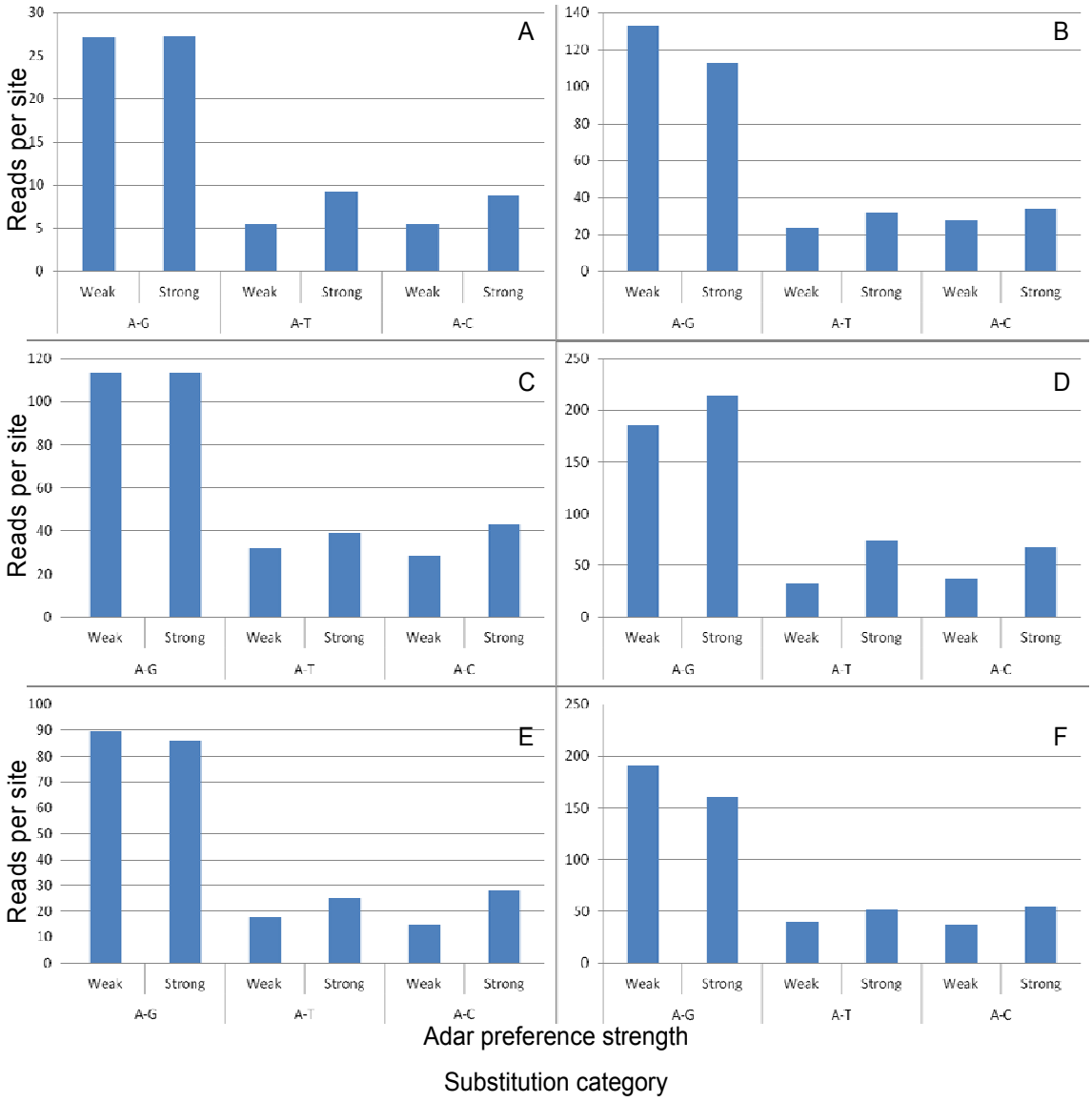

Figure 2-5. A-G site changes occurred disproportionally more often in all flies indicating adenosine deaminases acting on RNA (ADAR) activity. Sample-specific histograms summarizing mean numbers of minority-harboring reads for each mutation type (e.g. A-to-G) computed per site at 1747 A-consensus protein-coding sites. A = BC1, B = BC3, C = BC5, D = BC8, E = BC2 and F = BC7.

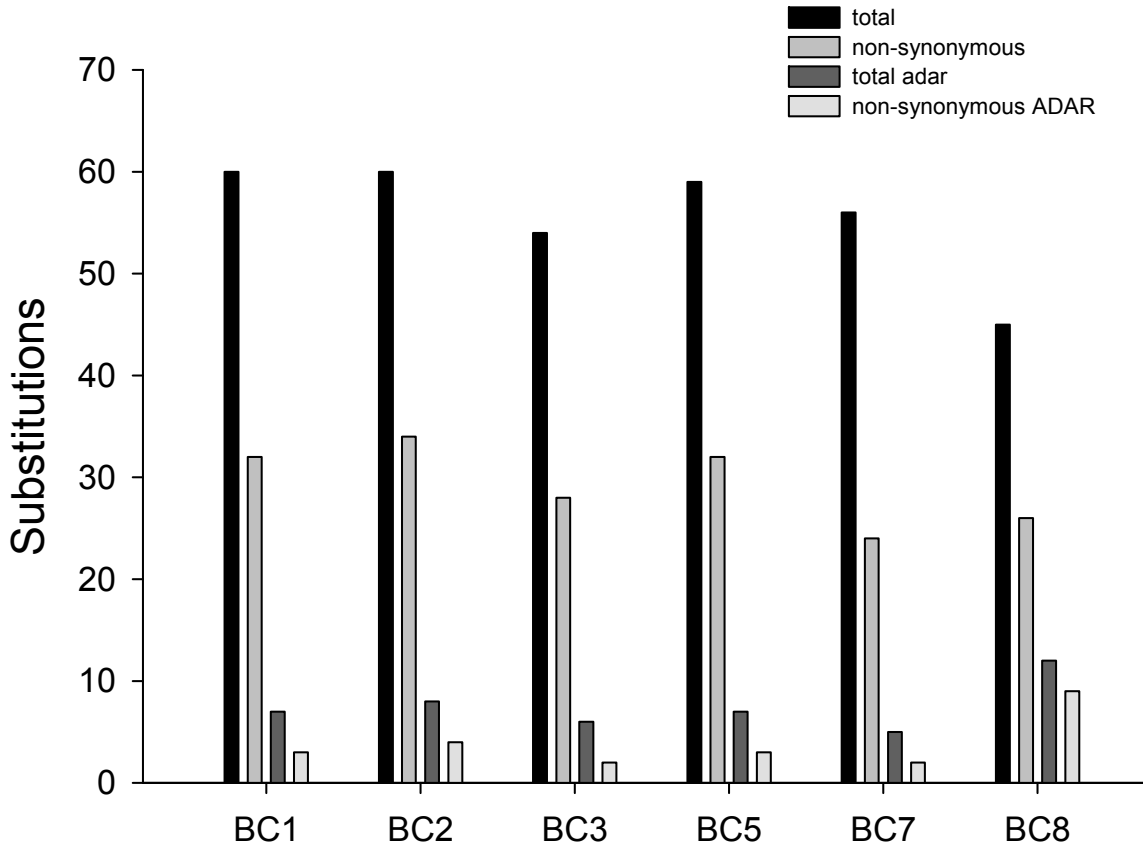

Figure 2-6. Total number of substitutions (that occur with frequencies of at least 5% in one of the samples) and type of substitution in each of the flies tested.

CHAPTER 3
EVOLUTION OF VIRULENCE IN THE SIGMA VIRUS FOLLOWING A HOST SHIFT IN DROSOPHILA

Background

Animal parasites, particularly viruses, regularly shift from their native wild animal hosts to pets, livestock or other wild animals that associate with humans, thus creating the potential for zoonoses and the evolution of virulence (BARBOZA *et al.* 2008; BULL 1994; MACKENZIE *et al.* 2001). However, it is expected that most instances of zoonosis go unnoticed for several reasons: virulence may be low, occurrence of disease may be rare, and most shifted parasites fail to establish stable populations in a human host. Even those parasites that do establish in one individual may not have the capacity for human-to-human transmission (COMBES 2005). Viruses that achieve human-to-human transmission can cause significant epidemics (PIALOUX *et al.* 2007) or even pandemics: for example, human immunodeficiency virus (HIV) (DENNEHY 2009; SALVAUDON *et al.* 2007) and, more recently, H1N1 influenza (FLYNN 2010; FRASER *et al.* 2009). It seems likely that viruses shift into humans on a regular basis; we cannot predict how each will affect the human population. Predicting the potential impact that viruses have on human populations requires a better understanding of the evolution of virulence following host shifts.

Much of our understanding of the evolution of parasite virulence has been developed using serial passage experiments (SPE), in which pathogens are passaged through single-celled hosts for many generations (EBERT 1998; LEVIN *et al.* 1999). Four generalizations have emerged from SPE studies concerning the evolution of parasite virulence after a host shift: 1) virulence

increases on the novel host, 2) attenuation in the former host occurs concomitantly with the increase in virulence on the novel host, 3) the same substitutions that cause increased virulence in the new host are responsible for attenuation in the original host, and 4) the changes generally occur most rapidly when the parasite in question is an RNA virus (EBERT 1998). However, the degree to which these results can be extended to multicellular hosts is unclear. In the current study we focus on the first generalization: does virulence increase in the novel host following a host shift?

Parasite virulence affects parasite fitness. Maximal parasite fitness is mediated by an intricate tradeoff between virulence and transmission (ANDERSON and MAY 1979; ANDERSON and MAY 1982). High virulence leads to high host mortality and thus decreased transmission. Therefore, in theory every parasite can reach its optimal fitness by simultaneously exhibiting optimal transmission and virulence levels (ALIZON 2008; ALIZON *et al.* 2009). However, some argue that the issue is more complex and requires the inclusion of host recovery (pathogen clearance), drift, genetic bottlenecks as well as group and kin selection (ALIZON 2008; ANDERSON and MAY 1982; FRANK 1992; FRANK 1996). Some argue that the tradeoff hypothesis alone is weak at best in explaining the evolution of virulence (EBERT and BULL 2003). Others point out that it is hard to understand exactly how the trade-offs work for individual parameters affecting the evolution of virulence (ANDERSON and MAY 1982; FRANK 1996); which suggests that more data on tradeoff curves are needed to help inform each of the parameters in the model (ALIZON *et al.* 2009; BOLKER *et al.* 2010). We expect that

in a naturally evolving system, high virulence will typically come at the cost of low viral fitness and reduced transmission (ANDERSON and MAY 1982). In SPE studies, however, virulence may increase because viral fitness is guaranteed by the experimenter and, consequently, virulence does not incur a cost on the parasite within the SPE system (EBERT 1998).

Here we use a dipteran model system to explore the evolution of virulence following an interspecific shift between multicellular hosts. Dipteran vectors like mosquitoes, sand flies, and blackflies carry RNA viruses that are the causal agents of diseases that cause extensive human suffering (HOGENHOUT *et al.* 2003). We use *Drosophila melanogaster* (native host) and its sibling species, *D. simulans* (novel host), which are separated by more than one million years of evolution. *Drosophila melanogaster* and *D. simulans* are reproductively isolated; crosses of *D. melanogaster* to *D. simulans* produce large numbers of sterile female offspring, while the reciprocal cross is much more difficult to accomplish (STURTEVANT 1920).

The parasite used in the model is the rhabdovirus sigma (DMelSV), which is endemic to wild *D. melanogaster* populations and is not found in *D. simulans*, even though the two species are sympatric over most of their range (FLEURIET 1982; FLEURIET 1988). DMelSV is a negative single-stranded RNA virus (Mononegavirales) from the Rhabdoviridae, a virus family that includes the causative agents of rabies, hemorrhagic septicemia, hematopoietic necrosis and several other diseases of humans and livestock (FAUQUET CM 2005; HOGENHOUT *et al.* 2003). DMelSV imparts a characteristic CO_2 sensitivity to infected

individuals that makes them easily identifiable (FLEURIET 1988; L'HERITIER 1958). The virus is vertically transmitted by both parents; however, male transmission is both incomplete and ineffective over multiple generations (FLEURIET 1996; WAYNE *et al.* 2011). Conversely, infected females pass the infection to most or all of their offspring; the latter produce 100% infected offspring and are termed stably infected (BRUN and PLUS 1980). Pizzatto et al. (2010) demonstrated that it is easier to create a persistent infection following a host shift with DMelSV in more closely related taxa than in more distant ones, similar to trends seen in vertebrate systems (BASS *et al.* 1997; KIMURA 1968; NISHIKURA 2010). Collectively, the traits of the coevolving virus and dipteran host make this an excellent system in which to test the generality that virulence increases after a host shift.

We used this system to test several questions pertaining to host shifts. First, is infecting a novel host more difficult than infecting a naïve native host? Second, is virulence be higher on the on the novel host than on the native host, as was the case in SPEs? Third, does virulence correlate with viral titer, or is virulence independent of titer and instead caused some other mechanism? To address these questions we artificially infected naïve, highly inbred lines of *D. melanogaster* and *D. simulans* with the same viral inoculum. Having successfully infected the flies, we selected fly lines with complete transmission (100% infected offspring) over three generations and determined infection prevalence (a proxy for transmission efficiency) through 16 generations. We determined host fitness (a proxy for virulence) and virus titer using strand-specific quantitative PCR (KOMURIAN-PRADEL *et al.* 2004) at regular intervals.

Materials and Methods

Fly Lines

We infected one effectively isogenic line of *D. simulans* (*MD106;* www.dpgp.org) and one highly inbred line of *D. melanogaster* [*Diallel 75*, (YANG and NUZHDIN 2003)] by injection with a common, pooled viral inoculum (see next section for description of inoculum). *MD106* is one of the effectively isogenic lines included in the *Drosophila* Population Genomics Project (www.dpgp.org), and has a complete light shotgun sequence available. *Diallel 75* is a highly inbred line (40 generations of full sib inbreeding (YANG and NUZHDIN 2003)).

Inoculum and Injections

We infected one effectively isogenic line of *D. simulans* (*MD106;* www.dpgp.org). The inoculum was prepared from 1000 DMelSV-infected flies (600♀ + 400♂) collected from twelve infected isofemale lines generated from infected females collected in Athens, GA (June 2008). The inoculum was prepared following Clark et al. (1979). Briefly, the flies were homogenized in buffer (HB: 0.005 M Tris-HCl, 0.25M sucrose, pH 7.5). The homogenate was centrifuged at 1200g for 15 min at 2ºC. The resulting supernatant was centrifuged at 6000g for ten min at 2ºC. The resulting supernatant then was filtered through a 0.45µm filter and centrifuged at 19,500g for one hour. The resulting pellet was suspended in 2ml of HB, aliquoted and stored at -80ºC (CLARK *et al.* 1979).

The flies (1-2 days old females) were injected using a Sutter instruments XenoWorks™ Analog Microinjector equipped with a 100 µl syringe and Narishige GD-1 needle blanks that were pulled with a Kopf Model 720 vertical needle

puller. Prior to injection, the microinjector tubing, needle holder, and the needle were back-filled with physiological grade mineral oil. The inoculum was thawed on ice, agitated, and the needle back-filled with inoculum. The flies to be injected were anesthetized on ice and kept on ice until they were injected. The needle was inserted about the second abdominal tergite and 0.1 µl of inoculum was dispensed. In total, 550 *D. melanogaster* and 550 *D. simulans* were injected. Additionally, 30 flies from each species were "mock-injected" with HB to control for any effects of the injection procedures. These mock-injected flies produced our uninfected control lines.

Rearing Conditions, Artificial Selection for Infection, and Relaxed Selection

All flies were reared at 24°C with a 16:8 light:dark photoperiod on Jazz-Mix *Drosophila* food following the manufacturer's instructions and using 7 ml of food per fly vial. New vials were set up every 21 days with constant density and sex ratio (5♀ + 5♂); females were allowed to oviposit for seven days. This schedule was followed for both species and was necessary because under our rearing conditions, *D. simulans* did poorly when kept on a 14 day schedule.

Artificial selection was conducted to produce lines in which 100% of the offspring died (complete transmission). After injection, each fly (Generation 0, G0) was deposited on a prepared vial (food, KimWipe, yeast, cotton) along with a male and allowed to recuperate and oviposit under standard rearing conditions (480 *D. melanogaster* and 485 *D. simulans*). The offspring (G1) of the injected flies (G0) emerged and were transferred to individual vials of fresh food 20 days after the injection date. These G1 flies were flipped onto new vials again 10 days later to prevent them from mixing with their offspring. When the G1 flies were 25

days old (±5 days), they were split haphazardly into two groups. One group was exposed to carbon dioxide for 5 min at room temperature and scored for CO_2 sensitivity. For any vial that contained some gas-sensitive flies, their un-gassed siblings were then set up in fresh vials as 1♀ + 1♂ under the assumption that some of these flies were carrying the virus. Vials that had no gas-sensitive flies were discarded. This process of flipping, gassing, and 1♀ + 1♂ setup continued for three generations, or until 100% infection was achieved, whichever came first (Figure 1, 2). In total, 2,487 fly vials were gassed during this selection period. After three generations of selection, all 100% infection lines (13 lines for each species, see results section below) were maintained without selection at constant density (5♀ + 5♂) on a 21 day schedule under standard rearing conditions as described above.

Fecundity and Hatchability

Egg production and per cent hatchability were determined by allowing flies to oviposit on green tinted food (for ease of egg detection; tinted with McCormick & Co., Inc, Hunt Valley, MD) on glass slides over a 48 hour period. Approximately 60% of each standard glass slide was covered with 1ml of the colored diet, allowed to set, and stored at 4°C. The slide were sprayed with a 4mg/ml yeast solution and provided to the flies for oviposition. After a 12hr oviposition period, the slides were removed from the vial and replaced with a fresh slide. The eggs were counted and the slide was incubated at room temperature for >24hrs. Subsequently, the eggs that failed to hatch on these same slides were counted to determine per cent hatchability.

Ribonucleic Acid (RNA) Purification

Ribonucleic acid was isolated from individual flies using TRIzol (www.invitrogen.com) and following the manufacturer's directions. Briefly, each fly was homogenized in 200 µl of TRIzol on ice using a Kontes Teflon homogenizer. Each sample was incubated at room temperature (RT) for 5 min. Each sample then received 40 µl of chloroform and was vortexed for five seconds, incubated at RT for 2-3 min, then centrifuged at 17,000g for 10 min at 2°C. The upper aqueous phase was collected and the RNA was precipitated by the addition of 100 µl of isopropanol followed by vigorous shaking and incubation at -20°C for 10 min. The precipitate was pelleted by centrifugation at 17,000g for 10 min at 2°C. The pellet was washed twice with 1ml of cold (4°C) ethanol, air dried for five min at RT, resuspended in 29 µl of DEPC-treated water with 1 µl of 40 µM RNAse inhibitor (www.neb.com) and incubated at 55°C for 10 min to aid resuspension. Samples were stored at -80°C until used.

Reverse Transcription

Ribonucleic acid was transcribed using the Promega AMV RT enzyme (www.promega.com) following the manufacturer's directions, except that the protocol was scaled down to use 50ng of template per reaction. Briefly, 50ng of template RNA and 12.5ng of a tagged reverse primer (GCAGTATCGTGAGTTCGAGTGT*CCGATGACCTGTCCGTAACT*; tag underlined and gene-specific portion in italics) and water (as needed to bring total volume to 11 µl) were incubated at 65°C for 5 min and then plunged into an ice bath. Subsequently, 5 µl of AMV 5X buffer, 1 µl of RNAse inhibitor, 2.5 µl of sodium pyrophosphate, 2.5 µl of 10mM dNTP mix (www.promega.com), 0.1 µl

of AMV RT enzyme and water were added to the tube to attain a final volume of 25 µl. The sample was agitated, centrifuged briefly, incubated for one hour at 42°C and stored at -20°C.

Quantitative Polymerase Chain Reaction

Strand-specific qPCR was done to insure that only negative strand RNA genomes were quantified (PURCELL *et al.* 2006). For each RT-product, a 50 µl reaction was made with 25 µl of TaqMan® Gene expression master mix (www.appliedbiosystems.com, Cat. #4369016), 0.9 µl of 50µM tag primer (GCAGTATCGTGAGTTCGAGTGT), 0.9 µl of 50 µM gene specific reverse primer (GAGTCGCAGCTTTGGAGTTC), 1.25 µl of a 10 µM hybridization probe (CATGAGATGGAGGAACTTTCTCTCCCA), 3.3 µl of RT product and 18.65 µl of water. The reactions were vortexed and centrifuged and 15 µl were plated in triplicate. A standard curve was made using a PCR-amplified and purified fragment of the N-gene that was diluted from 10^7 to 10^1 copies/ µl. The reactions for the standard curve were identical to the sample reactions. The samples were amplified and quantified on a StepOne Plus thermal cycler (www.appliedbiosystems.com) with the following conditions: 50°C for 2 min, 95°C for 10 min, then 40 cycles of 95°C for 15 seconds and 60°C for 1 min.

Statistical Analyses

Differences in the proportions of injected flies from each species that became infected, and the proportions of infected flies that transmitted the virus to 100% of their offspring (complete transmission), each were evaluated using *G* tests.

The data for transmission efficiency under three generations of selection were not normal even when the data for each line were averaged (Shapiro-Wilk W test, $P < 0.0001$). As such, the data were analyzed with a Wilcoxon test followed by a non-parametric comparison for all pairs using the Wilcoxon each pair test. The data for transmission efficiency under relaxed selection also were not normal (Shapiro-Wilk W test, $P < 0.0001$, generations 6-16) and were analyzed as above.

Data for titer were also not normal (Shapiro-Wilk W test, $P < 0.0001$) and failed Levene's homogeneity of variance test (P =0.0492); as such the data were analyzed with a Welch's ANOVA. Additionally, non-parametric Kendall's *Tau* correlation tests were used to compare infection rates and titer, both to avoid assumptions of normality and to reduce the impact of outliers.

The fecundity and hatchability data for the infected flies were standardized to the mean of their uninfected counterparts to control for environmental differences between generations. Both standardized data sets were not normal (Shapiro-Wilk W test, $P < 0.0001$). The average relative fecundity and hatchability were calculated for each fly line at each generation to eliminate non-independence effects and these data (also not normal, Shapiro-Wilk W test, $P < 0.0001$) were analyzed as above. All statistical analyses were conducted using JMP 10 (www.jmp.com).

Results

Artificial Host-Shift Success and Selection for Transmission

We injected similar numbers of *D. melanogaster* ($N = 480$) and *D. simulans* ($N = 485$) females. Injected *D. simulans* females produced infected

(CO_2-sensitive) offspring (*N* = 78, 16.1%) at a significantly higher rate than did *D. melanogaster* females (*N* = 45, 9.3%; *P* = 0.0121; Figure 3-1). Although a larger proportion of the infected *D. melanogaster* lines (28.9%, or 13 lines) than infected *D. simulans* lines (16.7%, or 13 lines) achieved the maximum transmission efficiency (100% infected offspring), these differences were not significant (*P* = 0.335, Figure 3-1). Therefore, the two species responded similarly to artificial selection for maximum transmission efficiency. However, there was variation within species for transmission efficiency, as in both species, some lines reached maximum transmission while others did not.

Overall transmission efficiency was higher in *D. melanogaster* than in *D. simulans* lines over the first three generations of artificial selection (*P* < 0.0001, Figure 3-2). The two species responded differently to selection across the three generations. While *D. melanogaster* lines exhibited significant increases in infection levels at each generation (*P* < 0.0001), statistically significant increases in infection levels were observed only from G2 to G3 in *D. simulans* lines (*P* < 0.0001, Figure 3-2

Infection under Relaxed Selection

The 26 lines (13 of each species) that attained maximum transmission within three generations of selection continued to evolve under relaxed selection for 13 additional generations (for a total of sixteen generations post injection). From G4-G16, transmission efficiency was significantly higher in *D. melanogaster* than in *D. simulans* (*P* = 0.0003, Figure 3-3). As was the case during selection, there was genetic variation within species: some lines from each species remained at or near 100% transmission efficiency, while others fell

to near 0% transmission efficiency. Both species experienced decreases in transmission efficiency; however, a significant decline (levels significantly lower than G3) occurred in *D. simulans* at G9 (P = 0.0105) while in *D. melanogaster* this significant decline occurred at G15 (P = 0.0167). Thus, transmission efficiency is erratic and decreases over time in both hosts, but decreases earlier and is overall lower in the novel host than in the native host.

Virus Titers

Ribonucleic was purified from each of three female flies per infected line from generations six and sixteen (i.e., G6 and G16), and viral titer was estimated by strand-specific quantitative PCR as described in the methods. The variance in viral titer was significantly greater in *D. simulans*, the novel host, than in *D. melanogaster* (Levene's test, P = 0.009), and the mean titer was also smaller (Figure 3.4). However, taking into account the unequal variances, the virus titers in the two species were statistically indistinguishable (Welch's ANOVA, P = 0.3820, Figure 3-4). Additionally, there were no differences in copy number for either species between generations (P = 0.7391). However, interesting trends are apparent from scrutinizing the data (Figures 3-3, 3-4, 3-5): there are more lines with low titer in *D. simulans* than in *D. melanogaster* at generation 6 and the titers in a number of *D. melanogaster* lines were markedly lower at G16 than at G6.

Virus titer and transmission efficiency were significantly correlated for both species at G6, though the correlation was negative for *D. melanogaster* (τ = -0.5394, P = 0.0289), while it was positive for *D. simulans* (τ = 0.4604, P = 0.0455; Figure 3-6). A significant positive correlation was observed between titer and

transmission efficiency for *D. melanogaster* at G16 (τ = 0.5002, *P* = 0.0265), but not for *D. simulans* at G16 (, τ = 0.3750, *P* = 0.1042; Figure 3-6).

Host Fitness

Host fitness (fecundity and hatchability) was examined at G6, G11 and G16 (Table 3-3 and Table 3-4). The data were standardized to the uninfected means to correct for differences in environmental heterogeneity between generations (hereafter referred to as relative fecundity and relative hatchability). The data were not normal, as such, the line means across vials were used in the statistical analysis to eliminate non-independence. The data were analyzed as described above using non-parametric tests.

Relative fecundity was statistically indistinguishable for infected and uninfected *D. simulans* at every generation tested (Compared to uninfected control mean of 1.0; *P* ≥ 0.2563). Similarly, relative fecundity was statistically indistinguishable for infected and uninfected *D. melanogaster* at G6 and G11 (*P* ≥ 0.3703). However, relative fecundity was significantly higher for infected G16 *D. melanogaster* than for the respective uninfected controls (*P* = 0.0265). Additionally, relative fecundity at G16 was significantly higher in infected *D. melanogaster* than in infected *D. simulans* at G16 (*P* = 0.0009). Although the relative fecundity of infected *D. simulans* females was statistically indistinguishable across the three generations tested (*P* ≥ 0.0513), it actually declined across the three generations (Figure 3-7). Conversely, the fecundity of infected *D. melanogaster* females was statistically indistinguishable between G6 and G11 (*P* = 0.1086) but rebounded at G16 (*P* ≤ 0.0025). This suggests that the native host (*D. melanogaster*), or virus in the negative host, evolved to reduce

virulence by G16 while the novel host (*D. simulans*) and/or its virus did not. Therefore, the virus is more virulent on the novel host.

Relative hatchability was significantly lower for infected than for uninfected *D. simulans* at G6 ($P = 0.0313$, Figure 3-8) but statistically indistinguishable ($P \geq 0.0528$, Figure 3-8) in subsequent generations of these flies. In *D. melanogaster*, relative hatchability was significantly lower for infected than for uninfected flies at G11 ($P = 0.0307$, Figure 3-8) but statistically indistinguishable ($P \geq 0.6761$, Figure 3-8) in the other two generations of these flies. Relative hatchability at G6 was significantly higher in infected *D. melanogaster* than in infected *D. simulans* ($P = 0.0313$, Figure 3-8) but there were not statistical differences between these two groups in subsequent generations ($P = 0.9783$, Figure 3-8). Finally, relative fecundity was significantly higher in infected *D. simulans* at G6 than at G16 ($P = 0.0455$, Figure 3-8) while there were not differences in these two generations for *D. melanogaster* ($P = 0.935$, Figure 3-8). Interestingly, hatchability was associated with virus titer. The eggs from flies with high virus titers hatched at less frequently than the eggs from flies with low virus titers ($P = 0.0442$, Figure 3-9).Thus, the virus is virulent in both species but it is more virulent in the novel host.

Discussion

Artificial Host-Shift and Selection for Transmission

The novel host (*D. simulans*) was easier to infect than the native host (*D. melanogaster*). The significant ease with which the DMelSV virus infected the novel host is certainly consistent with its ability to infect other *Drosophila* species (L'HERITIER 1958; PIZZATTO *et al.* 2010). However, it seems to be more difficult to

select for maximum transmission efficiency in the novel host; and moreover, transmission efficiency declined earlier and to a greater extent in the novel host than the native host following relaxation of selection. These trends are analogous to disproportionately higher dead-end, or sink, infections (i.e. the newly infected host is incapable of transmitting the pathogen) in the novel host; and may have occurred because of lower within-host reproduction, reduced between host transmission or both (BULLER *et al.* 2005; DENNEHY *et al.* 2006; MATROSOVICH *et al.* 1999; ROKYTA *et al.* 2005). Because DMelSV virus is vertically transmitted, virus-host interactions at the cellular level are critical for both within- and between-host transmission, and this interaction likely acted as a barrier to successful colonization of some of the novel host lineages (DENNEHY *et al.* 2007; PERLMAN and JAENIKE 2003).

However, another possibly complementary explanation for lower transmission efficiency in the novel host could be that DMelSV was more virulent in the novel host than the native host. Hatchability was significantly lower in infected than in uninfected *D. simulans* during the early stages of the infection. The failure of infected eggs to hatch would certainly have decreased transmission efficiency. However, virulence also decreased with time in some (but not all) lines. The result was a significantly different transmission trajectory for virus between native and novel host lines. It appears that, as with influenza and other host-shifted parasites (FRANK and SCHMID-HEMPEL 2008), the virulence of DMelSV attenuated with continual transmission in the novel host.

Transmission under Relaxed Selection

Under relaxed selection, transmission efficiency decreased faster in the novel host than in the native host. Unlike influenza, DMelSV is unipartite, so does not reassort, and there is no evidence that DMelSV recombines. As with other systems, it is likely that the number of virions transmitted between generations is low, leading to repeated bottlenecks (LAWRENCE and MATOS 2005). These are simple barriers to transmission that exist in both hosts. However, this combination of challenges caused a significant decrease in the transmission efficiency of the DMelSV earlier in the novel host than in the native host (Figure 3-3). Given that the success of parasites results from the interaction of the pathogen's genetics with its host, this result is perhaps unsurprising as it confirms that the virus is well-adapted to its native host. Additionally, this suggests that the predicted mutational capacity of RNA viruses in general, and DMelSV in particular, is insufficient to facilitate quick adaptation to the novel host (MATROSOVICH *et al.* 1999; SAMUEL 2011). Given that a genetically variable virus is expected to be more adaptable (ANDRE and HOCHBERG 2005; MOYA *et al.* 2000), by extension a more genetically robust virus should be less successful in an adaptive walk. DMelSV was extremely successful when aided by artificial selection, but experienced a drastic decrease in transmission efficiency once selection was relaxed and natural forces took over. Others have noted that DMelSV seems to have a low mutation rate (BRUSINI *et al.* 2012 in review; CARPENTER *et al.* 2009).

Virus Titers

Virus titers were highest in the native host, while variance in titer was higher in the novel host (Figure 3-4). Our results are in agreement with previous work showing that infection on the novel host can be inconsistent (ANTONOVICS *et al.* 2002). Moreover, once selection as relaxed following G4, the native host lines maintained high virus titers two generations later, while only 58% of the novel host lines did so (Figure 3-5). This again supports the notion that the cellular barriers to within-host replication and between-host transmission are stronger in the novel host (PERLMAN and JAENIKE 2003). It is likely that suppressed virus replication within the novel host resulted in decreased within and between host transmission, and led to the eventual crash of the virus population in several of the novel host lines by G6 and in most of the novel host lines by G16. This is in accord with the positive correlation between titer and transmission efficiency that was observed here (Figure 3-6) and in other systems (EWALD 1983; FRANK 1996; LENSKI and MAY 1994).

The high titers (high replication rates) suggest that the virus is more fit on the native host than on the novel host and that the lines with high titers have virus that successfully adapted to the host. However, DMelSV is vertically transmitted. As such, this virus should suffer the classic tradeoff between virulence and transmission (LIPSITCH *et al.* 1996; SHARON *et al.* 1999), but it does not, because higher titers within the host likely ensure increased transmission.

Virulence

The virulence of DMelSV was measured as a reduction in host fecundity and egg hatchability. Although the virus is virulent in both species, as predicted, it

is more virulent on the novel host, and it is more virulent in hosts with higher virus titers (EBERT 1998). Relative fecundity rebounds in the native host, despite high titers, indicating a reduction in virulence on the native host. In contrast, titer and fecundity both seem to decrease slightly with time in the novel host, suggesting continued virulence. Hatchability was lower on the native host at G6 but any species differences disappear subsequently indicating that the virus is adapting to the novel host. Interestingly, the fly lines with high virus titers also exhibited decreased hatchability regardless of species (Figure 3-9). However, hatchability was lower in the early generations of the novel host but rebounded by the end of the experiment. This shows that the virus adapted to both host but the adaptive walk took longer on the novel host. In conclusion, our data are consistent with predictions from serial passage experiments indicating that a parasite will be more virulent on a novel host.

Table 3-1. Sigma virus titers (mean ± standard error) in females of *Drosophila melanogaster* and *D. simulans* at six and 16 generations after artificial infection by injection.

	G6	G16
D. melanogaster	$1.3 \times 10^6 \pm 2.0 \times 10^5$	$1.1 \times 10^6 \pm 1.7 \times 10^5$
D. simulans	$1.0 \times 10^6 \pm 2.6 \times 10^5$	$1.0 \times 10^6 \pm 3.1 \times 10^5$

Titer is viral genomes per microgram of fly RNA.

Table 3-2. Mean sigma virus titers across species, line and generation (data presented in Figure 3-4).

Species	Line	Generation	
		Six	Sixteen
Drosophila melanogaster	DMelSV-207	599,984	555
	DMelSV-210	2,215,338	578
	DMelSV-211	1,563,223	551
	DMelSV-216	1,508,642	1,789,897
	DMelSV-235	976,914	843,461
	DMelSV-238	474,685	749,393
	DMelSV-302	964,328	2,668,256
	DMelSV-332	5,732,529	1,750,351
	DMelSV-339	2,747,062	1,511,670
	DMelSV-343	805,400	1,711,248
	DMelSV-349	1,012,250	2,474,306
	DSimSV-36	1,440,993	1,988
	DSimSV-52	1,506,997	3,544,808
Drosophila simulans	DSimSV-60	2,621	1,039
	DSimSV-61	3,120	173
	DSimSV-68	1,823	851
	DSimSV-70	580,547	118
	DSimSV-91	1,979,920	5,180,157
	DSimSV-138	125	694
	DSimSV-147	426	344
	DSimSV-151	3,101,075	18,143
	DSimSV-155	2,367,946	1,774,932
	DSimSV-177	804,883	1,767,926

Titer is viral genomes per microgram of fly RNA.

Table 3-3. Fecundity data (eggs/female/day) for sigma virus-infected and uninfected *Drosophila melanogaster* (native host) and *D. simulans* (novel host).

			G6		G11		G16	
		Line	Ave	StDev	Ave	StDev	Ave	StDev
Drosophila melanogaster	Infected	207	3.66	2.57	6.44	5.02	0.93	0.67
		210	3.38	4.00	3.94	2.46	3.71	3.70
		211	1.97	1.79	10.81	6.03	2.63	1.73
		216	4.88	2.32	7.13	4.17	3.31	1.46
		217	5.56	3.06	6.63	3.04	4.69	3.82
		225	2.69	4.11	--	--	--	--
		235	2.19	1.44	1.81	2.43	2.31	1.91
		238	4.22	2.06	4.56	2.11	3.13	2.62
		302	2.19	2.62	9.50	6.26	1.57	1.34
		332	5.47	2.65	4.75	2.12	1.63	1.55
		339	2.28	2.47	7.63	3.31	1.31	2.05
		343	3.41	1.60	2.75	1.34	3.31	1.83
		349	4.88	2.63	8.19	3.42	1.86	1.89
	Uninfected	201c	3.25	2.22	6.88	3.06	0.88	1.30
		204c	3.28	2.71	8.00	2.93	0.06	0.18
		205c	5.19	2.89	5.56	3.23	2.17	1.04
		211c	1.81	1.47	9.38	4.19	0.93	1.13
		214c	2.91	2.28	7.19	4.81	0.64	1.07
Drosophila simulans	Infected	31	8.53	2.55	1.00	0.65	6.88	2.53
		36	8.31	6.51	4.25	2.19	5.19	2.52
		52	5.03	2.54	3.50	3.05	1.75	1.16
		60	6.16	2.61	2.88	2.34	4.81	1.67
		61	6.28	2.42	2.44	1.35	3.69	1.58
		68	4.72	3.03	5.63	3.70	2.88	1.06
		70	0.78	0.59	0.75	1.25	1.86	1.03
		91	3.53	1.91	7.44	4.38	3.31	2.09
		138	7.97	3.38	5.38	2.85	3.93	1.06
		147	15.94	26.11	1.44	0.86	4.57	1.30
		151	4.47	1.22	5.38	2.88	2.25	0.71
		155	6.25	3.71	4.00	3.60	4.21	1.38
		177	5.88	1.94	11.25	7.03	5.00	2.12
	Uninfected	2c	6.56	3.56	3.56	5.25	5.19	4.54
		3c	5.78	2.74	2.74	3.63	3.69	1.62
		4c	2.13	1.99	1.99	4.63	5.88	2.62
		5c	4.34	2.83	2.83	1.63	1.50	1.16
		6c	3.03	1.92	1.92	2.44	3.71	1.75
		8c	5.81	2.31	2.31	2.56	4.64	0.63
		9c	6.25	2.01	2.01	5.38	6.38	2.83
		12c	6.81	2.16	2.16	1.94	3.88	1.06
		121c	3.91	2.56	2.56	4.75	3.56	2.80

Table 3-4. Hatchability data (% of eggs hatching) for sigma virus-infected and uninfected *Drosophila melanogaster* (native host) and *D. simulans* (novel host).

			G6		G11		G16	
		Line	Ave	StDev	Ave	StDev	Ave	StDev
Drosophila melanogaster	Infected	207	81.0	22.5	91.5	11.2	100.0	0.0
		210	55.2	39.8	85.0	14.0	78.1	25.6
		211	76.7	23.5	76.1	28.9	79.7	25.0
		216	77.0	19.9	79.8	18.8	68.3	18.5
		217	71.7	23.6	82.6	15.6	69.7	30.9
		225	70.4	25.7	--	--	--	--
		235	87.8	12.7	73.5	15.3	64.0	29.7
		238	66.2	29.1	71.1	28.8	86.7	16.3
		302	76.3	29.7	68.7	16.9	93.5	11.6
		332	64.9	35.0	71.7	16.8	72.9	28.4
		339	71.8	17.4	62.6	24.1	80.1	29.7
		343	70.4	27.6	91.0	10.7	54	33.4
		349	52.8	32.6	91.5	11.2	100	0.00
	Uninfected	201c	71.8	33.2	84.4	14.3	86.1	22.2
		204c	79.3	19.6	95.6	6.8	100.0	0.0
		205c	74.9	20.0	87.3	15.9	87.8	10.7
		211c	64.2	29.8	88.7	10.6	69.4	29.2
		214c	70.9	28.3	86.7	18.2	66.7	11.8
Drosophila simulans	Infected	31	60.5	22.1	94.4	16.7	56.1	30.4
		36	50.1	28.8	92.0	13.5	63.0	12.0
		52	77.0	17.9	69.3	21.2	76.9	23.5
		60	72.3	20.6	71.6	21.3	63.9	36.4
		61	70.3	22.2	65.2	25.9	84.1	19.8
		68	55.5	33.5	79.4	21.3	63.9	34.9
		70	72.3	25.0	66.7	31.2	72.5	30.9
		91	55.8	25.2	69.4	23.6	78.5	19.9
		138	71.7	22.4	76.2	18.9	64.5	38.2
		147	62.4	30.5	86.4	23.4	72.5	34.3
		151	58.3	24.7	95.4	10.1	60.7	31.7
		155	60.1	24.9	70.8	18.3	62.0	27.7
		177	52.0	24.2	57.6	23.6	54.9	40.1
	Uninfected	2c	74.0	26.2	85.9	15.9	65.9	32.1
		3c	78.5	18.9	85.5	14.1	52.5	32.0
		4c	61.4	31.7	76.4	19.8	65.8	26.0
		5c	75.9	29.6	95.5	10.1	60.5	29.3
		6c	62.0	35.0	89.5	21.3	77.6	26.4
		8c	88.6	17.5	79.9	21.3	71.5	31.7
		9c	65.6	31.8	82.1	20.3	58.4	25.4
		12c	67.6	34.0	86.9	19.0	90.6	18.6
		121c	83.9	20.8	93.1	9.1	70.9	26.8

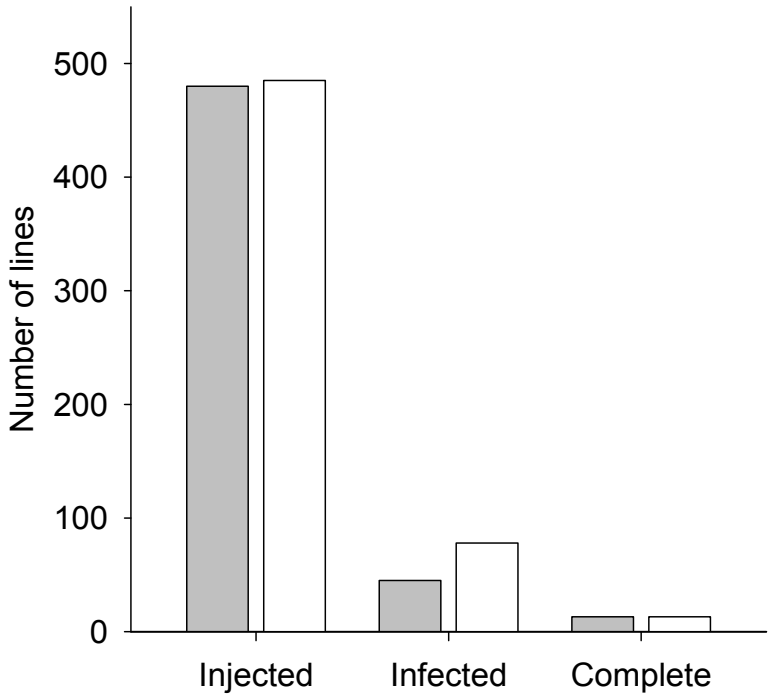

Figure 3-1. The novel host was easier to infect, but harder to stabilize. Total *Drosophila melanogaster* (grey bars) and *D. simulans* (white bars) females that were injected with sigma virus inoculum (leftmost bars), subsequently produced infected offspring (middle bars), and ultimately produced lines with 100% infected offspring (maximum transmission efficiency, rightmost bars). Significantly more *D. simulans* became infected than *D. melanogaster* ($P = 0.0015$); however, there was no significant difference in the proportion of infected flies that produced stable lines between the two species ($P = 0.1758$).

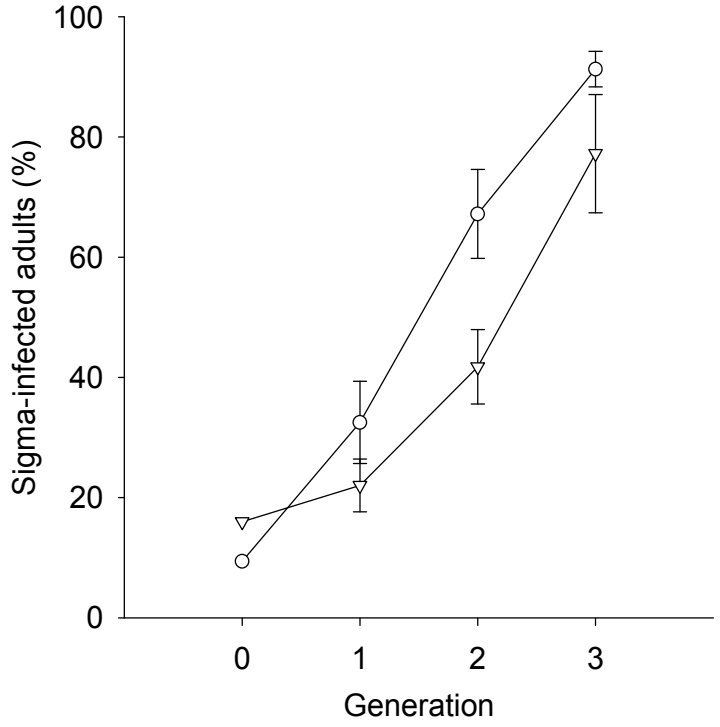

Figure 3-2. Maximum transmission efficiency (100%) was harder to achieve on the novel host. Per cent of *Drosophila melanogaster* (native host, *N* = 45 lines, circles) and *D. simulans* (novel host, *N* = 78 lines, triangles) adults infected with DMelSV following injection (G0) and during selection for transmission rate (G1 to G3). Transmission efficiency was higher in the native host than the novel host ($P < 0.0001$). Significant increase occurred between every generation in the native host ($P < 0.0001$) but only from G2 to G3 in the novel host ($P < 0.0001$). Error bars are two standard errors among lines.

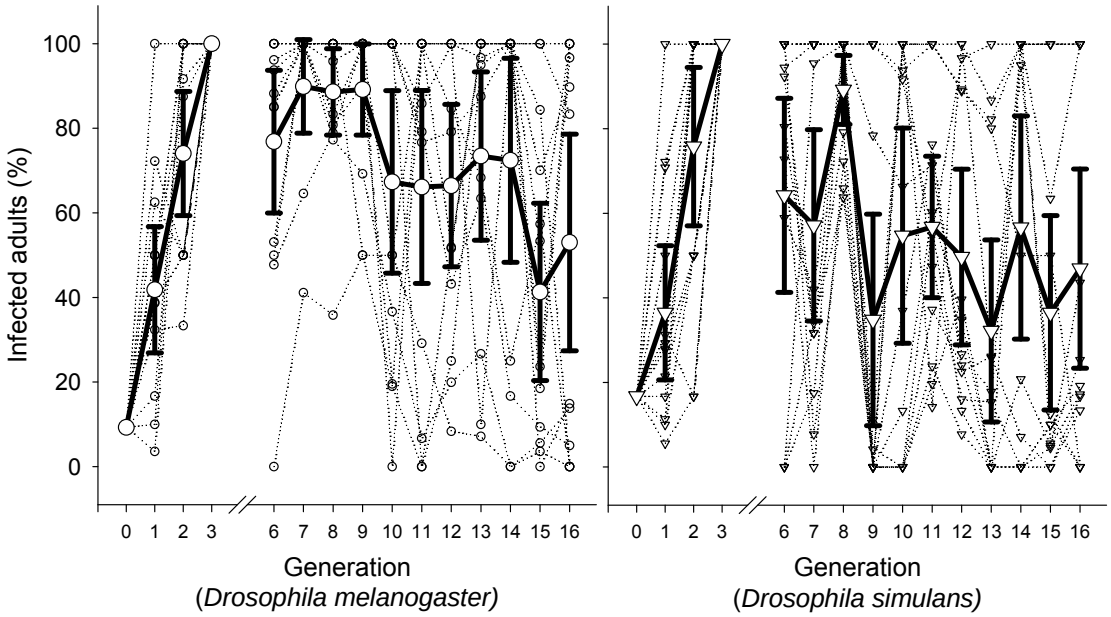

Figure 3-3. Under relaxed selection transmission efficiency decreased more rapidly on the novel host. Once selection was relaxed (after G3), the native host, *Drosophila melanogaster* (left panel), maintained higher average transmission efficiency (thick line with circles) than the native host (*D. simulans*, thick line with triangles) ($P < 0.0003$). Individual lines (thin dotted lines) were highly variable within each species. Significant decreases (from 100% at G3) occurred by G9 in *D. simulans* ($P < 0.0105$) and by G15 in *D. melanogaster* ($P < 0.0167$). Bars are two standard deviations (based on among line variance).

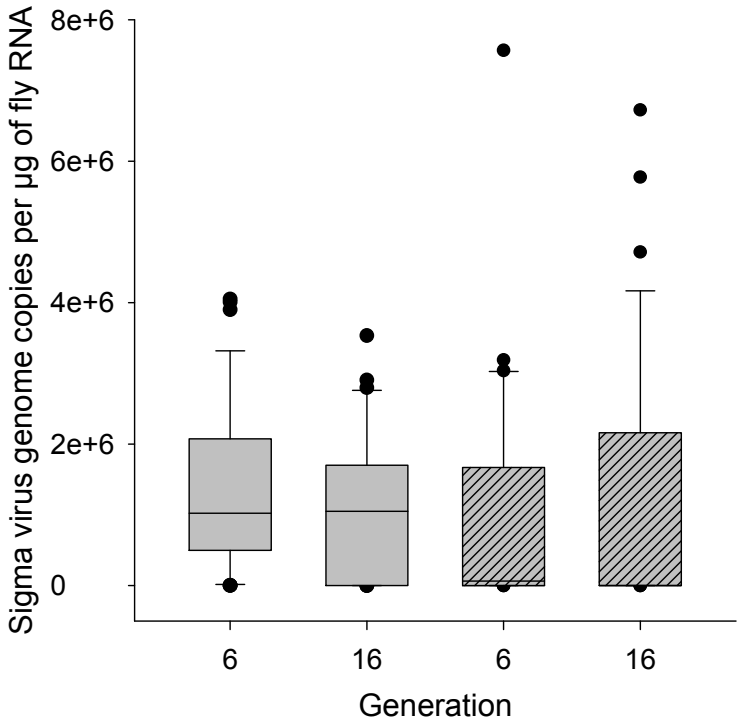

Figure 3-4. Virus titers were statistically indistinguishable in both species, but the means are lower and the variance is higher in females of *Drosophila simulans* (solid gray) than of *D. melanogaster* (hatched bars). The data failed Levene's homogeneity of variance test ($P = 0.009$); differences in titer ($P = 0.3820$) were tested using a Welch's analysis of variance. The summary of virus titers data are presented in Table 3-1.

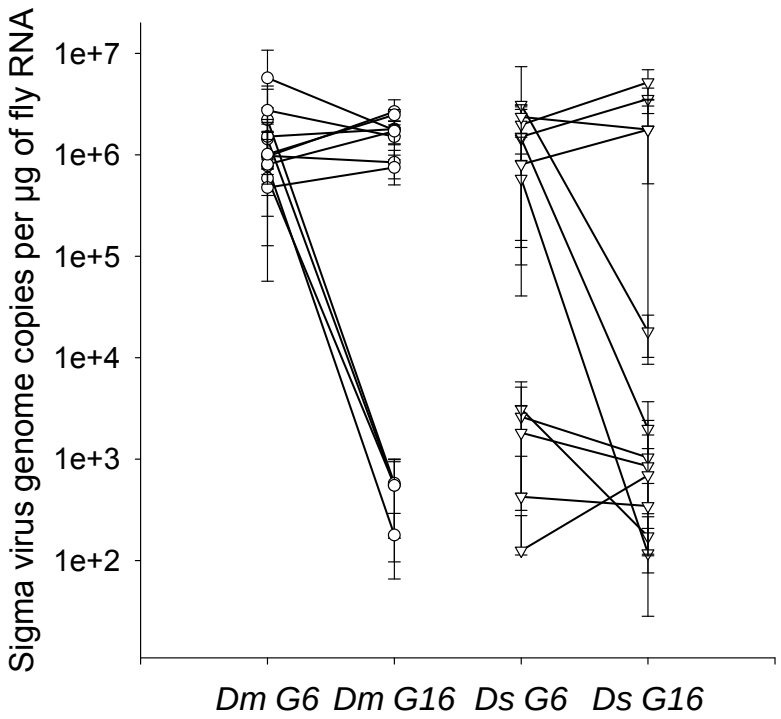

Figure 3-5. Heterogeneity in titer over time between lines for virus load. Titer in *Drosophila melanogaster* (*Dm*) and *D. simulans* (*Ds*) adults from unique lines was quantified at six and sixteen generations after artificial infection by injection. Raw data are presented in Table 3-2.

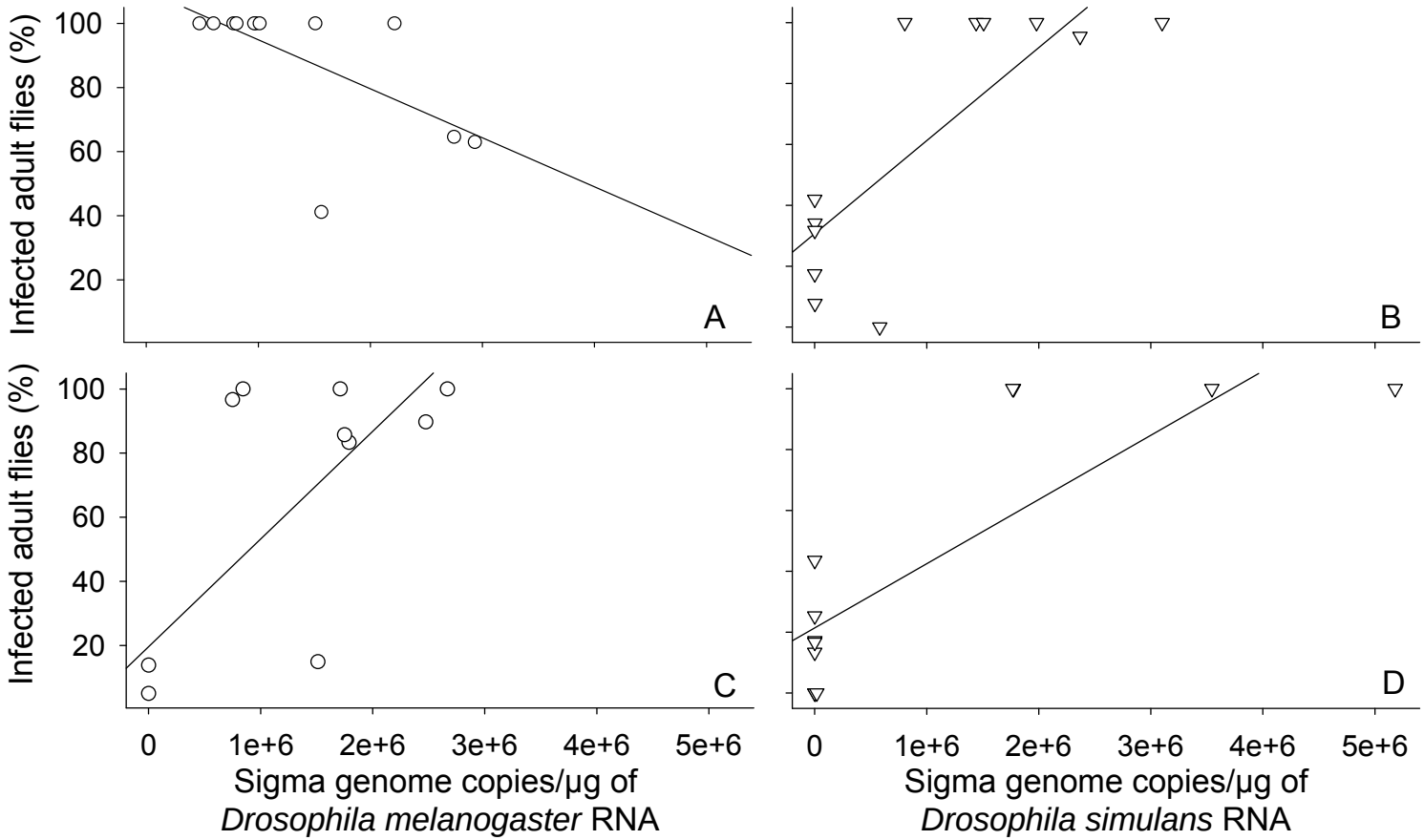

Figure 3-6. Virus titer is significantly correlated with transmission efficiency for both hosts, but the sign of the correlation changes with time. Kendall's *tau* correlation of DMelSV load and percent infected adults for *Drosophila melanogaster* (circles) and *D. simulans* (triangles) lines. A: *D. melanogaster* at G6, $\tau = -0.5394$, $P = 0.0289$; B: *D. simulans* at G6, $\tau = 0.4604$, $P = 0.0455$; C: *D. melanogaster* at G16, $\tau = 0.5002$, $P = 0.0265$; D: *D. simulans* at G16, $\tau = 0.3750$, $P = 0.1042$.

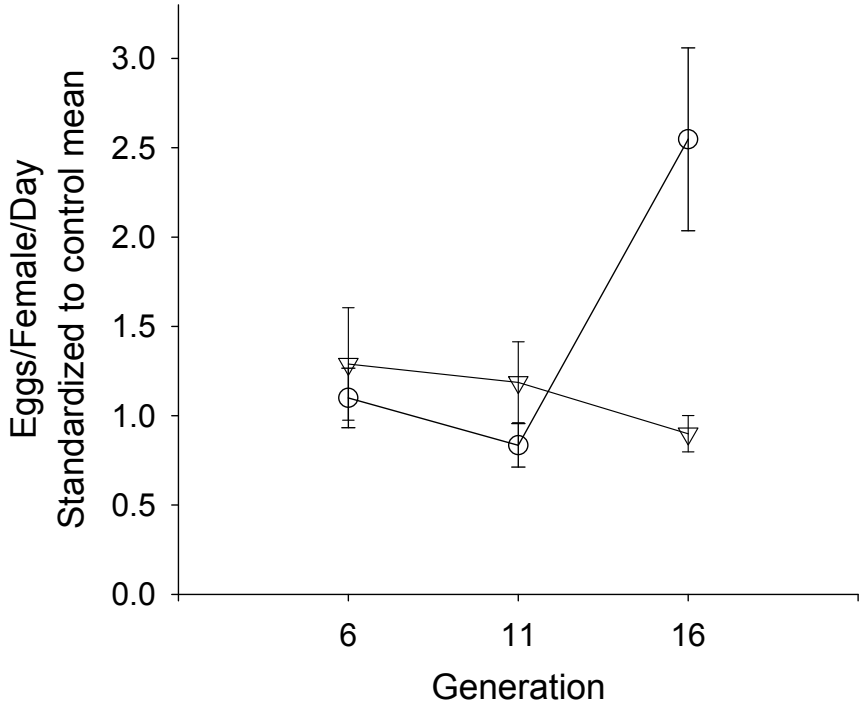

Figure 3-7 Fecundity was not negatively impacted by infection with sigma virus. Fecundity standardized to the relevant uninfected mean. Fecundity in the native host, *Drosophila melanogaster* (circles), rebounds in generation 16, while fecundity in the novel host, *D. simulans* (triangles), does not. Bars are two standard errors. Fecundity data are presented in Table 3-3.

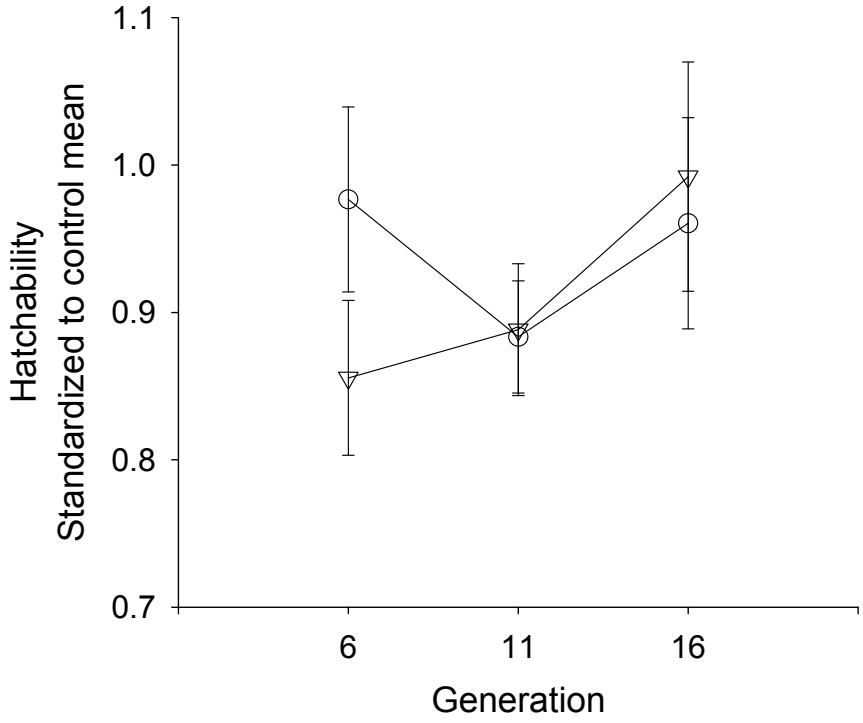

Figure 3-8. Hatchability was negatively impacted by infection with sigma virus in both hosts but at different generations. Hatchability standardized to the relevant uninfected mean. Hatchability in the native host, *Drosophila melanogaster* (circles), is higher in generation 6 than in the novel host, *D. simulans* (triangles). Hatchability tends to increase with time in *D. simulans*, but the trend is not significant. Bars are two standard errors. Hatchability data are presented in Table 3-4.

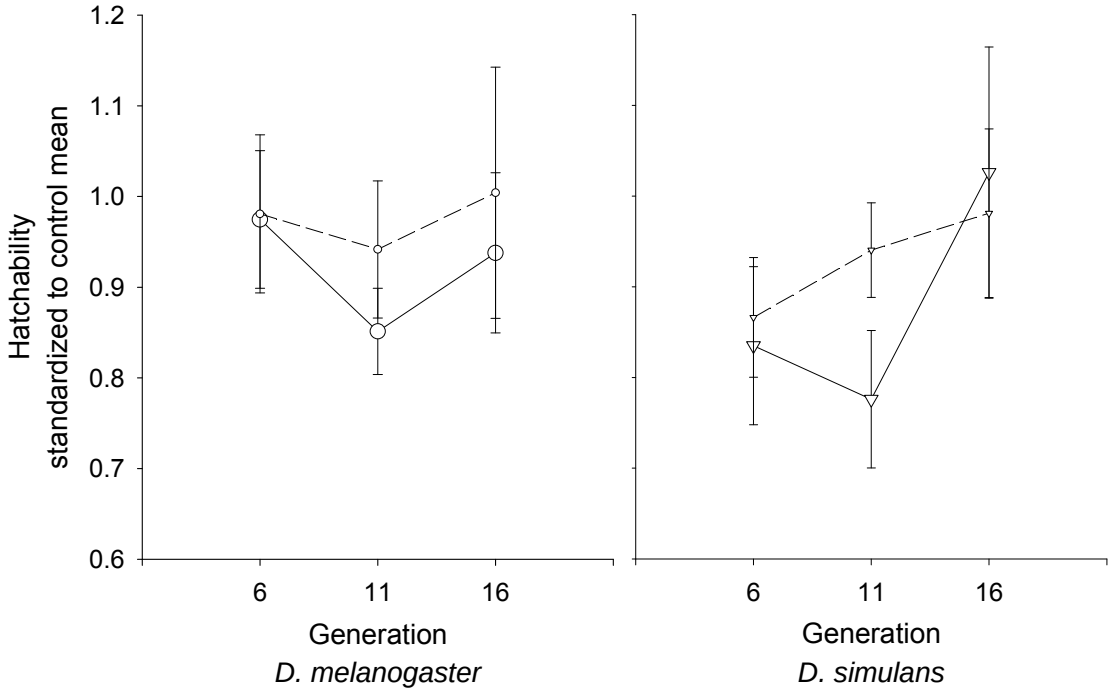

Figure 3-9. Hatchability standardized to the relevant uninfected mean for fly lines with high (large symbols) and low (small symbols) virus titers in *Drosophila melanogaster* and *D. simulans*. On average, eggs from high lines hatch at significantly lower levels than eggs from low lines ($P = 0.0442$), regardless of species. Bars are two standard errors.

CHAPTER 4
SEQUENCING AND ANALYSIS OF VIRAL GENOMES PASSAGED IN DROSOPHILA LINEAGES

Background

Emerging infectious diseases successfully colonize novel hosts because they can adapt to those hosts (LENDERBERGER ET AL. 1992). Therefore, understanding how parasites will evolve once they enter a novel host is of particular importance, particularly because adaptability has been identified as one of six major factors that can contribute to successful host shifts (LENDERBERGER ET AL. 1992). Most of the pathogens that we know are shifting successfully into novel hosts are RNA viruses (HOLMES and RAMBAUT 2004); therefore, it is necessary to gain a better understanding of how these types of viruses adapt to their novel host following a host shift.

Some general patterns of RNA virus evolution are known; RNA viruses tend to exhibit extensive convergent evolution with increased time spent in a novel host species (BULL et al. 1997; CUEVAS et al. 2002). This occurs because beneficial mutations are usually fixed early and contribute the most to viral fitness in the new environment (BULL *et al.* 1997; WAHL and KRAKAUER 2000). Additionally, mutation rate is constant and high, while genetic diversity is low during the course of adaptation and increases once fitness plateaus (WICHMAN *et al.* 1999a). Finally, mutations that are favorable in the novel host usually decrease fitness in the original host (BULL *et al.* 1997). Most of these observations were made in serial passage experiment studies (SPEs) (EBERT 1998); therefore, little is known about viral evolution in multicellular eukaryotes .

We will examine these generalities in the DM/DS/sigma system by sequencing the

major genomic variant from sigma virus strains passaged naturally in the novel host *D. simulans* for sixteen generations.

We have developed the sigma rhabdovirus and Drosophila spp. as a model system in which we can better understand viral adaptation following host shifts. The sigma rhabdovirus is a good model pathogen because like many other viruses of interest it has an RNA genome (- ssRNA). An ancestral sigma rhabdovirus inoculum was generated from wild-caught *D. melanogaster* females collected in Athens, GA (USA). This virus was artificially shifted by injection onto novel hosts (*D. simulans*). Additionally, the virus was mock shifted onto its native host (*D. melanogaster*) to control for any effects of procedures or the environment. Subsequently the virus was passaged naturally through sixteen host generations (Chapter 3). Here we test whether the shifted virus experienced convergence on the novel host.

Materials and Methods

Ribonucleic Acid (RNA) Purification

Ribonucleic acid from each inoculum was isolated using TRIzol (www.invitrogen.com), following the manufacturer's directions, as described previously (Chapter 3). Briefly, a 50 µl aliquot of inoculum was mixed with 200 µl of TRIzol on ice. The mixture was vortexed briefly and incubated at room temperature (RT) for 5 min. Each sample then received 40 µl of chloroform and was vortexed for five seconds, incubated at RT for 2-3 min, then centrifuged at 12,000g for 10 min at 2°C. The RNA in the aqueous phase was precipitated with isopropanol at -20°C for 10 min. The precipitate was pelleted by centrifugation at

12,000g for 10 min at 2°C. The pellet was washed with cold (4°C) ethanol, air dried for five min at RT, resuspended in 29 µl of DEPC-treated water with 1 µl of 40 µM RNAse inhibitor (www.neb.com) and incubated at 55°C for 10 min to aid resuspension. Samples were stored at -80°C until used.

Reverse Transcription

Reverse transcription of the DMelSV virus genome was performed using the SuperScript III First-Strand cDNA synthesis system (www.invitrogen.com) following the manufacturer's recommendations. Briefly, 200ng of inoculum RNA were mixed with 1 µl of primer (whole1F: 5' TAGAAGCATCCTCGGCTTTC 3') and 1 µl of 10mM dNTP mix and the reaction volume was brought to 10 µl with DEPC-treated water. This reaction (part 1) was incubated at 60°C for 5 min and then placed on ice. A cDNA synthesis mix (2 µl of 10X RT buffer, 4 µl of 25mM $MgCl_2$, 2 µl of 0.1mM DTT, 1 µl of 40U/ µl RNAseOUT and 1 µl of 200U/ µl SuperScript III enzyme, part 2) was prepared and added to part 1 and the mixture was incubated at 50°C for 50 min, then at 85°C for 5 min to stop the reaction and then chilled on ice. RNAse H (1 µl) was added to the tube and the synthesis reaction was incubated at 37°C for 20 min. The cDNA was stored at -20°C.

Viral Genome Sequencing

The genome of the major viral variant of the ancestral virus (DMelSV-A) and of the virus evolved in *D. simulans* (DSimSV-E) were sequenced in 19 overlapping fragments that were produced using 19 primer pairs (Table 4-1). All amplicons were amplified using the following PCR parameters: 94°C for 2 min followed by 40 cycles of 94°C for 30 seconds, 55°C for 60 seconds and 72°C for

60 seconds; a final elongation step of 5 min at 72°C was done after the cycling. The amplicons were electrophoresed to check for quality and sent for sequencing by Genewiz (www.genewiz.com) using their standard protocols. Each amplicon was sequence with the forward and reverse primer. The sequenced amplicons were assembled and compared using the Geneious 5.6 software package (www.geneious.com). Final genomes were assembled to a 99% quality score.

Further details on the materials and methods used here are presented in Appendix B.

Statistics

All statistical tests were implemented in MEGA 5.0 (TAMURA et al. 2011) unless otherwise noted in the text. The hypothesis of positive selection (HA: dN > dS) was tested for each of the six viral genes for all lines using the Z-test of selection. The bootstrap method (1000 replications) was used with a syn-nonsynonymous substitution setting using the Nei-Gojobori method (Jukes-Cantor). Finally, the transition/transversion bias was determined using a NJ (neighbor joining) tree examining nucleotides using the Kimura 2-parameter model. A phylogenetic tree of the twelve complete viral genomes was constructed using the maximum likelihood method with a Jukes-Cantor model, and robustness of the nodes was evaluated using the bootstrap method (1000 replications).

Results

In total, twelve viral genomes were sequenced: the ancestral genome DMelSV-A, five genomes evolved in *D. simulans* (DSimSV-E) and six genomes evolved in *D. melanogaster* (DMelSV-E). The sequence polymorphisms for all of

the eleven derived lines sequenced are shown in Table 4-2 and these data are summarized in Table 4-3. Three DSimSV-E lines had genomes carrying many differences when compared to the ancestral genome (10.73 ± 6.74 substitutions/Kb); predictably, only 1.82 ± 1.38 per Kb were non-synonymous substitutions. Two DSimSV-E lines had low numbers of synonymous and non-synonymous substitutions (0.153 ± 0.308 and 0.0267 ± 0.0923 substitutions/Kb, respectively) for the duration of the experiment. Finally, the DMelSV-E lines as a group had the lowest numbers of synonymous and non-synonymous substitutions (0.0.277 ± 0.619 and 0.0267 ± 0.092 substitutions/Kb, respectively).

For all the lines, synonymous changes outnumbered non-synonymous ones, as expected. Indeed, statistical analyses (*Z* test as described in Materials and Methods) found no evidence supporting positive selection (*P* = 1.0 for each of the genes). This is not surprising given the relative scarcity of non-synonymous changes. Finally, five of the six genes had a slight transition bias. Gene *M* was the exception, with a very large bias (Table 4-4). The eleven lines sequenced carry 10 polymorphisms in the M gene: five G-to-A, three U-to-C and two C-to-U, all transitions. The full sequence of the DMelSV-A is presented as a reference in Table 4-5.

In the phylogenetic analysis (Figure 4-1) the three DSimSV-E lines that had many polymorphisms grouped together on their own branch with 100% bootstrap support. The ancestral virus shared a node with a branch containing all of other sequences (six DMelSV and two DSimSV) that are extremely similar to

the ancestor. Interestingly, the two DSimSV lines grouped on different branches within this cluster (95% bootstrap support for both of these branches).

Discussion

It is clear that the ancestral virus was well adapted to its native host, as the virus changed little during 16 generations on naive native host lines; however, beyond that simple conclusion, the data are perplexing. Attempting to correlate the sequence data to transmission efficiency, viral copy number, and virulence in the three lines harboring the genetically different evolved virus did not yield any explanatory patterns (even if not significant). The three DSimSV-E virus lines that had very different genetics are neither more fit (higher titer or greater transmissibility, consistent with an adaptive change) nor less fit (lower titer or lower transmissibility). Additionally these viruses are not more or less virulent than the other two viruses that were very similar to the ancestral virus genome. The presence of two major classes (many substitutions vs. few substitutions) of virus passaged in the novel host suggests either a great deal of stochasticity such that different strategies were employed or multiple challenges were met. One possibility is that upon infection, some virions (in the low variants) encountered by chance favorable cell populations and were able to immediately establish themselves as the major variant in that fly lineage. This scenario is equally likely for the low SNP and high SNP variants, assuming that the genetic variants present in the three distinct, fast evolving DSimSV-E lines were low variants in the inoculum but also were viable and replication competent upon being injected into the novel host. Although sigma is less genetically diverse that one might expect for an RNA virus (BRUSINI *et al.* 2012 in review; CARPENTER *et*

al. 2009), we showed in Chapter 2 that the wild population from which the ancestral inoculum was generated still carried a significant amount of genetic diversity. Moya et al. (2000) demonstrated that bottlenecks reduce viral fitness and, as seen in Chapter 3, many of the artificially generated infections failed because viral fitness was low (low transmission). However, rare variants that may be maladapted to their native host may have had a fitness advantage over the common variants and, as such, were more fit in the novel host (relative to the common variant) and persisted. This seems unlikely, unless the model of adaptation underlying the *Z* test, which emphasizes the role of amino acid substitutions rather than synonymous changes, is not appropriate for virus evolution. Indeed, silent changes may be very important as they may lead to codon deoptimization and thus reduced virus fitness at a level that was not detectable here (MUELLER *et al.* 2006). An alternative explanation which cannot be falsified with our current data is that the mutation rate has increased in these three lines. More directly evaluating these hypotheses would require sequencing viral variants collected in generations of host immediately following the host shift to determine whether these polymorphic variants establish themselves soon after the host shift or evolve over time. Given that the two variants with few polymorphisms appeared to have equivalent fitness (i.e. transmission and titer) to the three fast evolving strains, it seems more likely that stochastic events such as a bottleneck produced by the host shift facilitated the rapid establishment of a completely new major variant within the novel host. Similarly, any mutational increase was not adaptive per se. In either case, bottleneck or mutational

speedup, evolution did not occur by small steps (BURCH and CHAO 1999; FISHER 1930).

Table 4-1. Sequences for 19 primer pairs using to produce overlapping amplicons from viral complementary DNA (BRUSINI *et al.* 2012 in review).

Forward primer ID and sequence		Reverse primer ID and sequence	
44-	TAGAAGCATCCTCGGCTTTC	976+	CTTCCCTGTCTTTGCACCTC
679-	ACACACTTGGTCGGAAGGAG	1638+	CAAATTGCAACTGGCTCTCA
1323-	TGTCCACAAAGTATTTCCATGTC	2273+	TGAGTAGGCTGGTTCAAAACG
1964-	TTCTGCCATATGTCCTTTGC	2916+	TTGCATCGCTTCAGTTGTTC
2640-	AATAAGGTGCAGGGCTTTCC	3608+	GGTTCCTCATGCATCTGACA
3302-	TAAACCACTTGCTGCCATGA	4275+	GGGAAGACTTGTGTCATCCAT
3989-	AGTCGGCTGAATGATTGTCC	4894+	CATGGCAGATGACCTCACAA
4610-	TTCATCAGATTGCCCTCTCC	5482+	GACGACGGTGGAGAAAGAGA
5262-	AAGGTAATGTGTGCCATCCA	6164+	AGATCACCTTGCCAGAGGAA
5925-	GGTCCAGCATCCTGTTGTG	6852+	TCAATTGTCGCACTGGTTGT
6559-	TGTGTCCATAGAGAACGGATCA	7522+	CCATCCTAACTCTGTGACATTCC
7209-	TCTCCTCCGTGTTCAGTGTT	8173+	GTCATACGACCCCAAGAAGC
7885-	GTACGCCTTCGTGTTCACTG	8799+	TTCAAGCCTATTTTATTGTTCGTG
8535-	CCTCAATGAAGGACTCGATAGG	9444+	GCTGCAAATGACACTGGATG
9181-	TCGTCCATGTAATGTATCTTCACTG	10105+	CACCGTGCTTTACCATGAAA
9842-	GGCCCTCTTGGAAGTGCT	10762+	TGGACCAGGGTAATGAGAGC
10516-	GGGGTGCTCCACATCACTAT	11431+	TCACCGGGAAGAAGATATGG
11167-	GAGCTGTCACACCCGAGTTA	12028+	GATCCTCGCTCCATTTTCAA
11848-	TAATCCGATGGTTCCAAAGC	12498+	GACACCAAGACCGAGTTCGT

Table 4-2. Sigma virus strains evolved in *Drosophila simulans* (DSimSV-E) or *D. melanogaster* (DMelSV-E) were compared to the ancestral virus (DMelSV-A). New bases are given for each of the viral strains when a single nucleotide polymorphism was detected. The new amino acid is given whenever the SNP was non-synonymous.

Gene	N	N	N	N	N	N	N	N	N	N	N	N	N	N	NC	NC
Codon Number	10	81	88	93	114	133	168	222	228	229	249	250	270	341		
DMelSV-A codon	GCU	UUA	AUC	AAC	AAG	UCU	GGA	AAU	GAU	GUG	UCU	GCC	CCA	ACC	A	C
SNP codon position	3	1	1	3	3	1	3	1	1	1	3	3	3	3		
Consensus AA	A	L	I	N	K	S	G	N	D	V	S	A	P	T		
DSimSV-E-52		C			A	G		G			C	U	G		G	U
New AA								D								
DSimSV-E-91																
New AA																
DSimSV-E-151		C			A	G		G			C	U	G		G	U
New AA								D								
DSimSV-E-155	C	C	G	U	A	G	G				C	U	G	U	G	U
New AA			V													
DSimSV-E-177																
New AA																
DMelSV-E-216																
New AA																
DMelSV-E-238																
New AA																
DMelSV-E-302																
New AA																
DMelSV-E-332									A	A						
New AA									N	M						
DMelSV-E-343																
New AA																
DMelSV-E-349																
New AA																

Table 4-2. Continued.

Gene	P	P	P	P	P	P	NC	NC	NC	X	X	X	X	X	X
Codon Number	479	481	482	550	636	695				773	780	882	893	905	914
DMelSV-A codon	GGU	GCG	UAU	GCA	CCC	AAC	G	C	G	AUU	AGA	AGC	AUC	CCU	ACG
SNP codon position	3	3	1	1	3	3				1	3	1	2	3	2
Consensus AA	G	A	Y	A	P	N				I	R	S	I	P	T
DSimSV-E-52	C	A	C	U		U	A	U	A	G	G	G	C	C	
New AA			H	S						V		G	T		
DSimSV-E-91															
New AA															
DSimSV-E-151	C	A	C	U		U	A	U	A	G	G	G	C	C	
New AA			H	S						V		G	T		
DSimSV-E-155	C	A	C	U	U		A	U		G	G	G	C	C	U
New AA			H	S						V		G	T		M
DSimSV-E-177															
New AA															
DMelSV-E-216															
New AA															
DMelSV-E-238															
New AA															
DMelSV-E-302															
New AA															
DMelSV-E-332															
New AA															
DMelSV-E-343															
New AA															
DMelSV-E-349															
New AA															

Table 4-2. Continued.

Gene	X	X	X	X	X	X	X	M	M	M	M	M	NC	G
Codon Number	929	967	976	1014	1022	1024	1058	1120	1167	1207	1242	1292		1340
DMelSV-A codon	CGC	CGA	ACU	GUG	UAU	AUU	ACA	UUA	ACU	UAC	GAG	CAG	C	CAU
SNP codon position	3	1	3	3	1	3	3	1	3	3	3	3		3
Consensus AA	R	R	T	V	Y	I	T	L	T	Y	E	Q		H
DSimSV-E-52			A	A			C		C		A	A		G
New AA														Q
DSimSV-E-91														
New AA														
DSimSV-E-151			A	A			C		C		A	A		G
New AA														Q
DSimSV-E-155	U	A	A	A	C	C	C			U	A			G
New AA					H									Q
DSimSV-E-177														
New AA														
DMelSV-E-216														
New AA														
DMelSV-E-238													U	
New AA														
DMelSV-E-302								C						
New AA														
DMelSV-E-332														
New AA														
DMelSV-E-343														
New AA														
DMelSV-E-349														
New AA														

Table 4-2. Continued.

Gene	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
Codon Number	1355	1358	1376	1384	1388	1390	1416	1421	1442	1449	1467	1510	1543	1570	1606
DMelSV-A codon	CUG	AUC	GUA	AGU	GUG	AAC	GUU	CUA	GAU	CCU	GAC	GAU	AAG	CGG	CAU
SNP codon position	1	3	3	3	3	3	1	1	3	3	3	3	3	3	3
Consensus AA	L	I	V	S	V	N	V	L	D	P	D	D	K	R	H
DSimSV-E-52			G	C	A	U		U	C	A	U		A	A	C
New AA															
DSimSV-E-91															
New AA															
DSimSV-E-151			G	C	A	U		U	C	A	U		A	A	C
New AA															
DSimSV-E-155	U	A	G	C	A	U	A	U	C	A	U		A	A	C
New AA							I								
DSimSV-E-177												C			
New AA															
DMelSV-E-216															
New AA															
DMelSV-E-238															
New AA															
DMelSV-E-302															
New AA															
DMelSV-E-332												C			
New AA															
DMelSV-E-343															
New AA															
DMelSV-E-349												C			
New AA															

Table 4-2. Continued.

Gene	G	G	G	G	G	G	G	G	G	G	G	NC	L	L
Codon Number	1654	1656	1728	1759	1769	1819	1852	1859	1859	1877	1895		1952	1997
DMelSV-A codon	CUU	GUC	AAC	GAG	UCU	GCA	AUC	UUU	UUU	AGG	UCC	C	GAG	ACG
SNP codon position	3	3	3	2	3	3	3	1	3	2	3		2	3
Consensus AA	L	V	N	E	S	A	I	F	F	R	S		E	T
DSimSV-E-52	C	U	U		C	G	U	C	C	A	A	U		A
New AA								L	L	K				
DSimSV-E-91														
New AA														
DSimSV-E-151	C	U	U		C	G	U	C	C	A	A	U		A
New AA								L	L	K				
DSimSV-E-155	C	U	U	G	C		U	C	C	A	A	U	G	A
New AA				G				L	L	K			G	
DSimSV-E-177														
New AA														
DMelSV-E-216														
New AA														
DMelSV-E-238														
New AA														
DMelSV-E-302														
New AA														
DMelSV-E-332														
New AA														
DMelSV-E-343														
New AA														
DMelSV-E-349														
New AA														

Table 4-2. Continued.

Gene	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Codon Number	2021	2025	2036	2039	2040	2042	2074	2087	2104	2118	2130	2132	2193	2198	2211
DMelSV-A codon	GGA	CUA	AAA	CUU	GAA	GCU	CUG	GAA	ACC	CCC	UUC	CCC	AUA	UUG	UUG
SNP codon position	3	1	3	3	3	3	1	3	3	3	3	2	1	1	1
Consensus AA	G	L	K	L	E	A	L	E	T	P	F	P	I	L	L
DSimSV-E-52	C	U	G		G	A	U	G	U	A	U	U	C	C	C
New AA												L	L		
DSimSV-E-91															
New AA															
DSimSV-E-151	C	U	G		G	A	U	G	U	A	U	U	C	C	C
New AA												L	L		
DSimSV-E-155	C	U	G	C	G	A	U	G	U	A		U	C	C	C
New AA												L	L		
DSimSV-E-177															
New AA															
DMelSV-E-216															
New AA															
DMelSV-E-238															
New AA															
DMelSV-E-302															
New AA															
DMelSV-E-332															
New AA															
DMelSV-E-343															
New AA															
DMelSV-E-349															
New AA															

Table 4-2. Continued.

Gene	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Codon Number	2237	2291	2334	2336	2377	2415	2436	2448	2455	2488	2507	2555	2601	2622	2626
DMelSV-A codon	UUC	GGG	GUU	CCU	GAC	CGG	GUG	AUC	AGA	CCG	UUG	CCG	AAG	AGA	ACU
SNP codon position	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Consensus AA	F	G	V	P	D	R	V	I	R	P	L	P	K	R	T
DSimSV-E-52		U	C	A	U	A	A	A	G	U	A		A	G	C
New AA															
DSimSV-E-91												A			
New AA															
DSimSV-E-151		U	C	A	U	A	A	A	G	U	A		A	G	C
New AA															
DSimSV-E-155	U	U		A	U			A	G	U	A		A	G	C
New AA															
DSimSV-E-177															
New AA															
DMelSV-E-216															
New AA															
DMelSV-E-238															
New AA															
DMelSV-E-302															
New AA															
DMelSV-E-332															
New AA															
DMelSV-E-343															
New AA															
DMelSV-E-349															
New AA															

Table 4-2. Continued.

Gene	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Codon Number	2635	2654	2738	2750	2812	2835	2855	2871	2886	2923	2945	2951	2957	2988	2989
DMelSV-A codon	GCC	CAG	ACG	UCU	GCA	UUC	UUU	CUU	GCU	GGA	GGA	GCU	UUA	AUU	AGA
SNP codon position	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Consensus AA	A	Q	T	S	A	F	F	L	A	G	G	A	L	I	R
DSimSV-E-52	A	A	A	C	C	U		G	A	G		C	G	C	G
New AA															
DSimSV-E-91															
New AA															
DSimSV-E-151	A	A	A	C	C	U		G	A	G		C	G	C	G
New AA															
DSimSV-E-155	A	A	A	C	C			G	G	G	G	C	G	C	G
New AA															
DSimSV-E-177															
New AA															
DMelSV-E-216															
New AA															
DMelSV-E-238															
New AA															
DMelSV-E-302							C								
New AA															
DMelSV-E-332															
New AA															
DMelSV-E-343															
New AA															
DMelSV-E-349															
New AA															

Table 4-2. Continued.

Gene	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Codon Number	3028	3068	3069	3089	3108	3113	3156	3164	3172	3202	3234	3296	3341	3359	3361
DMelSV-A codon	ACA	AAG	GGU	ACA	GCA	GGG	CGA	GGU	CCG	CUC	AUU	UUC	ACG	ACC	CUC
SNP codon position	3	1	2	3	3	3	3	1	3	3	3	3	3	3	3
Consensus AA	T	K	G	T	A	G	R	G	P	L	I	F	T	T	L
DSimSV-E-52	U	C	A	G	G	A	C	A	A	A			A		A
New AA		Q	D					S							
DSimSV-E-91											C			U	
New AA															
DSimSV-E-151	U	C	A	G	G	A	C	A	A	A			A		A
New AA		Q	D					S							
DSimSV-E-155	U	C	A	G	G	A	C	A	A	A		U	A		A
New AA		Q	D					S							
DSimSV-E-177											C	U		U	
New AA															
DMelSV-E-216											C	U		U	
New AA															
DMelSV-E-238											C			U	
New AA															
DMelSV-E-302											C			U	
New AA															
DMelSV-E-332											C	U		U	
New AA															
DMelSV-E-343											C	U		U	
New AA															
DMelSV-E-349											C	U		U	
New AA															

Table 4-2. Continued.

Gene	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Codon Number	3392	3399	3426	3430	3465	3474	3525	3531	3655	3666	3755	3769	3771	3808	3820
DMelSV-A codon	UCG	CCU	GCC	UCU	GCA	CUG	AUA	AGC	AAU	AGC	CAG	UCA	GUU	AUA	UUG
SNP codon position	3	3	3	3	3	3	2	3	2	3	3	3	3	1	1
Consensus AA	S	P	A	S	A	L	I	S	N	S	Q	S	V	I	L
DSimSV-E-52	A	C	G	C		A	C	A		U		G	G	C	
New AA							T	R						L	
DSimSV-E-91															C
New AA															
DSimSV-E-151	A	C	G	C		A	C	A		U		G	G	C	
New AA							T	R						L	
DSimSV-E-155	A	C	G	C		A	C	A				G		C	
New AA							T	R						L	
DSimSV-E-177									G		U				
New AA									S		H				
DMelSV-E-216					G										
New AA															
DMelSV-E-238															
New AA															
DMelSV-E-302															C
New AA															
DMelSV-E-332									G		U				
New AA									S		H				
DMelSV-E-343					G										
New AA															
DMelSV-E-349									G		U				
New AA									S		H				

Table 4-2. Continued.

Gene	L	L	L	L	L	L	NC
Codon Number	3825	3872	3881	3927	3996	3997	
DMelSV-A codon	GGG	UUG	UAC	CAA	GAU	AAU	C
SNP codon position	3	1	3	3	3	3	
Consensus AA	G	L	Y	Q	D	N	
DSimSV-E-52	A		U		C	C	U
New AA							
DSimSV-E-91		C					
New AA							
DSimSV-E-151	A		U		C	C	U
New AA							
DSimSV-E-155	A		U		C	C	U
New AA							
DSimSV-E-177							
New AA							
DMelSV-E-216							
New AA							
DMelSV-E-238				G			
New AA							
DMelSV-E-302		C					
New AA							
DMelSV-E-332							
New AA							
DMelSV-E-343							
New AA							
DMelSV-E-349							
New AA							

Table 4-3. Summary of synonymous and non-synonymous substitutions in ancestral virus evolved in *Drosophila melanogaster* (DMelSV-A) and in strains evolved in *D. simulans* (DSimSV-E) and *D. melanogaster*.

Evolved Strain	Gene	Total SNPs	Syn	Non-Syn	Total SNPs/Kb	Non-Syn /Kb
DSimSV-E-52	N	7	6	1	5.76	0.82
	P	5	3	2	5.31	2.12
	X	8	5	3	8.53	3.20
	M	3	3	0	4.49	0
	G	22	18	4	12.29	2.23
	L	68	60	8	10.80	1.27
	NC	8	--	--	25.81	--
DSimSV-E-91	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	0	0	0	0	0
	L	5	5	0	0.79	0
	NC	0	--	--	0	--
DSimSV-E-151	N	7	6	1	5.76	0.82
	P	5	3	2	5.31	2.12
	X	8	5	3	8.53	3.20
	M	3	3	0	4.49	0
	G	22	18	4	12.29	2.23
	L	68	60	8	11.87	1.62
	NC	8	--	--	25.81	--
DSimSV-E-155	N	11	10	1	9.05	0.82
	P	5	3	2	5.31	2.12
	X	13	8	5	13.86	5.33
	M	2	2	0	2.99	0
	G	25	19	6	13.97	3.35
	L	66	57	9	10.48	1.43
	NC	7	--	--	22.58	--

N= nucleoprotein, P = polymerase assisting phosphoprotein, X = unknown function, M = matrix protein, G =glycoprotein, L =polymerase, NC = non-coding sequence, SNPs = single nucleotide polymorphisms = synonymous, Kb = kilobase.

Evolved Strain	Gene	Total SNPs	Syn	Non-Syn	Total SNPs/Kb	Non-Syn /Kb
DSimSV-E-177	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	1	1	0	0.56	0
	L	5	3	2	0.79	0.32
	NC	0	--	--	0	--
DMelSV-E-216	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	0	0	0	0	0
	L	4	4	0	0.64	0
	NC	0	--	--	0	--
DMelSV-E-238	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	0	0	0	0	0
	L	3	3	0	0.48	0
	NC	1	--	--	3.23	--
DMelSV-E-302	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	1	1	0	1.50	0
	G	0	0	0	0	0
	L	5	5	0	0.79	0
	NC	0	--	--	0	0
DMelSV-E-332	N	2	0	2	1.65	1.65
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	1	1	0	0.56	0
	L	5	3	2	0.79	0.32
	NC	0	--	--	0	--

Table 4-3 Continued.

Evolved Strain	Gene	Total SNPs	Syn	Non-Syn	Total SNPs/Kb	Non-Syn /Kb
DMelSV-E-343	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	0	0	0	0	0
	L	4	4	0	0.64	0
	NC	0			0	0
DMelSV-E-349	N	0	0	0	0	0
	P	0	0	0	0	0
	X	0	0	0	0	0
	M	0	0	0	0	0
	G	1	1	0	0.56	0
	L	5	3	2	0.79	0.32
	NC				0	

Table 4-4. Summary of statistical analysis comparing each of the six genes from twelve viral variants. One was the ancestral virus and eleven were passaged in flies for sixteen host generations.

Gene	*dN* > *dS*		*R* (trans/trans)
	P	*Z* Statistic	
G	1	-4.246	5.56
L	1	-7.804	2.84
M	1	-2.107	10558160
N	1	-2.493	13.08
P	1	-1.514	5.02
X	1	-1.739	3.35

Table 4-5. Consensus sequence for ancestral virus collected from wild-caught *Drosophila melanogaster*.

1-50	GACACCAAGACCGAGUUCGUGGUCCGAACCCCUGAAUUUACAAGUCCGGU
51-100	AGAGUAUCCGUUCACCUGGUUUACAAACAAUAAGACAAAGCCACUAUUCA
101-150	AGAUCUCUGUGAUGGCUGACUGUAGCCUGGAAACGGCCAGAGUCGCAGCU
151-200	UUGGAGUUCCUUAUGGGAGAGAAAGUUCCUCCAUCUCAUGUGAUCGAUUA
201-250	CCUUUACGAAUUUUGUAAGACUAUGACUCAGGAAUUAGAUACCAACUGGG
251-300	AGAGUUACGGACAGGUCAUCGGUAAGAAGGGUGAUACAGUUACUCCUAUU
301-350	AAUCUCCUCCACGUCAUGAUAUCUCAAGACACGCGCAAGUACACACCAAA
351-400	AAAUCCGCGCUUACUGACCGAUGAGGUUGAUAUCUAUUUGGUGGGUAACC
401-450	UCCUUUGCUCAUAUAGGUAUAACAAGACACAUGAGAAAAUGCAAACGGCU
451-500	UACGGCACCAAGGUCGCUUCGAUCCUCGCUCCAUUUUCAACGCAUGAUAC
501-550	ACAAAACAUCCGAAUCUCUACAUUCUUGACUAACUCCAAAUCACUUGUAG
551-600	ACCAUCCAAACUUUGAGAUAAUGGCAUCUGCUAUUGAUAUGUUUUUAGAA
601-650	AGAUUCCCCUCGCAUGGAAUGGGAAAGCUGCGCUUUGGAACCAUCGGAUU
651-700	AAGGUACCAAGGGUGCUCUGGUCUUGUGGACUUAACCUACCUUAAACAAA
701-750	UACUUAAAAAGUCCGGAUUAGCAGACGCAAUAAAAUGGCUAUUUGCAGGG
751-800	UGCCUAGUGUCUGAAGUAUUCCAGAUGAUGAUCAAUCAUCAGGAUGAAAU
801-850	UGAUGUGAAACACUCUUACUUCCCAUACAUGAUCGGAUUCAAGAUAAGCA
851-900	ACAAGAGUCCUUUCUCUGCCGGGAGUAACCCCUCAGUCCACACCCUUGUC
901-950	CACCUGAUCGGUUCCUUGCUUGGAUCACCAAGGUCAAUCAAUGCGAUCAU
951-1000	GAUUGAAUAUGGUGUCAUAAGUGACAUUGUGGUGAAUGCAGCAAUAGUUU
1001-1050	UCUUUGCACACCGCAGCGAUAUCGGUGCUAAAGUUCGGUUUGGGGAGAAA
1051-1100	GGUGUGAUGGACGAGAUUCACCGGGAAGAAGAUAUGGCCAGACAAAGAAA
1101-1150	GAGAUCUGCUACUGCAGGCGGACUGCCGGCCACGCAACCUACCGACCCAC
1151-1200	GUGGUUGGCUAGCGGAAUAUCAAGAUAGAGAUUAUAAAUUCACACAGUCA
1201-1250	GAGGUGGAUGAUCUAAACAAAGUGAUUCAAGGUAUACCACAGGUACGCGC
1251-1300	CGGCACGGUGGGAGAGUGGGUGAAGUUGAACUUCAUCUCAAACCUUCCGA
1301-1350	AAGGGUUUGAUGUAACUCGGGUGUGACAGCUCAUGAAAAAAAUACAAAUA
1351-1400	AUCAAGAUAAAAUACAACUCCAAAGAGCCAACAUCACCAUCACACACCAC
1401-1450	AGCCAGGACUUGGAAAAGCACUAACAACAAAGAUGAAUCUGACUCAGGAG
1451-1500	GAUGAACGAAGGCUUGCACUACUACGUGAGAGAGCCGCUGCCUGGGAUCA
1501-1550	UGCAAAUGCCUUGGGUGAGUUUGAAAGCAAUGCAGAAUUGGAUGUAUACC
1551-1600	AUGAAGGUAUUGCGUAUCCCCCCUCGAUCGACGAGAUGGAGGUAGGUGAG
1601-1650	GAUAGACCCACUUUAGACAAGAUCUUGGAAUCCUCCAUAGGCUCGGAAUC
1651-1700	CUUACCGAAUCUAGAGGAUGGACCACCAGAUCAAGAGGCAGCCUACCAUG
1701-1750	AAGGUGAGAUGUAUUCGACGAUGACAACCAGCUACAUGGACCAGGGUAAU

Table 4-5. Continued.

1751-1800	GAGAGCAUGGACCUAAGUGCAGGACCAACCUGCUCUACUCCUCAACCUAU
1801-1850	UACCGGGAUCCAUCAUAAAGAGACCAUCAAUGGUCGGGUUACCCAGAUGG
1851-1900	UUUACAUCGGUGAAACCUCUGAUAUGGCACUCAUCAAUAGAUUGCAAGAU
1901-1950	GAAUUUGAUGUUUUGUACAACAUGUGUCCGAAUCGUAACAGGCGGGGAGA
1951-2000	GAUUGUACACUUGAUAGUGAUGUGGAGCACCCCUCUUAAACCGAGCACUC
2001-2050	GCUAUUCGUUAUCGCCAGCCCAACCCCCCGGCAAGAAGAGGAAGAUUGUU
2051-2100	CAACACUCAUCAGACCCAGAGCCAGAGAUAGUAUUUGUCCCGGAACCCAC
2101-2150	UUCAUCAGAACAGUCAUUACCUGAGCGUCAGAUAAAAGCGUAUCUCCAGC
2151-2200	GGGGGAUAGAACUUAAGGGGAAGUCGGCGGAAGUUCCCAACUUCAAGAUC
2201-2250	ACCAACGAGACAUUGGGUUUCUCUGAUGCGAACCUUAAGUUAUUAUACCC
2251-2300	CGACACCUCUUUGUAUCCGGACAACCCCGUGUUUAUAUUAGAAGAUGUGU
2301-2350	UCACAUCCACCAGACAAUUAAAUAAGAUCAAGGUAAAUUACCAGUUGUAU
2351-2400	GAACCGGUACUCCCCUCUAUAUAGGGGAGCGACCCCAUCCAUGCACCGUG
2401-2450	CUUUACCAUGAAAAAAAUGUCUCAACAACAAAUCAAAAUUCCCUCCACAA
2451-2500	UGUGCCACAGAACCUCAGUUCUACCCUUAGUCCUUUCCUUUUUCCAUCUA
2501-2550	UAUUAUUAUGGACAUGUUGCAUCUGUGUCUAUCCCGUUAUCAAUAAAUAU
2551-2600	AACAGACAGGAUAUUGCCAGUACAAUUGAACACUACAAUCUUGAUGGAGG
2601-2650	GUGCAGAAGAGCUGAAUAUUACUCAGCAGUUAAAUGUGCAGCACUUCCAA
2651-2700	GAGGGCCGGCUCACACGGGAGGAAAAUUUUGUCUGUAGUCACAUGACAAC
2701-2750	AACAGAUCUAGCCACUGAUUUCUUCCAACUUUCAAAACUAGCCGGAGCUU
2751-2800	AUGAUGCCGCGGGUAGCCAAGACCAAAGAGUGUACUCUCCCUCUUCCAUC
2801-2850	UCCACGUUGGCCCCAAUAUUUCCUGAAACCACACCUGCAACCCUCAGCCC
2851-2900	UGAUGGAGUGACGCUAAACAUCAGGACCUCGGUGAUGGAUAUUAACCCAU
2901-2950	UGACUCGCUUCACCAAAGCAGUUAAAGAUCUUAGGGAAGACAUUACACCA
2951-3000	GGGUAUGAUCUGACCAACGGAAGGAAGUUUGGUCGGGAGUUUUUAAGAGC
3001-3050	CCUUGAGGAGAAUGGGGUACGAAGCAGGCACAAGAGGGAUUCAACUACUG
3051-3100	AUAUGCUGCAAAUGACACUGGAUGUGACUGAGUCUAGAGAAAAAUACGUU
3101-3150	GAAAUCAGGGACCAGUGUGCAAGCGUGCUAGGUGCUCUGAUCAAGAGUAU
3151-3200	GGUUCCAUCAGUGUCGGAACAUUGCUUAUAUCACUAUCACAUUAAGAAUA
3201-3250	UAGUAAAUUGUUUCACUAGUCAUUACAUCAGACUCCACAAUGAGCCAGAU
3251-3300	UUCCCGUUGGUUAUGGCAUUAUACUCUGAGUACCUAUCACUCACAGUGAA
3301-3350	GAUACAUUACAUGGACGAGAUUUUAGUCAGCACCCUAAGGGCAUGAAAAA
3351-3400	AACUAGCAACAUCAAUUACAAACAUCACUGACAAAUUCUCUGUGAAACCG
3401-3450	CCAUGAAUAAGAUGAACCAAUUGGUCAGGUUUGUUAAAGACACCGUUGCG
3451-3500	GUGAGAAAACCACAGAGUGAAGAUCAAUUAUAUCUCCCUAUCCCCAGCAC
3501-3550	GAUCGGAGGACAUGAGGUGGACUCACCAUUCGCCGAGCCGACCGCCCCGA
3551-3600	CACUAGGAAUCAUCCAAUCCAAGUGCAAAAGGGCCGACUGGCUGAUCAAG

Table 4-5. Continued.

3601-3650	UCGCAUCUCACGAUCACAACUAAUUAUGAGAUAAAAGAGUGGGAGACCUG
3651-3700	GGACAGAGCAAUUAGUGAUAUUCUGGAUCUGUAUGAUGGCAAUCCAGUAU
3701-3750	UCAAGCCUAUUUUAUUGUUCGUGUAUUAUGUAUUAGCCUACAACGCCCGG
3751-3800	AAAAUUCCUGGCCCAUCCAACGGGGUAAGGUAUGGGGCAUAUUUUGACGA
3801-3850	ACUAACUACUGUAUGGCAUGCCAUCCCGGAACUAAUGAACCAGGAGACUG
3851-3900	AUUAUUCAUACAAUCAUCGGGUAAUCCAUAGAAAGAUCCAGUAUGUCAUC
3901-3950	UCAUUUAAGAUUCAGAUGUCAUCUACUAAGCGGCGAACUUCACCUAUCGA
3951-4000	GUCCUUCAUUGAGGUUACAUCUGAAGGACUUAAACACACUCCACAGUUUA
4001-4050	CCACCAUUCUAGAUCGUGCACGGUUUGUCUAUUCGUUGACAGGAGGGCGU
4051-4100	UAUGUUAUCCAUCCUUUCUAAAACUAACCUACGCAGUAACAACAGAUUAU
4101-4150	UCCAUCUGCAUGAAAAAAACUCCUUCUGCUGAAAUGGUACAUUACGAGAC
4151-4200	UCAUAUUCUAUUCAUCCAUCUCUGGAUGUUGGCACUGAUAUUUAUCACCA
4201-4250	CCUCAGUAUGGUUGGCCGCCAGCCAAAAAACUUUUACCCCGGAUUUGGUA
4251-4300	UUCCCUGAAAUGAAUCGUAAUAGUUCUUGGAGUGUGGCAAACUAUGGAGA
4301-4350	GAUAUUAUGCCCAACAUCCUUUCAGUCAUACGACCCCAAGAAGCACCAGA
4351-4400	UCCUGACACGUGUCCUAGUUGAGAGACCAAGCCUAAACACUGACACAAAG
4401-4450	GUGGAGGGCUACACAUGCCAUAAGGUGAAGUAUGAGACCAUCUGUGAUAU
4451-4500	GCCAUGGUAUUUCUCCCCUACCAUUUCACACUCCAUCUCUCCGCUUAGGG
4501-4550	UAAAAGAGUCUGAGUGUAAAGACGCUAUAGCAGAGCACCAAUUGGGCACC
4551-4600	CAUGUCUCGCUGAGCUUUCCCCCUGAGGACUGUUCUUGGAACUCAGUGAA
4601-4650	CACGAAGGCGUACGAGGAUAUUAUAGUUAAAGACCAUCCUGUAAUGUUGG
4651-4700	AUCCAUACACUAAUAAUUAUGUUGACGCAAUAUUUCCAGGUGGUAUUUCA
4701-4750	AGCCCAGGGAUGGGGGGGACCAUUCAUGACGAUAUGAUGUGGGUAUCCAA
4751-4800	GGAUCUGGCAGUGAGCCCAGAAUGCUCUGGGUGGCAACAAAGUAUGGGGC
4801-4850	UUAUUUAUUCAUCUCGGCUGUAUGGAGAGCGGGAGCCGAUGCUUGAAGUA
4851-4900	GGCUCCAUCCACAUCGAGGGCCACAGGGACAAGAACCUCACUUUAGCCUG
4901-4950	CCGGAUUUCCUUCUGUGGUGAGAUAGGGGUGAGGUUUCAUGAUGGGGAGU
4951-5000	GGAUGAAGGUAUCUGUAAAUCUUGACCAUCCUAACUCUGUGACAUUCCAA
5001-5050	GUAACUGAUUUCCCUCCGUGUCCUCCCGGUACCACUAUCCAAACUGCCGU
5051-5100	UGUGGAAAAUAUAAACCCAGAGAUUCAAGAGCUUACUGUCAACAUGAUGU
5101-5150	ACAGAUUGAAGUGUCAAGAGACCAUCUCAAAAAUGGUAUCUGGGCUUCCG
5151-5200	ACCUCUGCCCUUGAUCUCUCCUAUUUGAUUCAAGUGCAGGAGGGACCUGG
5201-5250	UAUUGUUUAUAAGAGAGAGAAGGGUGUCUUGUACCAAAGUGUCGGAAUGU
5251-5300	ACCAAUACAUAGACACAGUAACACUGAACAAGGAGGAGAACCAACUGGGU
5301-5350	GAAAACGCGAGGGGGCAAAAGGUCUUUUGGACCGAAUGGAGCGAUUCACC
5351-5400	GACACGCCCCGACCUCCAGGAAGGGAUCAAUGGGAUUGUUAAAUAUGAGG
5401-5450	GGCAGAUAAGAGUGCCUUUAGGAAUGUCUCUGAGAUUAGAGGCUGCAACA

Table 4-5. Continued.

5451-5500	GAAUUAAUGUGGGGACACCCUGUCCACACCGUGUCUCACCCUAUACUUCA
5501-5550	UGUUAUCAGUAAUCACACAGAACAGUCUGUAACAACUUGGAACAGGGGUG
5551-5600	UUAAUUCCACAAACCUGAUAGGCCUGGCAACUCGAUCAAUAUCAGGUUUC
5601-5650	UACAAUGAUCUGAAGUUGUAUUUGAUCUUGGCACUGAUAGUCGUUUCAAU
5651-5700	UGUCGCACUGGUUGUCUUAGAUGUGAUCCCCUUUAAGUAUAUUUUAUUUA
5701-5750	UACUAUGCCCACCUCUGCUGUUGUGUAGGUUCAUAAAAUGUUCACGAAGG
5751-5800	AGGCCUGAAACUGGGGACAGGUAUCACGUAGAAUAUAACCGACCUGGACA
5801-5850	AGUGUCCAGCGCGUUCUAAGGUGGUCAACAUGGCCAUGAAAAAAACAACA
5851-5900	CCCAUUAAGCUGUGACAACUCCAGGUCAUCUUGCAAUUAUCAUGGAUUUU
5901-5950	GAAAUAGAAGAUCCAUAUGAUCCGUUCUCUAUGGACACAUAUUUGGACCC
5951-6000	ACAAGACCCAUCCUUCGGGGAUCUUGAGUCAAUGAGGCAUUUAAGUAAUG
6001-6050	UUGACUACUCUCUGAAUUCUCCGAUGAUUGCAGACGAGUUGGAGGCGUUC
6051-6100	AUUCGUUGGUUGCAAUGCGGAUGCACUGACCCCCGAUGGAACGAGGAUCG
6101-6150	AUGGGUCCGGACGAAACAAGGACUGUUCUCUGGGCAAUCUCCUACAACAA
6151-6200	UAGAAGGUGCAGCUACUUUCACCGGAUGGUUUGGAAACUUUAAUCUAAAA
6201-6250	CGCAGGUACUAUAUUGUGAGACAGUUUAAAAUGAUACUUGAAAAAGCUCA
6251-6300	GGCAGACUCAGAAGAGACCAAACCUGUCGUGGAUGCAUUCCUACGGGGAU
6301-6350	GGAUCAAUCACAAAGGAGUCACCCUGACAUCUAAGAUCACCCUGCCAGAG
6351-6400	GAAGAGUUGAAAUGGGGAUAUUAUUUUUGGGAAUUGCACAUAGUGACUCU
6401-6450	CCACCUAAAUUGUACAACCGACCAAGAGAGGACCCACUUGAUCAAAAGCU
6451-6500	UCAAAUCCAAGUCUCGUGGGUUGCCCGAUGUGUUCGAUUUCACAUUGUAC
6501-6550	ACGAGAAACUUCGGGCCCUUGAGUAUAGCAGGGGGGUAUGUUUACAUGUU
6551-6600	UGAUCACAACAGGAUGCUGGACCGCAAUGCAAUUCUGAUGAUGAAGGAUA
6601-6650	CCUACGUUGCUCGGUUCAACUCAUUCCUUGCACUAAGCAAUAGGGCAGAC
6651-6700	UGUGUCUUCCCUGAAGACGCCCAUUACCGACUACAAAUGCUAUAUGAGAU
6701-6750	AGGGGACAUGGUCUUGGAUGAGGGCGGCACAUCUGGAUACAAUGGUCUGA
6751-6800	AGUUGCUGGAGGCUAUGUGCAGUAGCAGGAUUACAGACUUGGCACAGUCU
6801-6850	AAGAAGCCUCUAAUACCUGACUUUCCUGACUUCCGUUCCCAUGUUAGAGC
6851-6900	CAAGGUAAGAGAAGAGUCUGCGAACACACCAUCGAUUGGUAAAAUGUAUG
6901-6950	AGCUGAUUGAAGGAACAACCAGCUAUGACACUCUUCUCACAUUCUAUGGG
6951-7000	UCCUUUCGUCACUGGGGACACCCUUAUAUAAAUUACUUAGCCGGGUUAGA
7001-7050	GAAAUUGUACAUCCAGACGACGGUGGAGAAAGAGAUUGACCAGGAGUACG
7051-7100	UAGAGAAAUUGGCCAGUGAUCUCGCGUUCCUUGUUAUCCAGGAUAGGUUU
7101-7150	CGUAAAACAAAGAAAUGGCCCGUUGAUCCUCUUCUAAUAGAUAAGGACCA
7151-7200	UCCUCUGGUAGAGUAUAUCCGGACAUCAUCAUGGCCCAACAACUCCAUCA
7201-7250	UUAAAAAUUUUGGGGAUGGAUGGCACACAUUACCUUUGACUAAAUGUUAU
7251-7300	GACAUCCCUGACGUGAUUGAUCCAUCACUCUUGUAUUCAGACAAGAGUCA

Table 4-5. Continued.

7301-7350	UUCUAUGACGAGAUCGGAAGUCAGGGGAUGGAUGACAUCACACCCUGGGA
7351-7400	AACCGAUCCCCUCCCGGAAAGUGCUGUCAACUCUCUUAAACUCCCCAUCG
7401-7450	ACGAACUGGCCCGUGUUCUUACAACAGGUGAACGAUUCAGGCAUACCCAU
7451-7500	UGAACAGCUUAUUAUCGGAUUGAUGGCAAAAGAAAGAGAACAGAAAAUAG
7501-7550	ACGGGAGGUUCUUCUCUUUAAUGUCUUGGGACAUCCGCGACUAUUUUGUC
7551-7600	AUGACAGAAUACCUCAUUAAGACCCACUUUGUCCCGUUAUUCAAAGGGCU
7601-7650	GACCAUGGCAGAUGACCUCACAACAGUGAUUGGUAAGAUAUUGGAGAACU
7651-7700	CCCGGGGUCAAGGAGAAGCAGACUAUGAGAACUUGACUAUCACUGAUCAC
7701-7750	AUUGACUAUGAAAAAUGGAACAACCAUCAGAGAGGGGAAGCAAAUAAUCC
7751-7800	AAUUUUUUUGGUCAUGGGGAAGUUUUUAGGUUACCCGCAUCUCAUAGAAC
7801-7850	GGACCCAUGAGAUAUUUGAGAAAUCAUGGAUCUAUUAUCUAAAUAGGGCU
7851-7900	GACCUGAUGGACUUCGAUGGAGAGGGCAAUCUGAUGAAUAGAACCGAACU
7901-7950	CCGGGUCUGCUGGAAUGGGCAGAAGGGUGGACUGGAGGGGCUGAGACAAA
7951-8000	AGGGAUGGAGCAUUUGCAACUUGCUAGUAUUGAGGAGAGAGAGCCUAGCU
8001-8050	ACUAACACAGUGGUUAAAACGCUCGCCCAAGGAGAUAAUCAGGUAUUGAG
8051-8100	UUCUAGAUAUAGGAUACGCACAUCACGAGAUCAGAAUCAGUUGCAGUCGA
8101-8150	ACAUUGAGGAUAUCUGUAGAAAUAAUCGAAGCUUAAUGGAGAGAAUUAGG
8151-8200	AUAGGGACAGGGAAACUAGGACUCAUUAUAAAUCAUGAUGAGACCAUCAA
8201-8250	AAGCACCGAGUACAUGAAUUACGGGAAGACUUGUGUCAUCCAUGGUAAUA
8251-8300	UUAGAAACCUUGAGACAAAGAGAUGGUCUAGAGUAACCUGUGUAACUAAU
8301-8350	GAUCAGUUACCUACCCUAUCAAAUGUCAUGGCCACGAUUGGAAGCAAUGC
8351-8400	CCUAACAGUCUCCCACUACUCUGACUCCCCGAUCAAUUCGAUGUAUCAUU
8401-8450	AUAAUUUUAUGGGUAAUUUUGUGAGGAUCAUGAAUGAGAUACACAACCCA
8451-8500	GCAUUGCGUGGACCAGUGUCCUCUAUAGAAGGGGUAACCGGACAAUCAUU
8501-8550	CAGCCGACUGUCUUACUUAUUGGCGGUCUUGUACUUAGACCCCUCUAUGG
8551-8600	GAGGUGCAUGUGGUAUGUCUUUAACCAGGUUCCUGAUUAGGAUGUUCCCG
8601-8650	GACCCGAUAACAGAAAGUCUUACUUUCUUGAGGAUCGUAGCGAUGAAUGU
8651-8700	UCACUCUGAUGAGGUCCGACAGACAUUUAUCCAGUUUGGUAAUCCUAAAC
8701-8750	UGAAGACAUUCUCACCGGAAGACUUAUCUAAGCUUUUAGAAGACCCUCUC
8751-8800	UCUAUCAACGUACCGAAAGGUUUGAGUGCUACUAAUCUGAUAAAGGAUGC
8801-8850	UAUAAAACUGUCGCUACACAGAUCAGUUGACGAAAUUGCCAAUGAAAUCA
8851-8900	UUGCAGAAGCUGUGAUCCACCAAAAAGAUCAUGAAGAAGGAUUCCUCAUG
8901-8950	CAUCUGACACAGAUCAGUCCCUUGUUCCCGAGAUUUUUGAGUGAAUAUAA
8951-9000	AGCUGGAACUUAUUUGGGCAUAGCUGAGGGUCUAAUCGGAUUAUUUCAAA
9001-9050	ACUCUAAAACCAUCAGGAACCAAUUUAGGAGGAACCUCGAUAUUGGAUAU
9051-9100	GACAGUAUUGUGAUUAAGUCUGAGAUUGCAACGAUUAGAGACUUGACUGG
9101-9150	GUACCGAUUAGAGGAUGCUGAAAGAGUGGAGAUGUGGCCAUGUUCAUCCA

Table 4-5. Continued.

9151-9200	CGCAAGCAGAUUACUUGAGGAGAGUGUCAUGGCAGCAAGUGGUUUAUGGG
9201-9250	GCGACAAUACCUCACCCUGCCGAGUUAUUCGGACUUCCUCUAAGGGCCGC
9251-9300	CCCGACUUGCCCAAAUUGCACUACAACUUUCCCGAUGAACCUGUACAUCU
9301-9350	CUGUAUUGAUUCCACUUGGGUUCAAGGGUUUGAAAGAUACCCGAGGAACA
9351-9400	UGUGUGGCAUAUUUGGGAUCUAGUACGACAGAAUCCACAGGUAUCGUUAA
9401-9450	UCCAUGGGAGAAGGAGGCGGUUGUGCCAGUGAUCAAAAGGGCAGCAUCGC
9451-9500	UGAGAAAUGGGAUAGGAUGGUUCAUUGAACCUGGAUCAAACCUGGCACAG
9501-9550	AGUAUUUUAAACAACUUGCAAUCUUUGACAGGCGAGUCCUGGAGCCAAAA
9551-9600	CUCCGGUGGGGUGAGACGAACUGGAUCUGCAUUGCAUCGAUUCAGUUGUU
9601-9650	CUCGGCAGAGUGGUGGUGGUUACACAGCUCAGAACCCGUCGAAGUUGACA
9651-9700	AGGAUGAUUGCCACGACCAAUUAUCUUGCUGAUCUGGGAGAUGAGAACUA
9701-9750	UGAUUUCAUGUACCAGAGCUGCUUGCUCAACGCUUUAAUAUCAGUGGGGG
9751-9800	AGAUCCACCCUGUAGAUGGGUCUCAAGGUUAUUAUCAUCAACACGUCAAU
9801-9850	UGUACGUCCUGCCUCAGACCUAUUAAGGAGGUGACACUGGAAAGCCCUGC
9851-9900	ACCUUAUUCCCAUACCAUAACAUCCAAUCUUCUGGACAAAUGGAAGCCAG
9901-9950	AUGGCUCGAAGUGGUCAGUCUCCCGACCCUCUAUUCCAUUGAGAUCAGGG
9951-10000	AAGUGGGAAUGUGUGUCUCAUGAUCGUCAAUCCUAUCAUGUUGGCUUCAU
10001-10050	ACAAGGGUUCAUUUAUGGAGAUUCUGUAUGGGGAAUCCGAUCUAUGGCUG
10051-10100	AUGAUCCCGCAUUAUUCCCAUUGAGUUUCCGGAAUAAGGUUAACCCUCGA
10101-10150	GCCUACCUACUGGGAAUCUUGCACGGACUACUUCGCAGCUGCACGGUUUC
10151-10200	GGUUGUCCAUCAACGAUGUUUUCGAAGCAGUAGAGCCGUGAAACAAACCA
10201-10250	CCCUCGGACUCUGUAGCAUGACUGUGAGUAGGCUGGUUCAAAACGAUGGA
10251-10300	UUUCUAAACAUACUCCGGGAUGAGCAGUUUACGGCAGUUUUCAGAUCGAU
10301-10350	CCCACACAGGAUACCUCCUUCUUAUCCGAUGGUGACCCAAGACAUAGGGG
10351-10400	ACUUGGCAUCAAACUACCUAAAGCGCCAGCUGAUGACUGAAGGGCUGGCC
10401-10450	UACUUCAGGUCUGUCAAAUCGGGAUCUGGAAAUGAAGCAUGGAUCUUUGC
10451-10500	UGAUGCUAACCACCCUAUGAUCGUGAGCCUAAUAACAGUUAGUGGGCUAA
10501-10550	UGACCAAAAUUUUAGCAAAGGACAUAUGGCAGAAGAGAGACCUGGAAGAA
10551-10600	UUGAAGGAGCUUCGAAGCUUGAGCGUACAGGCGCGAGAACAGGAUGACCC
10601-10650	GCAUACUGACAUCACCGCCAUGGAGUCACUCGCCGCUGAGUGGGUUGUUU
10651-10700	GCUCUGAUCAGGAGACCAGACACGCAGUAAAGUAUAACACUCGAAUAGAG
10701-10750	UCCGUAACAGACAGCGGGCUGACAGUGACAUGGGGUGAUGAAUACACUUC
10751-10800	AUCUGCUGAUUUCGUAACAGUGGUGUUCUCUGUAGAAGAAAUCCGUACUC
10801-10850	CACCGGGUAUGCUGAUCCCCAGGAUUCAAAACCCUUUAAUAUCUGGCCUA
10851-10900	AGAACAGCACAAAUUGCAACUGGCUCUCAUUACAAACUAAGGAGUAUUCU
10901-10950	UUCAAAACUCCGCCUCAAUGUAAGGGGCGCCCUAGUGGGAGGAGACGGGU
10951-11000	CAGGGGGGCUGACUGCCUUAGUGUGUAGAAUGUACCCCACUAGCAGAGUA

Table 4-5. Continued.

11001-11050	AUCUUUAACAGUAUAUGCGAUUUCUCAGAUGUACGCCUGAAGGGCACAAC
11051-11100	UCCUGCACCUCCGUCUGCCCUCUCUCACUCAUUGAAUGAUUGUACCCAAG
11101-11150	UUGUGAAUUAUUCCGAUAGCUGGGCACAUCCUAGCGAUCUUACUGACACA
11151-11200	AAGACAUGGAAAUACUUUGUGGACAUAACUAAAUCUAAAUCCAUACAAGU
11201-11250	GGAUCUGAUAAUACUGGAUAUGGAAGUAGUAGACGAGAGCUCAAUAUCCA
11251-11300	AGAUUGAAGACAACUUGAUGAGGUAUGGACCUCAACUGCUGACAAGAGAU
11301-11350	GGAGUGAUCCUAUUUAAAACGUACCUCACCAGGAUAUUCAAGGCCCAGGA
11351-11400	GAUGAUAUUGACCAAAUGUGGACAUGUCUUCUCUCAGGUCGAGCUAUGGU
11401-11450	ACUCAGAUCUAUCUUCAUCCCAAACAUCAGAGGUUUACGUCUUGAUGUCA
11451-11500	GGGCAGACAAAACUCUCCCACCUUCAGGUGCGUAAGCCGGAUCUGAUGCA
11501-11550	AUUGAGGACAGAUGUCUCAUCCUUCCCUGUCUUUGCACCUCCGAUAGUGG
11551-11600	AAUUCAAGCGGGCCAGAAAGGUCGCCGGGUUGGAUUUGACAGUAGGGGUA
11601-11650	CCCCCUCUACUACUGCCAGAACCAGGCCCUGAAAUGAUUAACCUCUUAUC
11651-11700	AUCAUUAGGAGUUCGCCCUGAUAUAUCAUUUUCGAUCACUAGAUCAUUUG
11701-11750	GGGACAACAUGUCAACAGAAUUCCUACCUCUACACUUGUUUCUACUAGCG
11751-11800	UUGAAUGGGAUAUACGAUGUCACAACCGGCUACAGGACCCCUCCUGGCCC
11801-11850	UCCUUCCGACCAAGUGUGUAUAAAGUUGGGCAUCUGGAUAGUGGGAUAUA
11851-11900	GGAUAUGGUCAGGAUAUGUCCAAGAUUCAUACGCAAAAGCCUCAUUCGGU
11901-11950	CAAAAACUCAUUGAUAACUUUGUUCCGUUCAACUUUUGGAACAAACAGAU
11951-12000	UGGACGAUCUUGGUUCCCGCAUUGGUCACUUAUUAAACCACAGACAUAUA
12001-12050	CGAAAAAUAUACAAUUGGAUUCUGAGAUGGCCCAUAUUGGUUCAGUGAUA
12051-12100	CGUAUCCUGCAUCGAAAUUUCCCCAAAGCCAGAAGAAACCCUCCCGGUGA
12101-12150	UUUGGUCGAUAAUAACUGUUCUCGAAUCAACAAGGGGAUAUCCUGUCUCA
12151-12200	ACAUGGACAUGAAAACGGGGAUAUUACGGUGGAUGGGGGAUGGGCGAGAC
12201-12250	CUGUCUGGAGUUAAUGUCACCACCCGGAUCUCAAACUUCUAUGUCACCCA
12251-12300	AGAUGAUCAAACCGCUAGCUUCAGAAAUUAAUGGACUCCUCGUCCCUUUU
12301-12350	GCCUCUCAUGAAAAAAACAGUUCUGGAAGCACUAUCCAUGCAAUCGAUUC
12351-12400	CGCAGGCAUCACUGAAUAACAGAUUCCGCGUACAAAAUUGAAUCAGU

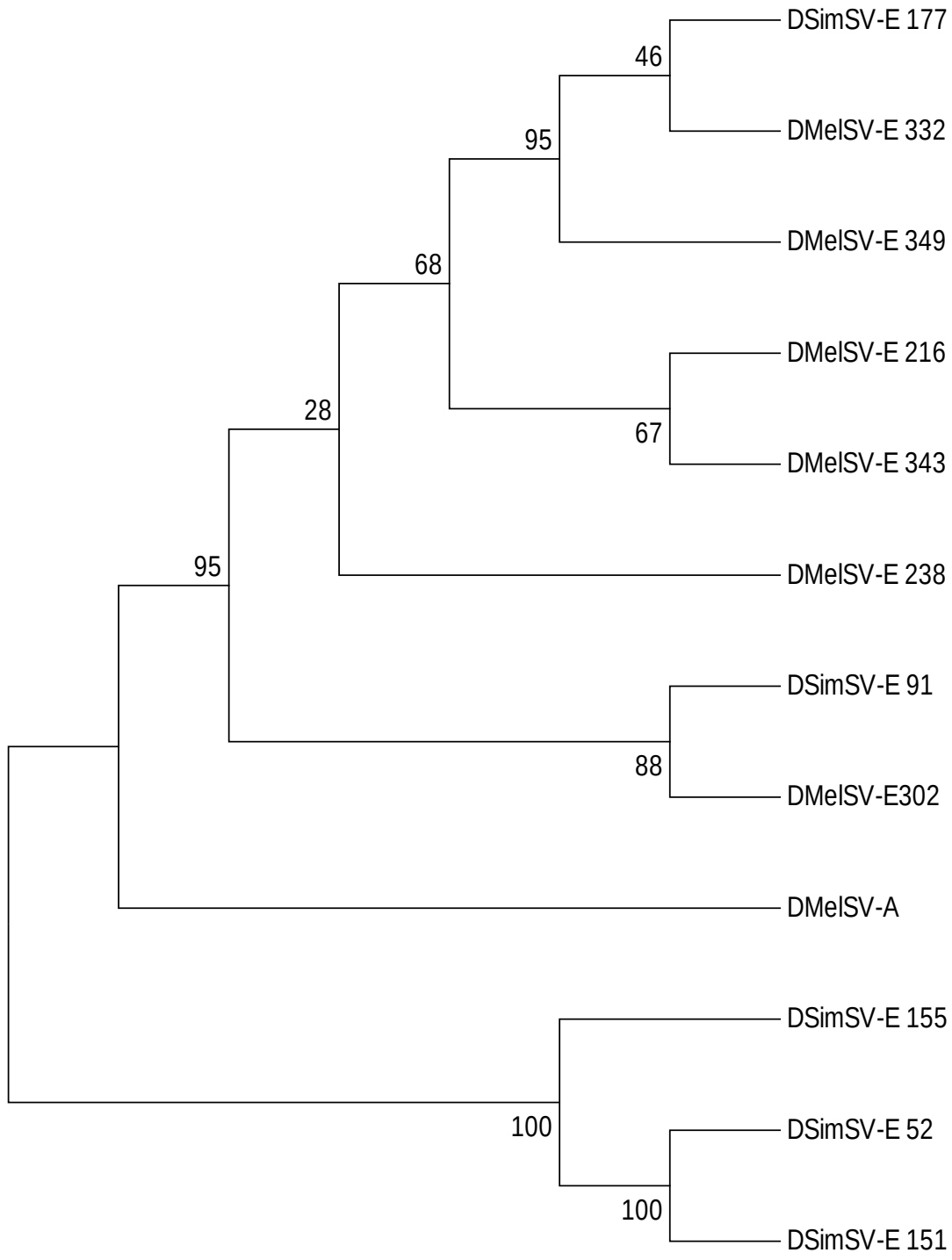

Figure 4-1. Maximum likelihood phylogenetic relationship among ancestral sigma virus from *Drosophila melanogaster* (DMelSV-A) and sigma virus evolved in naive native hosts (DMelSV-E) and in novel *D. simulans* host (DSimSV-E) lineages. Bootstrap values are indicated at the branches.

CHAPTER 5
TESTING ATTENUATION BY CROSS-INFECTING NAÏVE NATIVE AND NOVEL HOSTS WITH EXPERIMENTALLY-EVOLVED VIRUS

Background

Interspecific host shifts are a major source of zoonosis in humans. Recent examples of shifts into humans include HIV, SARS and influenza (CLAAS *et al.* 1998; GAO *et al.* 1992; HOLMES and RAMBAUT 2004). The debate continues as to what biological processes govern the evolution of virulence; it is clear that a "one size fits all" theory is unlikely to exist (ALIZON *et al.* 2009; ANDRE and HOCHBERG 2005; EBERT 1998). Thus, we need to continue exploring the evolution of virulence under different conditions and in natural as well as model systems.

The evolution of virulence following host shifts has been studied primarily in serial passage experiments (SPEs) (EBERT 1998). Two major conclusions from SPE studies are that pathogen virulence increases in the novel host (as observed in Chapter 3) and that this increase in virulence in the novel host is accompanied by attenuation in the native host (EBERT 1998). Attenuation is defined as a reduction or even elimination of the ability of the evolved pathogen to re-infect, replicate and induce virulence on is native host. Since pathogen success is guaranteed under SPE regimens at every generation, which could influence the experimental outcomes (EBERT 1998), and the majority of SPE experiments involve single-celled novel hosts, the degree to which SPE findings will be replicated in natural pathogen evolution in a complex multicellular host is unknown.

The *Drosophila*/sigma system provides a testing ground for the hypothesis that attenuation to the native host occurs following a shift to a novel host in a complex, multicellular host. *Drosophila melanogaster* is the native host of this virus and is an

excellent model organism with many genetic and molecular tools available (SCHNEIDER 2000). The DMelSV rhabdovirus of *D. melanogaster* (DMelSV) is an RNA virus that is very similar to the causative agents of many human disorders, such a rabies (ROSE and WHITT 2001). DMelSV is readily passed from parent to offspring (FLEURIET 1988) and can be successfully shifted onto new host species (As in Chapter 3; LONGDON *et al.* 2011). This virus/host system thus eliminates the need for experimenter-assisted passage across host generations.

In a previous study (Chapter 3), DMelSV was artificially host shifted onto a novel host (*D*. *simulans*). Many infected novel host lines were produced. The first three generations after injection underwent selection for transmission efficiency. After this, the virus was passaged naturally (without experimenter selection) for 13 host generations. While transmission efficiency declined in many of these infected host lines, four sustained a steady infection with high DSimSV titers and infectivity (Table 1 and Chapter 3). High titer was defined as at least 5.0×10^5 viral genome copies per microgram of purified host RNA. We regarded the virus strains in these fly lines as well-adapted to their novel host. Here we tested whether attenuation would occur in these four strains by injecting the *D. simulans*–evolved (DSimSV-E) virus strains, as well as the ancestral, unselected isolate from *D. melanogaster*, into naïve *D*. *melanogaste*r and *D. simulans* hosts. We tracked infection success, transmissibility and virulence across three host generations.

Materials and Methods

Naïve Hosts

We infected one effectively isogenic line of *D*. *simulans* (*MD106;* www.dpgp.org) and one highly inbred line of *D*. *melanogaster* (*Diallel 75*; YANG and NUZHDIN 2003) with

DSimSV-E. *MD106* is one of the effectively isogenic lines included in the *Drosophila* Population Genomics Project (www.dpgp.org). *Diallel 75* is a highly inbred line (40 generations of full sib inbreeding; YANG and NUZHDIN 2003). These lines were used previously to generate the virus evolved in *D*. *melanogaster* and *D*. *simulans* (Chapter 3).

Inocula and Injections

Inocula were prepared from four DSimSV-E-infected fly lines exhibiting high virus titers and infectivity of *D*. *simulans*-evolved virus (Chapter 3; see Table 1 for titers and inoculum names). These viral genotypes were the result of 16 generations of experimental evolution (first 3 generations with selection for transmission efficiency; subsequent 13 generations under relaxed selection) in novel (*D*. *simulans*) hosts. Each inoculum was prepared from ~200 females from each evolved line as in Chapter 3.

The *D. melanogaster* and *D. simulans* flies (1-2 days old females) were injected as described previously (Chapter 3). Briefly, a Sutter instruments XenoWorks™ Analog Microinjector equipped with a 100 µl syringe and needles made from Narishige GD-1 blanks were used. Each inoculum was thawed individually on ice and 0.1 µl was injected into the abdomens of females from each host species. In total, 720 *D*. *melanogaster* and 720 *D*. *simulans* survived the injection procedure and were placed on food. Additionally, 144 *D*. *melanogaster* and 144 *D*. *simulans* were injected with the ancestral un-evolved inoculum (DMelSV-A) harvested from wild-collected flies. Finally, 20 females from each species were “mock-injected” with homogenization buffer to control for any effects of the injection procedures. The mock-injected flies produced the uninfected control lines.

Rearing Conditions and Artificial Selection for Infection

After the injection procedure, each fly was placed in a vial with 7 ml of Jazz-Mix *Drosophila* food that was prepared following the manufacturer's instruction, and prepared with additional Baker's yeast and a strip of KimWipe®. A male fly was added to each vial. This is referred to as generation zero (G0). All flies were housed at 24°C with a 16:8 light:dark photoperiod.

Artificial selection was conducted for transmission efficiency, retaining only those lines which transmitted successfully. The offspring (G1) of the injected flies (G0) emerged and were transferred to vials of fresh food 20 days after the injection date. These G1 flies were flipped onto new vials again 10 days later to prevent them from mixing with their offspring. When the G1 flies were 25 days old (± 5 days), they were set up as 1♀ + 1♂ and females were allowed to oviposit for two days. After this oviposition period the flies were removed from the vials, and their fecundity and hatchability was assayed as described below. Finally, the parent flies' infection status was determined by CO_2 assay. The vials containing the eggs of infected females were retained, while the vials containing the eggs from uninfected females were discarded. This process of 1♀ + 1♂ setup, aging, flipping, fitness assaying and gassing was continued for three generations. At each generation only the offspring of the infected females were kept.

Fecundity and Hatchability

Egg production and per cent hatchability were determined by allowing flies to oviposit for 24 hours on standard food tinted with green food coloring (for ease of egg detection; McCormick & Co., Inc, Hunt Valley, MD) which had been poured in a thin layer onto standard glass microscope slides. Approximately 60% of each slide was covered with 1ml of the colored diet. The slides were sprayed with a 4mg/ml yeast

solution and provided to the flies for oviposition within the confines of a wide *Drosophila* vial. Subsequently, the female was gassed to determine her infection status. All of the eggs from infected flies were counted and the slides were incubated at room temperature for >24hrs. Subsequently, the eggs that failed to hatch on these same slides were counted to determine per cent hatchability. Over 3000 females were set up on slides and scored for infection but many were uninfected (did not die upon gassing). In total, 630 flies were identified as being infected and had their fecundity and hatchability assayed.

Ribonucleic Acid (RNA) Purification

Ribonucleic acid from each inoculum was isolated using TRIzol (www.invitrogen.com), following the manufacturer's directions, as described previously (Chapter 3). Briefly, a 50 µl aliquot of inoculum was mixed with 200 µl of TRIzol on ice. The mixture was vortexed briefly and incubated at room temperature (RT) for 5 min. Each sample then received 40 µl of chloroform and was vortexed for five seconds, incubated at RT for 2-3 min, then centrifuged at 12,000g for 10 min at 2°C. The RNA in the aqueous phase was precipitated with isopropanol at -20°C for 10 min. The precipitate was pelleted by centrifugation at 12,000g for 10 min at 2°C. The pellet was washed with cold (4°C) ethanol, air dried for five min at RT, resuspended in 29 µl of DEPC-treated water with 1 µl of 40 µM RNAse inhibitor (www.neb.com) and incubated at 55°C for 10 min to aid resuspension. Samples were stored at -80°C until used.

Reverse Transcription

Reverse transcription of the DMelSV virus genome was performed using the SuperScript III First-Strand cDNA synthesis system (www.invitrogen.com) following the manufacturer's recommendations. Briefly, 200ng of inoculum RNA were mixed with 1 µl

of primer (whole1F: 5' TAGAAGCATCCTCGGCTTTC 3') and 1 µl of 10mM dNTP mix and the reaction volume was brought to 10 µl with DEPC-treated water. This reaction (part 1) was incubated at 60°C for 5 min and then placed on ice. A cDNA synthesis mix (2 µl of 10X RT buffer, 4 µl of 25mM $MgCl_2$, 2 µl of 0.1mM DTT, 1 µl of 40U/ µl RNAseOUT and 1 µl of 200U/ µl SuperScript III enzyme, part 2) was prepared and added to part 1 and the mixture was incubated at 50°C for 50 min, then at 85°C for 5 min to stop the reaction and then chilled on ice. RNAse H (1 µl) was added to the tube and the synthesis reaction was incubated at 37°C for 20 min. The cDNA was stored at -20°C.

Statistics

The proportion of injected flies that became infected, and the proportion of these flies that transmitted the virus to their offspring, were evaluated using *G* tests. All of the samples were larger than five, thus no data corrections were applied. Because the residuals for infection, fecundity and hatchability data were not normal, the data were analyzed using nonparametric statistics as relevant (descriptions in text) with JMP version 10.0.

Results

Cross-Infection Success and Selection for Transmission

We injected naïve *D*. *melanogaster* and *D*. *simulans* females with the four high titer strains of sigma resulting from passage for 16 generations in *D*. *simulans* (DSimSV-E), as well as the ancestral inoculum from which they were derived (DMelSV-A; Chapter 3). Each of the five inocula was injected successfully (i.e., the injected flies survived) into 144 females from each species (720 females successfully injected). As observed previously (Chapter 3), the DMelSV-A inoculum infected *D*. *simulans* more effectively

than *D*. *melanogaster* (P < 0.0001; Figure 5-1). The DSimSV-E and the DMelSV-A inocula had similar infectivity in *D*. *simulans* (*G* = 0.8361, *P* = 0.3605; Figure 5-1). Interestingly, the DSimSV-E and the DMelSV-A inocula also infected *D*. *melanogaster* similarly (*G* = 3.15, *P* = 0.0759; Figure 5-1), suggesting that the evolved virus was not attenuated to its native host with respect to within-host replication.

Since a successful host shift requires that the pathogen not merely replicate within novel hosts, but be transmitted *between* novel hosts, we determined the transmission efficiency of the infected flies (Figure 5-2). *D*. *simulans* flies infected by DSimSV-E transmitted the virus to their offspring at significantly higher frequencies than conspecifics infected with DMelSV-A (*G* = 6.57, *P* = 0.0104). However, there was quite a bit of variation among viral strains within *D. simulans*, with two strains (DsE155 and DsE177) generally higher than DMelSV-A, while the confidence intervals of the other two strains (DsE52 and DsE91) overlapped DMelSV-A (Figure 5-2).Transmission efficiency for the *D*. *melanogaster* females that were infected with DSimSV-E was statistically indistinguishable from conspecifics infected with DMelSV-A (*G* 1.016, *P* = 0.3135).

The nature of the data (heteroscedastic residuals) precludes a straightforward parametric analysis of transmission efficiency; however, several meaningful results were clear. DSimSV-E was transmitted at higher efficiency in *D*. *simulans* than in *D*. *melanogaster,* but only at G1 (Figure 5-3). Infection rates at each generation were nearly identical for *D*. *melanogaster* infected with DSimSV-E and DMelSV-A virus: 95% confidence intervals for DSimSV-E (G1: 57.7, CI 36.4-78.9; G2: 88.9, CI 58.3-119.4; G3: 87.5, CI 42.1-132.9) overlapped with the point estimate as well as the 95% CI for

DMelSV-A (G1: 58.3, CI 19.3-98.3; G2: 100.0, CI 100.0-100.0; G3: 100.0, CI 100.0-100.0; see Table 5-2). As observed previously (chapter 3), the few *D. melanogaster* lines that were infected with DMelSV-A tend to respond to selection for transmission efficiency more quickly than the *D. simulans* lines infected with DMelSV-A (Figure 5-3, Table 5-2).

Components of Host Fitness as a Proxy for Virulence

Two components of host fitness, fecundity and hatchability, were examined for three generations following infection by injection. Data for both fitness components were standardized to the uninfected means of the relevant host species to correct for differences in environmental heterogeneity between generations.

Fecundity as a Proxy for Virulence

Neither the error bars nor the confidence intervals for DSimSV-E-infected *D. melanogaster* and *D. simulans* overlap unity in G1. Thus, the virus is virulent with respect to fecundity in both hosts (that is, fecundity is lower in infected animals than in uninfected, sham-injected conspecifics; Table 5-3, Figure 5-4). DSimSV-E is also virulent in the *D. simulans* host in G2, but not in *D. melanogaster.* Fecundity is not reduced in either host in G3.

Interestingly, fecundity of DSimSV-E infected *D. melanogaster* is lower than that of DSimSV-E-infected *D. simulans* in G1 (*D. melanogaster*: 0.132, CI -0.013-0.278; *D. simulans*: 0.480, CI 0.128-0.833), but higher in G2 (*D. melanogaster*: 0.956, CI 0.265-1.647; *D. simulans*: 0.504, CI 0.131-0.876), and indistinguishable in G3 (*D. melanogaster*: 0.656, CI -0.801-2.114; *D. simulans*: 0.899, CI 0.099-1.698; see Table 5-3, Figure 5-4). Interestingly, although the small sample size (three strains in G1) precludes formal Wilcoxon signed rank tests, an analogous examination of the data

shows that *D. simulans* has higher fecundity in G1 for all three of the viral strains that were transmitted successfully in both hosts (Table 5-3). This difference is not trivial, with a fecundity reduction of 32% in *D. melanogaster*. Fecundity in *D. melanogaster* rebounds in the next generation, at least in the two strains that are present in both species in G2 (fecundity in *D. melanogaster* is more than twice as high as in *D. simulans*, which is unchanged; Figure 5-4). Thus, although the sample size is small, the data are consistent with rapid change between G1 and subsequent generations in *D. melanogaster*, but not in *D. simulans.*

*D*rosophila *melanogaster* infected with DSimSV-E had lower relative fecundity at G1 (0.132, CI -0.013-0.278) than did conspecifics infected with DMelSV-A (0.755 CI 0.037-1.473; *i.e.,* standardized fecundity for DSimSV-E standard errors did not overlap with DMelSV-A; Table 5-3, Figure 5-4). Again, DSimSV-E-infected *D*. *melanogaster* appear to rebound quickly, as their relative fecundity is higher than that of DMelSV-A-infected conspecifics at G2 (0.956 CI 0.265-1.647 and 0.315 CI -0.065-0.695, respectively; Table 5-3) and is indistinguishable at G3 (0.656, CI -0.801-2.114 and 0.678, CI –0.855-2.212, respectively; Table 5-3).

Errors for DSimSV-E in *D. simulans* do not overlap with point estimates for DMelSV-A in conspecifics, in any of the three generations, though the means appear to be closer to DMelSV-A than in *D. melanogaster* (Table 5-3). Additionally, errors for the DSimSV-E-infected *D. simulans* do not overlap 1 in G1 or in G2 (*i.e*., fecundity is reduced in infected flies relative to uninfected flies; the virus is virulent). However, they do include 1 in G3, again suggesting that host, virus, or both have changed in such a manner as to reduce virulence (Table 5-3).

Hatchability as a Proxy for Virulence

Hatchability for *D. melanogaster* and *D. simulans* infected with DSimSV-E was similar in G1 and G3 (*D. melanogaster*: G1: 0.757, CI 0.036-1.479; G2: 1.443, CI 0.292-2.594; G3: 0.316, CI -0.271-0.904; and *D. simulans*: G1: 0.709, CI 0.346-2.000; G2: 0.553, CI 0.098-1.009; G3: 0.346, CI -0.176-0.868; Table 5-4, Figure 5-5), but higher for *D. melanogaster* in G2.

Hatchability was not decreased in *D. melanogaster* infected by DSimSV-E relative to sham-injected, uninfected conspecifics in the first two generations, but was in the third (*i.e.* both confidence intervals and error bars overlapped unity; see above and Table 5-4, Figure 5-5). Surprisingly, hatchability of *D. simulans* was reduced relative to uninfected controls but only at G3 (G1: 0.709, CI; 0.346-2.000, G2: 0.553, CI 0.098-1.009; G3: 0.346, CI -0.176-0.868). Thus, DSimSV-E virulent with respect to hatchability in the host in which it had evolved for 16 generations (*D. simulans*), but not in the ancestral host, *D. melanogaster*.

Further evidence that DSimSV-E was avirulent with respect to hatchability in *D. melanogaster* comes from the observation that our point estimate of hatchability for *D. melanogaster* infected by DMelSV-A (0.936) was included in both the confidence intervals and the error bars for conspecifics infected with DSimSV-E for all three generations (above and Table 4). Hatchability for *D. simulans* infected by DSimSV-E was similarly indistinguishable from conspecifics infected by DMelSV-A, so the protracted period of evolution in this species did not result in “rescue” of virulence with respect to hatchability.

Discussion

We tested the generality that, after a host shift, a pathogen becomes attenuated and its infectivity toward its native host is reduced. We compared the ability of the evolved virus and the unevolved ancestral virus to infect naïve native hosts (*D. melanogaster*) after 16 host generations of evolution in a novel host, *D. simulans*. The ancestral virus (DMelSV-A) and the evolved virus (DSimSV-E) infected injected flies at similar rates. This suggests that the infectivity toward the native host was not attenuated, despite 16 host generations in the novel host.

However, successful infection requires the successful infection of the next host. Thus we examined whether the flies that were successfully infected by injection were able to produce infected offspring. We also tested whether native host females infected with DSimSV-E transmitted the virus to their offspring with lower efficiency than the flies infected with DMelSV-A, as predicted by the attenuation hypothesis (EBERT 1998). As seen previously (Chapter 3), *D. melanogaster* infected by injection with DMelSV-A transmit the virus to their offspring at relatively low levels. Interestingly, *D. melanogaster* females infected by injection with DSimSV-E actually transmitted the evolved virus to their offspring at higher levels (12.1% vs 7.0%; not statistically significant) than their conspecifics transmitted the ancestral virus.

Finally, we tested whether transmission efficiency would be similar across three host generations for native hosts infected with either DMelSV-A (ancestral virus) or DSimSV-E (evolved virus). Infection levels were nearly identical at every generation for *D. melanogaster* females infected with either ancestral or evolved virus, consistent with lack of attenuation.

However, absence of attenuation does not necessarily mean absence of adaptation in the novel host. *D. simulans* females injected with DSimSV-E had uniformly higher infection levels at G1 than did the *D. melanogaster* flies that received the same virus type suggests that DSimSV-E did, in fact, adapt to the novel host but that this adaptation did not come at the cost of attenuation in the native host.

Taken together, the results on the infectivity of evolved virus do not support the attenuation generality that was developed based on results from SPE studies (EBERT 1998). Instead of the predicted attenuation, the evolved virus actually was more infective than the ancestral virus during the early stages of the infection. Subsequent infection rates were indistinguishable at each generation. There are few, if any, other examples where infectivity was not attenuated but instead increased when a pathogen evolved on a novel host was returned to its native host (ANTONOVICS *et al.* 2002).

We also tested the whether virulence would be attenuated by comparing components of host fitness in hosts infected with either ancestral or DSimSV-E (Figure 5-4 and 4-5). In *D. simulans,* fecundity was unaffected by virus type (Figure 5-4): both ancestor and DSimSV-E reduced fecundity similarly. However, in *D. melanogaster*, DSimSV-E was *more* virulent than the ancestor in G1, but less virulent than the ancestor in G2 and indistinguishable in G3. Higher virulence than the ancestor in the native host is once again inconsistent with attenuation. Moreover, the fecundity cost incurred by infection with DSimSV-E was higher in *D. melanogaster* than in *D. simulans*. Thus, even though fecundity in the novel host was not significantly different for the ancestral virus and the evolved virus, the virus may have evolved in a way that was (in

the context of this study, not in an absolute sense) neutral in the novel host but deleterious on reintroduction to the native host.

In contrast, hatchability of *D. melanogaster* infected with the ancestral virus or with DSimSV-E was indistinguishable; moreover, hatchability was the same between infected and uninfected flies in this species. Interestingly, in *D. simulans*, hatchability was adversely and similarly affected by infection with DSimSV-E and DMelSV-A. Again we did we not see attenuation.

This result is contrary to the SPE conclusion of attenuation but is not entirely novel (EBERT 1998). A hypervirulent bacterial insect pathogen (*Xenorhabdus nematophila*) exhibited significant increases in virulence when evolved on a novel host and then returned to its native host (CHAPUIS *et al.* 2011).

Our results lead us to conclude that the predicted outcome of pathogen attenuation following a host shift is not ubiquitous, as demonstrated when the DMelSV rhabdovirus is evolved in a novel host (*D. simulans)* and reintroduced into its native host, *D. melanogaster*. The lack of attenuation is contrary to the predictions developed from SPE studies and may lend support to the suggestion that the experimenters in SPE studies may be inadvertently influencing at least some experimental outcomes by guaranteeing the success of the pathogens being studied (EBERT 1998). Another factor that could explain our contrary results is the use of a complex multicellular host. Pathogens infecting complex host are challenged differently than pathogens infecting unicellular host and, as such, we might expect different evolutionary strategies. At the least, more work with complex hosts is necessary to further explore these results.

Table 5-1. The virus lines used for cross infection inoculum production exhibited high virus titers.

	Titer	Infectivity (%)
DSimSV-E 91	5.18 $x10^6$	100
DSimSV-E 52	3.54 $x10^6$	100
DSimSV-E 155	1.77 $x10^6$	100
DSimSV-E 177	1.77 $x10^6$	100

Lines of virus evolved in *Drosophila simulans* (DSimSV-E) that exhibited high virus titers (genome copies/µg of host RNA) and infectivity (>50% of flies infected) in *D. simulans* after passaging in this novel host for 16 host generations.

Table 5-2. Confidence intervals for per cent infection in *Drosophila simulans* (DSim) and *D. melanogaster* (DMel) adults infected with either evolved (DSimSV-E) or ancestral (DMelSV-A) virus.

	Generation	Estimate	StDev	StErr	Upper CI	Lower CI	N
DSimSV-E/DSim	1	76.9	25.0	1.7	101.4	52.4	4.0
	2	90.5	19.7	1.6	109.8	71.3	4.0
	3	94.7	15.4	1.8	112.2	77.3	3.0
DSimSV-E/DMel	1	57.7	18.8	5.2	78.9	36.4	3.0
	2	88.9	22.0	7.3	119.4	58.3	2.0
	3	87.5	23.1	8.2	132.9	42.1	1.0
DMelSV-A/DSim	1	62.8	22.0	3.2	106.0	19.6	1.0
	2	89.3	20.7	2.8	129.9	48.7	1.0
	3	76.3	25.6	5.9	126.6	26.0	1.0
DMelSV-A/DMel	1	58.3	20.4	8.3	98.3	18.3	1.0
	2	100.0	0.0	0.0	100.0	100.0	1.0
	3	100.0	0.0	0.0	100.0	100.0	1.0

Table 5-3. Confidence intervals for fecundity standardized to the uninfected control mean in *Drosophila simulans* (DSim) and *D. melanogaster* (DMel) females infected with either evolved (DSimSV-E) or ancestral (DMelSV-A) virus.

	Generation	Estimate	StDev	StErr	Upper CI	Lower CI	N
DSimSV-E/DSim	1	0.480	0.360	0.024	0.83	0.13	4.0
	2	0.504	0.380	0.031	0.88	0.13	4.0
	3	0.899	0.707	0.081	1.70	0.10	3.0
DSimSV-E/DMel	1	0.132	0.129	0.036	0.28	-0.01	3.0
	2	0.956	0.499	0.166	1.65	0.26	2.0
	3	0.656	0.744	0.263	2.11	-0.80	1.0
DMelSV-A/DSim	1	0.610	0.466	0.068	1.52	-0.30	1.0
	2	0.392	0.261	0.035	0.90	-0.12	1.0
	3	1.114	0.793	0.182	2.67	-0.44	1.0
DMelSV-A/DMel	1	0.755	0.366	0.150	1.47	0.04	1.0
	2	0.315	0.194	0.052	0.69	-0.07	1.0
	3	0.678	0.782	0.319	2.21	-0.86	1.0

Table 5-4. Confidence intervals for hatchability standardized to the uninfected control mean in *Drosophila simulans* (DSim) and *D. melanogaster* (DMel) females infected with either evolved (DSimSV-E) or ancestral (DMelSV-A) virus.

	Generation	Estimate	StDev	StErr	Upper CI	Lower CI	N
DSimSV-E/DSim	1	0.709	0.371	0.026	0.760	0.83	0.13
	2	0.553	0.465	0.039	0.631	0.88	0.13
	3	0.346	0.461	0.058	0.461	1.70	0.10
DSimSV-E/DMel	1	0.757	0.638	0.202	1.213	0.28	-0.01
	2	1.443	0.830	0.277	2.082	1.65	0.26
	3	0.316	0.300	0.134	0.689	2.11	-0.80
DMelSV-A/DSim	1	0.781	0.380	0.055	0.893	1.52	-0.30
	2	0.557	0.419	0.060	0.677	0.90	-0.12
	3	0.391	0.530	0.132	0.673	2.67	-0.44
DMelSV-A/DMel	1	0.936	0.451	0.184	1.409	1.47	0.04
	2	1.812	1.194	0.331	2.533	0.69	-0.07
	3	0.722	0.574	0.287	1.635	2.21	-0.86

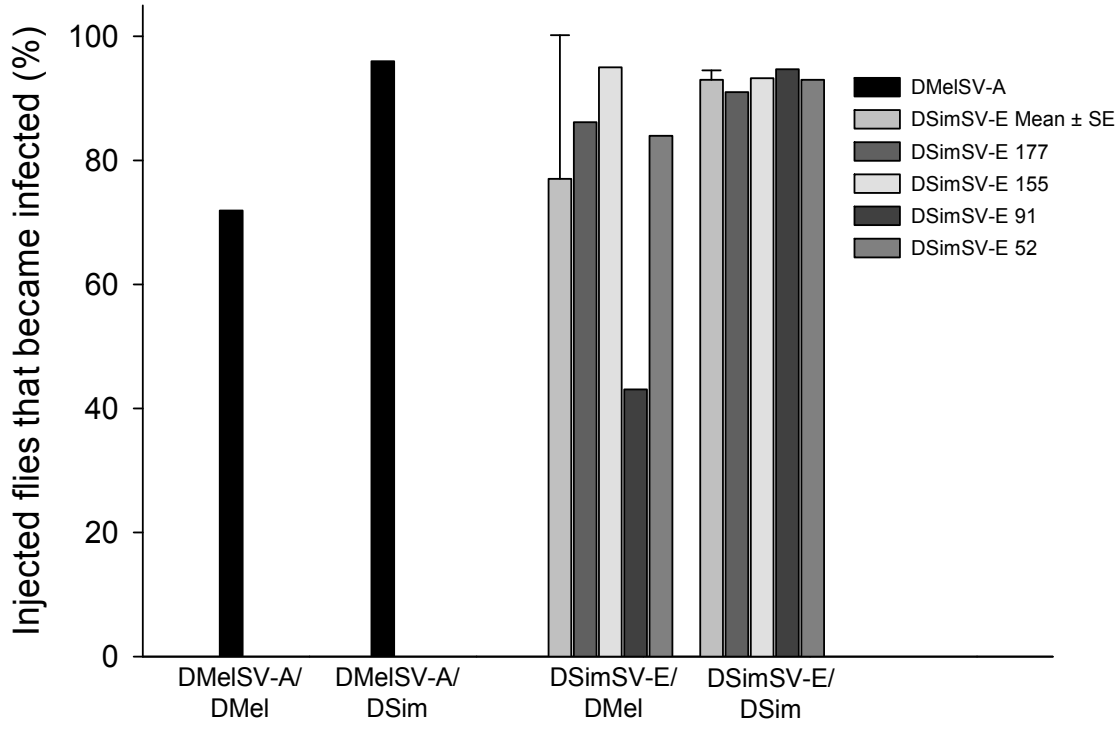

Figure 5-1. Infectivity of virus passaged in the novel host, *Drosophila simulans* (DSimSV-E), was not attenuated to the native host, *D*. *melanogaster*. Per cent of naïve injected flies that became infected is shown for each inoculum x host combination. DSimSV-E and the original ancestral virus (DMelSV-A) infected naïve *D*. *melanogaster* at similar levels ($P = 0.3605$). DMelSV-A infected *D*. *simulans* more effectively than *D*. *melanogaster* ($P < 0.0001$), while DSimSV-E and DMelSV-A infected *D*. *simulans* or *D*. *melanogaster* similarly ($P = 0.3605$ and $P = 0.079$, respectively).

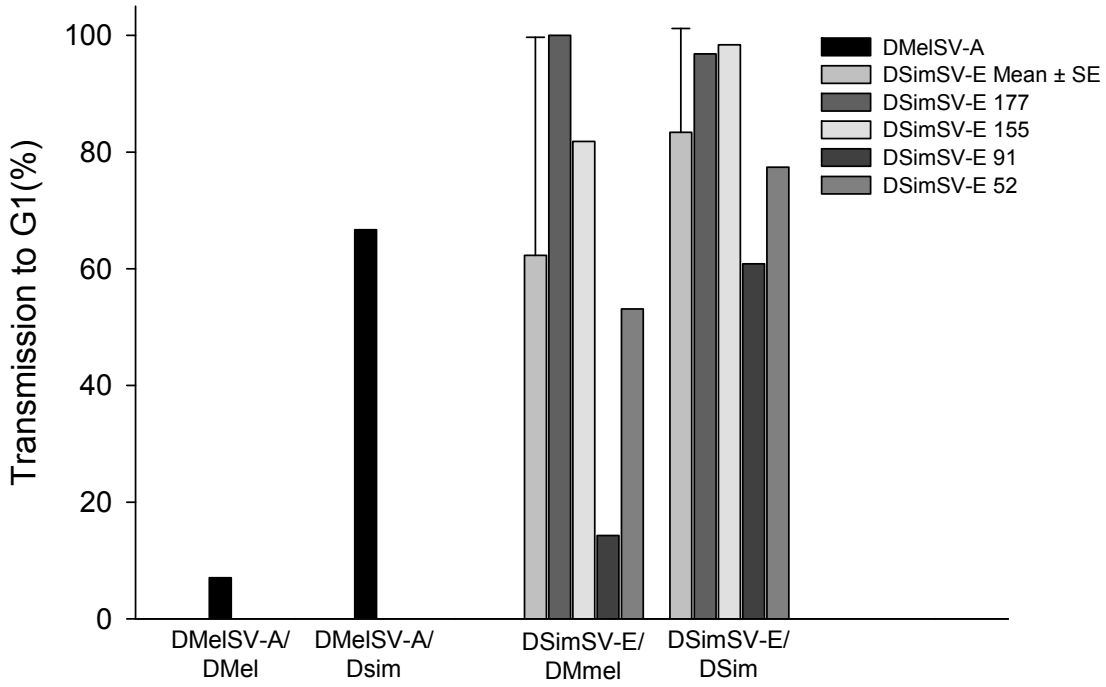

Figure 5-2. Transmission efficiency of virus evolved in *Drosophila simulans* (DSimSV-E) was not attenuated to the native host, *D*. *melanogaster*. Per cent of infected flies that successfully infected their offspring is shown for each inoculum x host combination. DSimSV-E-infected flies transmitted the virus to their offspring at significantly higher levels than the *D*. *melanogaster* flies initially infected with ancestral virus (DMelSV-A, $P = 0.0104$). DSimSV-E and DMelSV-A infected *D*. *melanogaster* females transmitted virus to their offspring at similar levels ($P = 0.3135$)

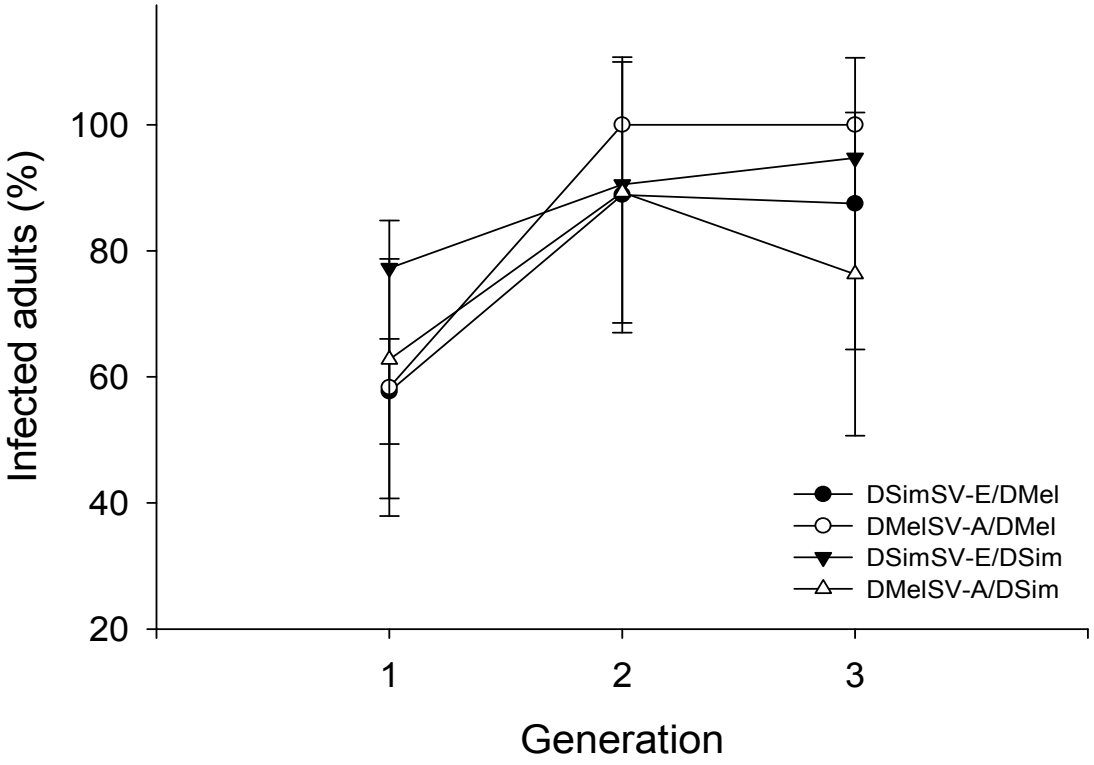

Figure 5-3. The ability to produce increasing infection levels in successive generations was not attenuated when virus evolved in *Drosophila simulans* (DSimSV-E) and re-introduced to its native host, *D*. *melanogaster*, when compared to the performance of the unevolved ancestral virus *(*DMelSV-A*)* in the native host. Per cent of infected adult flies shown for each generation. Infection levels were statistically indistinguishable at each generation for *D. melanogaster* infected with either DSimSV-E (black circles) or DMelSV-A (white circles). Additionally, DSimSV-E was more infective in *D*. *simulans* (black triangles) than in *D*. *melanogaster*, but only at G1.

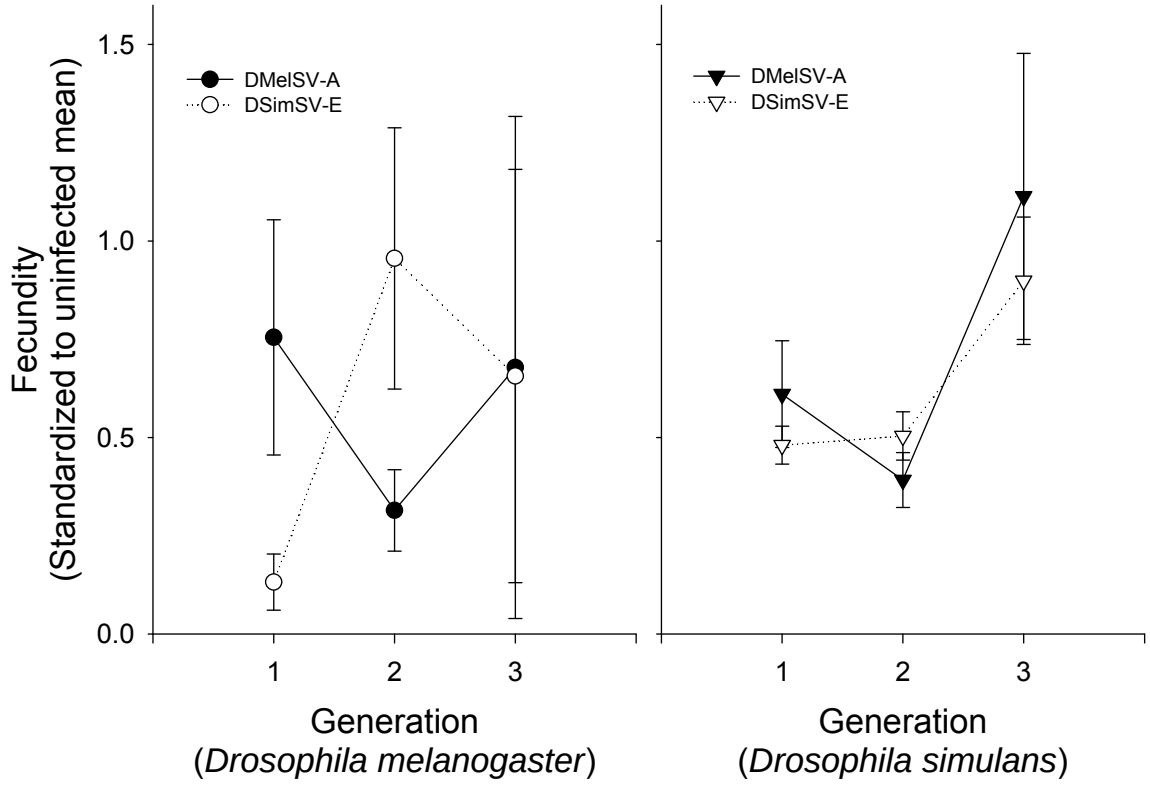

Figure 5-4. Contrary to prediction, the virus passaged in *Drosophila simulans* (DSimSV-E, white circles) appears to be more virulent in the native host (*D. melanogaster*) than the ancestral virus (DMelSV-A, black circles). Fecundity standardized to the control mean is show for *D. simulans* and *D. melanogaster* infected with DSimSV-E or DMelSV-A. DSimSV-E-infected *D. melanogaster* had significantly lower fecundity at G1 and conspecifics infected with DMelSV-A. DMelSV-A-infected *D. melanogaster* appear to rebound quickly as their fecundity is significantly higher that DSimSV-E infected conspecifics at G2 ; no difference in fecundity was observed for these two groups at G3. Fecundity of DSimSV-E and DMelSV-A-infected D. simulans was statically indistinguishable at all three host generations.

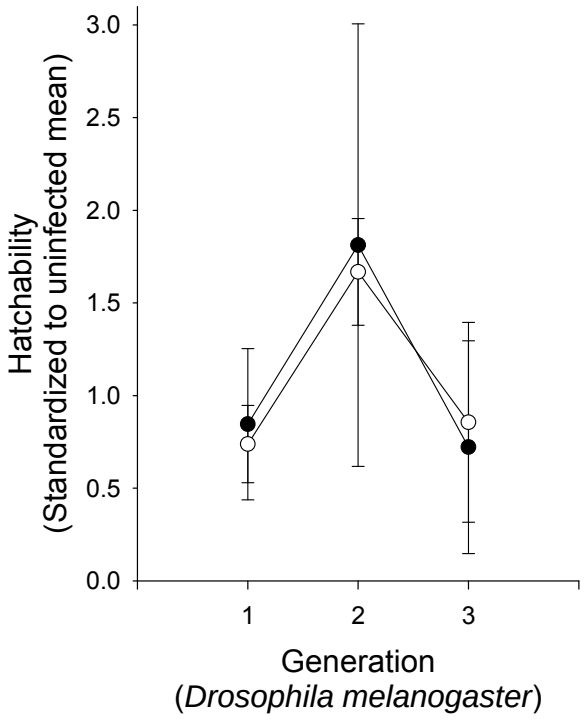

Figure 5-5. Hatchability was indistinguishable for *Drosophila melanogaster* receiving unevolved ancestral virus (black circles) or virus evolved in *D. simulans* (white circles).

CHAPTER 6
CONCLUSIONS

Summary of Findings

Understanding the evolution of virulence continues to be an important area of research as we strive to prevent and cure diseases of humans, non-human animals, and plants. Of particular interest are the events that occur after a pathogen invades a novel host (i.e., a host shift). Most of the currently accepted theory on the evolution of virulence following a host shift is based on data developed using single celled hosts in serial passage experiments (EBERT 1998). Based on these data, it is expected that after a host shift, a pathogen will be highly virulent to its novel host and will grow progressively more attenuated (i.e., less infection- and replication-competent) to its native host as time passes.

Studying host shifts in nature is difficult because natural host shifts happen in the absence of observation and because the standing genetic variation present in the pathogen population immediately prior to the shift is unknown. Therefore it becomes difficult to know whether the polymorphisms observed in the novel host are de novo mutations or simply genetic variants that were present in those parasites of the original host that successfully colonized the novel host. An additional complication is whether or not these ancestral variants are actually advantageous in the new host, or simply increased in frequency due to founder effects; resolution of this conundrum is beyond the scope of the current work. Additionally, the host shifts of interest (particularly those with potential to infect humans) involve complex multicellular hosts. To address these issues, I developed a tractable model system using *Drosophila* spp. and the sigma rhabdovirus, which, respectively, are good models for dipteran disease vectors and

dipteran-carried pathogenic RNA viruses. Using this system, the genetic variation in sigma virus from natural populations of the native host was determined using next generation SOLiD™ sequencing, a shift onto a novel host was induced and tracked in the laboratory, and the resulting evolved virus was returned to its native host. Finally, virus lines that successfully colonized the novel host were sequenced.

I found that the sigma rhabdovirus has lower genetic variation than is typically expected of RNA viruses and that a large part of the variation likely is due to ADAR activity. The sigma virus was more infective on the novel host than on the native host and, as predicted by serial passage studies (SPE), it was also more virulent on the novel host. Contrary to prediction, I found no evidence of attenuation when the evolved virus was returned to its native host; in fact, the opposite was true early in the infection. Sequencing the major variant from virus strains that successfully infected the novel host identified some viruses that were identical or nearly identical to the ancestral virus, while other viruses that were genetically drastically different from the ancestral virus. However, both types of virus successfully achieved and maintained high transmission efficiency on the novel host. These results indicate that the evolution of virulence following a host shift in multicellular animals is more complex than what is seen in microbial systems, and indicate that this new model system can be used successfully to learn about the evolution of virulence.

Adenosine-Dependent RNA Deaminases

Although a large part of the genetic variation identified through sequencing was attributable to ADAR activity, a better understanding of ADAR activity within *D. melanogaster* is needed to determine whether or not this is an antiviral response favorable to host fitness and deleterious to viral fitness, as previously suggested

(CARPENTER *et al.* 2009). In *D. melanogaster*, dADAR activity helps control nervous system integrity and function in adult flies and is not necessary during development (PALLADINO *et al.* 2000b). The eggs and sperm of infected flies contain virions (BRUN and PLUS 1980); however, the detailed mechanics of inter-generational and inter-cellular transmission of sigma virus are unknown. Determining whether or not dADAR-mutated virions are transmitted to the next generation of flies and moreover successfully replicate in them would first require determining conclusively that the production of transmitted virions occurred in the absence, or at least decrease, of dADAR activity. Additionally, the major variant from the contingent of virions that are vertically transmitted within flies for 15 successive generations could be sequenced and compared to determine whether dADAR affected the transmitted viral genomes, in other words, if A-to-G changes are detectable when the viral genomes present at each generation are compared. In vertebrates, ADAR1 does help modulate the host response to infection (GEORGE and SAMUEL 2011; WARD *et al.* 2011); however, no such activity has been determined for ADAR2, which is the homolog of dADAR (GEORGE and SAMUEL 2011; PALLADINO *et al.* 2000a). Moreover, there may be host genetic variation for dADAR activity. Further analysis will be required to determine which, if any, mutations in the virus genomes following 16 host generations of host evolution resulted from ADAR changes. It would be particularly interesting to attribute differential virulence and/or titer to the presence or absence of ADAR changes; however, this remains to be seen.

Transmissibility and Virulence Following a Host Shift

The novel host (*D. simulans*) was easier to infect than the native host (*D. melanogaster*), but many lines of evidence suggested that selection against the virus was stronger in the novel host than the native host. Virus-host interactions at the cellular

level are critical for successful colonization of novel host lineages (DENNEHY *et al.* 2007; PERLMAN and JAENIKE 2003); this is one possible explanation for why more infections failed on the novel host than on the native host. Although the virus was successful in reaching new hosts in both species, response to selection for transmission efficiency was faster on the native host and decreased more rapidly in the novel host following relaxation of selection for transmission efficiency. Thus, it is likely that the adaptive walk required fewer steps (mutations) in the naïve native host than in the novel host (ANDRE and HOCHBERG 2005; BULL and OTTO 2005). This makes sense given that DMelSV is more mutationally robust than previously expected (BRUSINI *et al.* 2012 in review; CLAAS *et al.* 1998). As such, the virus would be less able to adapt to a novel host than to a naïve native host when all else is equal (ANDRE and HOCHBERG 2005; MOYA *et al.* 2000). Once successful transmission was established, the virulence of the virus on the native and hovel hosts was measured. The virus was virulent on both hosts but, as predicted by SPE results (EBERT 1998), it was more virulent on the novel host. Interestingly, in both hosts, groups with higher viral titer had higher virulence than groups with lower titer. A positive association between titer and transmission efficiency has been observed for many horizontally transmitted parasites (EWALD 1983; FRANK 1996; LENSKI and MAY 1994). Higher virulence is expected to be associated with lower transmission, even though higher titer may be associated with higher transmission; accordingly, an association between titer and virulence should result in a virulence/transmission tradeoff (LIPSITCH *et al.* 1996; SHARON *et al.* 1999). In DMelSV, a vertically transmitted parasite, we see that virulence decreased over time but transmission also decreased over time. These observations are not consistent with our expectation that virulence and

transmission should be negatively associated. It is possible that we are not effectively capturing the primary cause of virulence with our measurements of hatchability and fecundity, or that there is sufficient environmental variation that we are unable to distinguish signal from noise statistically, but selection (unconstrained by concerns of normality) can do so. It is also possible that for this system, the interaction between transmission and virulence was not a tradeoff but a compromise.

Sequencing Evolved Virus

The expectation from evolution of virulence theory is that a virus evolved on a novel host will possess genetic differences that correspond to its adaptation on that host (EBERT 1998). This suggests that conspecifics pathogens should converge on a common genotype if given enough time. Here we saw that two major genotypes occurred in viral strains that were equally successful in the novel host. One of the genotypes was nearly indistinguishable from the ancestral virus while the second was quite different. The three strains that had the different genotype suggest that either convergence occurred or some other mechanism was at play. It is likely that that different genotype was favored early in the infection process and became the predominant variant in the three lines were it was present. Conversely, that genotype was either absent or outcompeted in the two novel host lines where the major variant was very similar to the ancestral virus. In either case it is clear that this was not a case of simple convergence onto a main genotype.

Testing Attenuation

The expectation from evolution of virulence theory is that a virus will become attenuated to its native host as a result of passage on a novel host (EBERT 1998). Substitutions that facilitate transmission in the novel host could literally replace

substitutions that facilitated transmission in the native host. Attenuation is thus defined as a decreased ability to infect and/or persist on the native host. Accordingly, we tested attenuation by comparing the ability of the evolved virus (16 generations on the novel host) and the unevolved ancestral virus to infect naïve native hosts (*D. melanogaster*). The ancestral virus (DMelSV-A) and the evolved virus (DSimSV-E) infect injected *D. melanogaster* (native) flies at similar rates. Surprisingly, native host females infected with DSimSV-E transmitted the virus to their offspring with higher efficiency than the flies infected with DMelSV-A. Neither infectivity nor transmission then are consistent with the predictions of the attenuation hypothesis (EBERT 1998). A similar but different results was observe in a hypervirulent bacterial insect pathogen (*Xenorhabdus nematophila*) instead of being attenuated actually grew more virulent toward its native host (CHAPUIS *et al.* 2011). Therefore it is clear that the attenuation results from SPE studies do not hold for all systems, further work is needed to broaden our understanding of pathogen attenuation in complex multicellular organisms following host shifts.

In conclusion, it is clear that the questions on standing genetic diversity within pathogen populations and questions on the evolution of virulence and attenuation following a host shift require more attention. Although our study was small (one genotype of one novel host), it is compelling and suggests that conclusions generated in one model system may not be directly applicable to all other systems in question. Therefore, these questions need to be tested further on a larger scale using more genotypes of novel and native hosts, more species of host from within the same genus as the native host, hosts from different genera within the same family, or even using hosts from different families within the same order. With such a hierarchical approach,

one can consider standing genetic variation in both host and parasite populations and thoroughly address questions of virulence and attenuation. These studies will further our understating of the processes regulating the evolution of virulence, particularly as it relates to the phylogenetic length/distance that is spanned by a host shift. Using such a design may also permit us to evaluate over what taxonomic distances generalities may be constructed, or may be believable. Increasing our understanding in this area is critical because all of the parasites that jump into humans come from genera other than our own. In reality, only a few known cases of viral host shifts originated within the Hominidae [HIV for example (CALVIGNAC-SPENCER *et al.* 2012; GAO *et al.* 1992)] and while some zoonoses originated in mammals [SARS, for example (HOLMES and RAMBAUT 2004)], diseases like influenza originated from multiple steps: from other vertebrate orders to other mammals, either through direct contact with humans or through insect vectors that themselves became infected (BAIGENT and MCCAULEY 2003; GODDARD *et al.* 2003)].

Thus far, our work has demonstrated that at least for this system, some of the generalities from microbial systems do not hold. Our results are important because over a quarter of the emerging infectious diseases cataloged in a 50 year period were vector-borne (JONES *et al.* 2008); although many insects groups can vector disease, most insect vectors of significant human importance are dipterans just like *D. melanogaster* (BEATY and MARQUARDT 1996). Therefore, this study paves the way for larger future studies that can include a wide range of host species within the Insecta. Such studies will provide information on the evolution of virulence as it relates to host shifts in general

and as it relates to pertinent taxa in particular, like the Culicidae, whose members are the vectors of many diseases of humans and livestock.

APPENDIX A
MOLECULAR BIOLOGY PROTOCOLS

RNA Isolation from One Fly

RNA isolation was conducted using TRIzol® and following the manufacturer's guidelines (www.invitrogen.com).

1. Homogenization
    a. Chill the fly and put it into a clean tube with 200 µl of TRIzol® (for multiple flies add 100 µl per extra fly up to 1ml; up to 50 flies can be done in 1 ml of TRIzol. If doing more than one fly, ramp all volumes accordingly (except for ethanol washes and final suspension, those stay constant).
    b. Homogenize ON ICE with Kontes homogenizers until only the wings are distinguishable.
    c. Incubate the homogenized samples for 5 min at room temp.
2. Phase separation
    a. Add 40 µl of chloroform for the first fly and 20 µl more for each additional fly.
    b. Cap sample tubes securely.
    c. Vortex tubes for 5 seconds and incubate them at room temperature for 2 to 3 min.
    d. Centrifuge the sample at no more than 12,000 × g for 10 min at 2 to 8°C. Following centrifugation, the mixture separates into a lower red, phenol-chloroform phase, an intermediate phase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The volume of the aqueous phase is about 60% of the volume of TRIZOL® Reagent used for homogenization.
    e. Transfer the aqueous phase to a fresh tube. Do not contaminate the aqueous phase with the organic phase.
3. RNA precipitation
    a. Add 100 µl of isopropyl alcohol (50 µl more for each additional fly) to the aqueous phase and **shake vigorously** but do not vortex**.**

b. Incubate samples at -20ºC for 10 min.

c. Centrifuge at 12,000g for 10 min at 2 to 8°C.

i. This should produce a gel-like pellet on the side and bottom of the tube.

ii. Remove and discard the supernatant.

4. Ethanol wash

a. Add 1 ml of cold (-4ºC) 75% ethanol (this volume remains the same whether you do 1-10 flies).

b. Mix the sample by vortexing and centrifuge at no more than 7,500 × g for 5 min at 2 to 8°C. RNA can be left in alcohol at 4ºC for 1 week or longer at-20ºC.

c. Discard alcohol

d. Quick spin the tubes and use a pipette to discard extra alcohol at bottom of tubes.

5. Re-suspension of RNA

a. Air-dry the RNA pellet for 5 min. (pop the lid and leave in rack on bench top).

b. Dissolve RNA

i. add 29 µl of DEPC-treated RNAse-free water

ii. add 1 µl of RNAse inhibitor and mix by flicking the tube (DO NOT mix by pipetting)

iii. Centrifuge the tube briefly to collect the solution at the bottom.

iv. Incubate for 10 min at 55°C.

c. Cool the tube on ice, give a quick spin and store it at -80°C.

Reverse Transcription

The whole genome of the sigma rhabdovirus was reverse transcribed using a primer complementary to the 3' end of the RNA genome (whole1F: 5’ TAGAAGCATCCTCGGCTTTC 3’) and Superscript III first strand synthesis kit (www.invitrogen.com).

1. Combine 200ng of genomic RNA (4 µl of 50ng/ µl), 1 µl of whole1F primer, 1 µl of 10mM dNTP mix and 7 µl of DEPC-treated water for a final volume of 13 µl.

2. Incubate this mix at 65°C for 5 min and then chill on ice.

3. Combine the following items from the Invitrogen first strand synthesis kit: 4 µl of the 5X first strand buffer, 1 µl of DTT, 1 µl of RNAseOUT, 1 µl of Superscript III RT enzyme.

4. Add the mixtures from steps #2 and #3 and incubate at 50°C for 50 min.

5. Successful reverse transcription can be confirmed by PCR amplifying a fragment of the viral genome from each end of the viral genome using the newly synthesized cDNA as template; subsequently.

The Polymerase Chain Reaction

Routine PCR was performed using a basic master mix (www.promega.com).

1. Reactions were composed of 12.5 µl of 2X master mix, 1 µl of forward and 1 µl of reverse primer, 1 µl of template and 9.5 µl of DEPC-treated water.

2. Reactions were incubated in a thermal cycler for two min at 94°C followed by 40 cycles of 94°C for 30 seconds, 55°C for one min, and 72°C for one min.

3. Subsequently, the reactions were incubated for an additional 5 min at 72°C and then stored at 4°C.

4. PCR reactions were electrophoresed in 1% agarose gels made with 1X Tris acetate EDTA (TAE, 96.8 g Tris, 22.84 mL Acetic Acid,14.88 g EDTA), 2L of dH_20).

5. The gels were submerged in 1X TAE and electrophoresed at 80 volts for 45 min.

Quantitative Polymerase Chain Reaction

The N gene was amplified from each viral cDNA template using a tagged primer (NplusTag: 5'*GCAGTATCGTGAGTTCGAGTGT*CCGATGACCTGTCCGTAAC3', 22bp of non-Sigma sequence (italicized) followed by 20bp of Sigma-specific sequence). Virus titer quantification was achieved using strand-specific quantitative PCR and a Taqman® primer/probe set designed against the sigma virus *N*. The forward primer matched the tag sequence from the RT primer (tag:

5’GCAGTATCGTGAGTTCGAGTGTt3’; derived from Purcell et al. (2006) while the Taqman probe (Tprobe: 5’CATGAGATGGAGGAACTTTCTCTCCCA3') and the reverse primer (R: 5’GAGTCGCAGCTTTGGAGTTC3’) were specific to the Sigma *N* gene. The tagged primer was complementary only to the viral genome and, as such, allowed for quantification of genomic RNA but not mRNA or the full length genome intermediate (both are complementary to the genome).

1. 2µL of cDNA (equivalent to a quantity of 50ng of total RNA) were added to 25 µL of Real-Time PCR master mix (Applied Biosystems, Foster City, CA), 0.9 µL of both of primers at 50mM, 1.25 µL of the Taqman probe at 10mM and 19.95 µL of DEPC water for a total volume of 50 µL.
2. This reaction was split into 15 µl triplicates.
3. A serially diluted PCR amplicon containing the region being quantified was used for absolute quantitation using a five point serial dilution standard curve(10^7 to 10^3).
4. The reactions were amplified using a *StepOnePlus™ Real-Time PCR* System (Applied Biosystems, Foster City, CA).

APPENDIX B
FLY HANDLING PROTOCOLS

Drosophila Rearing

All flies were reared at 24°C with a 16:8 light:dark photoperiod in a Percival I36VL incubator (www.percival-scientific.com). The flies were reared on Jazz-Mix *Drosophila* food (www.fishersci.com) using standard narrow polystyrene drosophila vials.

1. The food was prepared according to the manufacturer's recommendations.
    a. 226.8 grams of food were mixed with 1.2 liters or water, brought to a boil and cooked at a low rolling boil for 10 min.
    b. The food was poured into narrow polystyrene vials (7ml per vial) using a Wheaton Unispense (www.fisher.com pump unit and a 3mm diameter hose. The food was cooled on the bench top overnight.
2. The vials were prepared to receive flies.
    a. Each vial was sprinkled with some baker's yeast (Fleischmann's brand was always used; 4-6 grains).
    b. A piece of Kim Wipe (1/8 of a sheet) was inserted into each vial until it reached the bottom (between the food and the side of the vial).
    c. The vial was capped with a cotton ball and refrigerated until needed.
3. New vials were set up with flies on a 21 days schedule with constant density and sex ratio (5♀ + 5♂).
    a. Adults were added on day 1 and the females were allowed to oviposit.
    b. The adults were cleared from the vials on day 7.
    c. The next generation was setup on day 21 (new day 1).

Artificial Injection of Female *Drosophila*

Inoculum was prepared according to Clark et al. (1979). Flies were collected such that only flies < 2 days old were injected. This is because at this age the abdomen is still “soft” and will distend to accept the inoculum. All flies were injected using a Sutter

instruments XenoWorks™ Analog Microinjector http://www.sutter.com/products/product_sheets/Xenoworks_Analog_Injector.html) equipped with a 100 µl syringe. The needle blanks used were Narishige GD-1 (Serial #08NG1025) purchased from Tritech Research, Inc. All blanks were pulled using a Kopf Model 720 vertical needle puller.

Inoculum Production

1. Place 100 adults in a 1.5ml Eppendorf tube. If you have more than 100 adults per line, make multiple homogenizations.
2. Add 500 µl of cold (4°C) homogenization buffer (HB, 0.005M Tris, 0.25M sucrose, pH 7.5)
3. Homogenize flies using the Kontes homogenizer until the flies are disrupted and only wings are visible. Vortex the sample as needed to dislodge flies from the bottom.
4. Centrifuge the homogenate for 15 min at1200xg and 0°C and collect the supernatant.
5. Centrifuge the supernatant from step 4 at 6000xg and 0°C.
6. Filter through a 0.45µm spin column.
    a. Add up to 800 µl of supernatant to the column.
    b. Spin at 1000 x g for 1 min.
    c. Collect filtrate.
    d. Homogenate that does not pass through the column should be transferred into a clean column and re-spun.
    e. Aliquot the filtrate into 100 µl tubes (30 µl per tube).
    f. Store at -80°C.

Artificial Injection

1. The microinjector tubing, needle holder, and the needle were back-filled physiological grade mineral oil (Sigma M8410-100ML; lot #019K0125). The system was ready for injection when the oil meniscus within the needle moved without lag, which would have suggested an air bubble (when an air bubble is present in the system the meniscus will not move exactly when dispensing occurs and the meniscus will continue to move after the mechanical dispensing

action has ceased. This is because of the bubble contracting and then relaxing slowly.

2. The inoculum was thawed on ice (and kept on ice) and the needle was back-filled. The tip of the needle was placed into the inoculum and the plunger on the syringe that was mounted on the microinjector was reversed.

3. An anesthetized fly was placed on the microscope stage with its head facing the person injection and its legs downward. This facilitated inserting the needle between the second and fourth abdominal tergites and dispensing 0.1 µl of inoculum. After the inoculum was dispensed the needle was left in place for about 5 seconds. In preliminary injections it was found that when the needle is removed immediately after the inoculum is dispensed some of the inoculum would “leak” out of the fly.

Fecundity and Hatchability

Fecundity and hatchability were done using food tinted with green dye that was poured directly onto a standard glass slide to produce a thin layer of food over one half of the slide. The females were allowed to oviposit, the eggs were counted and then incubated at room temperature for 24 hours. Then the eggs that failed to hatch were counted.

1. The food was prepared as detailed above.

2. Green food dye (McCormick brand) from local grocery store was added (9-10 drops per 100ml of food.

3. The food was dispensed using a repeating pipette and corresponding combitips (www.eppendorf.com). One milliliter of food was dispensed at one end of each glass slide.

4. The food was allowed to set (~60 min), covered in plastic wrap and stored at 4°C until its use.

5. On the day of use the slides were labeled and sprayed with a 4g/L yeast solution.

6. The slides were placed into a large polystyrene *Drosophila* vial that had been prepared earlier by the addition of 2ml of 2% agar (as the bottoms of the vials are not flat, this was necessary to prevent the flies from moving to the side of the slide that did not have food).

7. One male and one female were added to the vial on the side of the slide that contained the food.

8. The vials were kept at standard rearing conditions (see above).

9. The slide was replaced every twelve hours for up to three days.

10. The eggs on the slides removed from the vial were counted (fecundity measure).

11. The slides were sprayed with the yeast solution to keep them moist, lightly covered with plastic wrap and kept at the room temperature for 24hrs.

12. After 24hrs the slides were kept at 4°C and the unhatched eggs were counted as time permitted within the week following the experiment [(total eggs-unhatched eggs)/total eggs = percent hatchability].

LIST OF REFERENCES

ALIZON, S., 2008 Transmission-recovery trade-offs to study parasite evolution. Am Nat **172:** E113-E121.

ALIZON, S., A. HURFORD, N. MIDEO and M. VAN BAALEN, 2009 Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. J Evol Biol **22:** 245-259.

ANDERSON, R. M., and R. M. MAY, 1979 Population biology of infectious diseases. Nature **280:** 361-367.

ANDERSON, R. M., and R. M. MAY, 1982 Coevolution of hosts and parasites. Parasitology **85:** 411-426.

ANDRE, J.-B., S. GANDON and J. KOELLA, 2006 Vaccination, within-host dynamics, and virulence evolution. Evolution **60:** 13-23.

ANDRE, J. B., and M. E. HOCHBERG, 2005 Virulence evolution in emerging infectious diseases. Evolution **59:** 1406-1412.

ANTONOVICS, J., M. HOOD and J. PARTAIN, 2002 The ecology and genetics of a host shift: Microbotryum as a model system. Am Nat **160:** S40-S53.

BAIGENT, S. J., and J. W. MCCAULEY, 2003 Influenza type A in humans, mammals and birds: Determinants of virus virulence, host-range and interspecies transmission. Bioessays **25:** 657-671.

BARBOZA, P., A. TARANTOLA, L. LASSEL, T. MOLLET, I. QUATRESOUS *et al.*, 2008 Viroses émergentes en Asie du Sud-Est et dans le Pacifique. Méd Maladies Infect **38:** 513-523.

BASS, B. L., 2002 RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem **71:** 817-846.

BASS, B. L., K. NISHIKURA, W. KELLER, P. H. SEEBURG, R. B. EMESON *et al.*, 1997 A standardized nomenclature for adenosine deaminases that act on RNA. RNA **3:** 947-949.

BEACH, N. M., R. B. DUNCAN, C. T. LARSEN, X. J. MENG, N. SRIRANGANATHAN *et al.*, 2009 Persistent infection of turkeys with an avirulent strain of turkey hemorrhagic enteritis virus. Avian Dis **53:** 370-375.

BEATY, B. J., and W. C. MARQUARDT, 1996 *The biology of disease vectors*. University Press of Colorado, Niwot, Co.

BENAGLIO, P., and C. RIVOLTA, 2010 Ultra high throughput sequencing in human DNA variation setection: a comparative study on the NDUFA3-PRPF31 region. PLoS ONE **5**.

BOLKER, B. M., A. NANDA and D. SHAH, 2010 Transient virulence of emerging pathogens. J Roy Soc Interface **7:** 811-822.

BRUN, G., and N. PLUS, 1980 The viruses of *Drosophila*, pp. 625-702 in *The genetics and biology of drosophila*, edited by M. ASHBURNER and T. R. F. WRIGHT. Academic Press, London.

BRUSINI, J., Y. WANG, L. F. MATOS, L.-S. SYLVESTRE, B. M. BOLKER *et al.*, 2012 in review Virulence evolution in a host-parasite system in the absence of viral evolution. Am Nat.

BULL, J. J., 1994 Perspective - Virulence. Evolution **48:** 1423-1437.

BULL, J. J., M. R. BADGETT, H. A. WICHMAN, J. P. HUELSENBECK, D. M. HILLIS *et al.*, 1997 Exceptional convergent evolution in a virus. Genetics **147:** 1497-1507.

BULL, J. J., and S. P. OTTO, 2005 The first steps in adaptive evolution. Nat Genet **37:** 342-343.

BULLER, R., B. ARIF, D. BLACK, K. DUMBELL, J. ESPOSITO *et al.*, 2005 Poxviridae in *Virus Taxonomy, The VIIIth Report of the International Committee on Taxonomy of Viruses. (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fr-fst-g.htm.).* edited by M. M. FAUQUET CM, MANILOFF J, DESSELBERGER U, BALL LA. Elsevier Academic Press, Oxford, UK.

BURCH, C. L., and L. CHAO, 1999 Evolution by small steps and rugged landscapes in the RNA virus phi 6. Genetics **151:** 921-927.

CALVIGNAC-SPENCER, S., S. A. J. LEENDERTZ, T. R. GILLESPIE and F. H. LEENDERTZ, 2012 Wild great apes as sentinels and sources of infectious disease. Clin Microbiol Infect **18:** 521-527.

CARPENTER, J., 2008 The evolution and genetics of Drosophila melanogaster and the sigma virus, pp. The University of Edinburgh.

CARPENTER, J., L. KEEGAN, L. WILFERT, M. O'CONNELL and F. JIGGINS, 2009 Evidence for ADAR-induced hypermutation of the Drosophila sigma virus (Rhabdoviridae). BMC Genetics **10:** 75.

CARPENTER, J. A., D. J. OBBARD, X. MASIDE and F. M. JIGGINS, 2007 The recent spread of a vertically transmitted virus through populations of *Drosophila melanogaster*. Mol. Ecol. **16:** 3947-3954.

CHAPUIS, É., S. PAGÈS, V. EMELIANOFF, A. GIVAUDAN and J.-B. FERDY, 2011 Virulence and pathogen multiplication: a serial passage experiment in the hypervirulent bacterial insect-pathogen *Xenorhabdus nematophila*. PLoS ONE **6:** e15872.

CLAAS, E. C. J., A. OSTERHAUS, R. VAN BEEK, J. C. DE JONG, G. F. RIMMELZWAAN *et al.*, 1998 Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. Lancet **351:** 472-477.

CLARK, M. A., W. B. MCCRADY and C. L. FIELDING, 1979 Independent production of CO_2 sensitivity by nuclear gene-Dly and a transmissible agent from delayed-recovery *Drosophila melanogaster*. Genetics **92:** 503-510.

COLLIGNON, P. J., 2010 A pandemic response to a disease of predominantly seasonal intensity. Med J Aust **192:** 358-358.

COMBES, C., 2005 *The art of being a parasite*.

Council of Sci. Eds. Virulent [Internet]. MSN Encarta U.S. English Dictionary. [cited 2011 Jan 28]. Available from: http://encarta.msn.com/dictionary_1861734303/definition.html, pp.

COWPERTHWAITE, M. C., J. J. BULL and L. A. MEYERS, 2006 From Bad to Good: Fitness reversals and the ascent of deleterious mutations. PLoS Comput Biol **2:** e141.

CUEVAS, J. M., S. F. ELENA and A. MOYA, 2002 Molecular basis of adaptive convergence in experimental populations of RNA viruses. Genetics **162:** 533-542.

DARWIN, C. R., 1909-14 The voyage of the beagle, pp. in *The Harvard Classics*, edited by C. E. ELLIOT. P.F. Collier & Sons. www.bartelby.com/29/. (Jan. 26, 2011), New York, NY.

DENNEHY, J. J., 2009 Bacteriophages as model organisms for virus emergence research. Trends Microbiol **17:** 450-457.

DENNEHY, J. J., N. A. FRIEDENBERG, R. D. HOLT and P. E. TURNER, 2006 Viral ecology and the maintenance of novel host use. Am Nat **167:** 429-439.

DENNEHY, J. J., N. A. FRIEDENBERG, Y. W. YANG and P. E. TURNER, 2007 Virus population extinction via ecological traps. Ecol Lett **10:** 230-240.

DOLGIN, E., 2010 Better forecasting urged to avoid drug waste. Nat Med **16:** 5-5.

DOMINGO, E., and J. J. HOLLAND, 1997 RNA virus mutations and fitness for survival. Annu Rev Microbiol **51:** 151-178.

EBERT, D., 1998 Experimental evolution of parasites. Science **282:** 1432-1436.

EBERT, D., and J. J. BULL, 2003 Challenging the trade-off model for the evolution of virulence: is virulence management feasible? Trends Microbiol **11:** 15-20.

EWALD, P. W., 1983 Host-parasite relations, vectors, and the evolution of disease severity. Annu Rev Ecol Syst **14:** 465-485.

FAUQUET CM, M. M., MANILOFF J, DESSELBERGER U, BALL LA, 2005 *Virus Taxonomy, The VIIIth Report of the International Committee on Taxonomy of Viruses. (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fr-fst-g.htm.).* Elsevier Academic Press, Oxford, UK.

FINE, P. E. M., 1975 Vectors and vertical transmission: epidemiologic perspective. Ann N Y Acad Sci **266:** 173-194.

FISHER, R. A., 1930 *The genetical theory of natural selection*. Oxford University Press, Oxford., Dover, New York.

FLEURIET, A., 1981a Comparison of various physiological traits in flies (*Drosophila melanogaster*) of wild origin, infected or uninfected by the hereditary rhabdovirus sigma. Arch Virol **69:** 261-272.

FLEURIET, A., 1981b Effect of overwintering on the frequency of flies infected by the rhabdovirus sigma in experimental populations of *Drosophila melanogaster*. Arch Virol **69:** 253-260.

FLEURIET, A., 1982 Transmission efficiency of the sigma virus in natural populations of its host, *Drosophila melanogaster*. Arch Virol **71:** 155-167.

FLEURIET, A., 1988 Maintenance of a hereditary virus: the sigma virus in populations of its host, *Drosophila melanogaster*. Evol Biol **23:** 1-30.

FLEURIET, A., 1996 Polymorphism of the *Drosophila melanogaster* - sigma virus system. J Evol Biol **9:** 471-484.

FLYNN, P., 2010 The Handling of the H1N1 Pandemic: More Transparency Needed, pp. in *Parliamentary Assembly of the Council of Europe. http://assembly.coe.int/CommitteeDocs/2010/20100329_MemorandumPandemie_E.pdf.*

FRANK, S. A., 1992 A kin selection model for the evolution of virulence. Proc R Soc Lond, Ser B: Biol Sci **250:** 195-197.

FRANK, S. A., 1996 Models of parasite virulence. Q Rev Biol **71:** 37-78.

FRANK, S. A., and P. SCHMID-HEMPEL, 2008 Mechanisms of pathogenesis and the evolution of parasite virulence. J Evol Biol **21:** 396-404.

FRASER, C., C. A. DONNELLY, S. CAUCHEMEZ, W. P. HANAGE, M. D. VAN KERKHOVE *et al.*, 2009 Pandemic potential of a strain of influenza A (H1N1): early findings. Science **324:** 1557-1561.

GAO, F., L. YUE, A. T. WHITE, P. G. PAPPAS, J. BARCHUE *et al.*, 1992 Human infection by genetically diverse SIV_{sm}-related HIV-2 in west Africa. Nature **358:** 495-499.

GENTON, B., I. BETUELA, I. FELGER, F. AL-YAMAN, R. F. ANDERS *et al.*, 2002 A recombinant blood-stage malaria vaccine reduces *Plasmodium falciparum* density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. J Infect Dis **185:** 820-827.

GEORGE, C. X., and C. E. SAMUEL, 2011 Host response to polyomavirus infection is modulated by RNA adenosine deaminase ADAR1 but not by ADAR2. J Virol **85:** 8338-8347.

GODDARD, L. B., A. E. ROTH, W. K. REISEN and T. W. SCOTT, 2003 Vertical transmission of West Nile virus by three California *Culex* (Diptera: Culicidae) species. J. Med. Entomol. **40:** 743-746.

GOULD, S. J., 1990 *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton.

GOULD, S. J., and R. C. LEWONTIN, 1979 Spandrels of San-Marco and the panglossian paradigm: a critique of the adaptationist program. Proc R Soc Lond Ser B-Biol Sci **205:** 581-598.

HARISMENDY, O., P. C. NG, R. L. STRAUSBERG, X. Y. WANG, T. B. STOCKWELL *et al.*, 2009 Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol **10**.

HOGENHOUT, S. A., M. G. REDINBAUGH and E.-D. AMMAR, 2003 Plant and animal rhabdovirus host range: a bug's view. Trends Microbiol **11:** 264-271.

HOLMES, E. C., and A. RAMBAUT, 2004 Viral evolution and the emergence of SARS coronavirus. Philos Trans R Soc Lond, Ser B: Biol Sci **359:** 1059-1065.

ISAACS, D., 2010 Lessons from the swine flu: Pandemic, panic and/or pandemonium? J Paediatr Child H **46:** 623-626.

JONES, K. E., N. G. PATEL, M. A. LEVY, A. STOREYGARD, D. BALK *et al.*, 2008 Global trends in emerging infectious diseases. Nature **451:** 990-U994.

KEEGAN, L. P., A. GALLO and M. A. O'CONNELL, 2001 The many roles of an RNA editor. Nat Rev Genet **2:** 869-878.

KELLY, H. A., 2010 A pandemic response to a disease of predominantly seasonal intensity. Med J Aust **192:** 81-83.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624-626.

KIRCHER, M., and J. KELSO, 2010 High-throughput DNA sequencing - concepts and limitations. Bioessays **32:** 524-536.

KNIPE, D. M., and P. M. HOWLEY (Editors), 2001 *Fundamental virology*. Raven Publishers, Ltd., New York.

KOMURIAN-PRADEL, F., M. PERRET, B. DEIMAN, M. SODOYER, V. LOTTEAU *et al.*, 2004 Strand specific quantitative real-time PCR to study replication of hepatitis C virus genome. J Virol Methods **116:** 103-106.

KUIKEN, T., E. C. HOLMES, J. MCCAULEY, G. F. RIMMELZWAAN, C. S. WILLIAMS *et al.*, 2006 Host species barriers to influenza virus infections. Science **312:** 394-397.

L'HERITIER, P. H., 1958 The hereditary virus of *Drosophila*. Adv Virus Res **5:** 195-245.

L'HERITIER, P. H., and F. HUGON DE SCOEUX, 1947 Transmission par greffe et injection de la sensibilite herditaire au gaz carbonique chez la *Drosophile*. Bull Biol Fr Belg **81:** 70-91.

LANDER, E., and M. WATERMAN, 1988 Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics **2:** 231-239.

LAWRENCE, P. O., and L. F. MATOS, 2005 Transmission of the *Diachasmimorpha longicaudata* rhabdovirus (DlRhV) to wasp offspring: an ultrastructural analysis. J. Insect Physiol. **51:** 235-241.

LEHMANN, K. A., and B. L. BASS, 2000 Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry **39:** 12875-12884.

LENDERBERGER ET AL., 1992 Emerging infections: Microbial threats to health in the United States, pp. Institute of Medicine, Washington.

LENSKI, R. E., and R. M. MAY, 1994 The evolution of virulence in parasites and pathogens - reconciliation between two competing hypotheses. J. Theor. Biol. **169:** 253-265.

LEVIN, B. R., 1996 The evolution and maintenance of virulence in microparasites. Emerg Infect Dis **2:** 93-102.

LEVIN, B. R., and J. J. BULL, 1994 Short-sighted evolution and the virulence of pathogenic microorganisms. Trends Microbiol. **2:** 76-81.

LEVIN, B. R., and C. S. EDEN, 1990 Selection and evolution of virulence in bacteria: an ecumenical excursion and modest suggestion. Parasitol **100:** S103-S115.

LEVIN, B. R., M. LIPSITCH and S. BONHOEFFER, 1999 Population Biology, Evolution, and Infectious Disease: Convergence and Synthesis. Science **283:** 806-809.

LIPSITCH, M., M. A. NOWAK, D. EBERT and R. M. MAY, 1995 The population dynamics of vertically and horizontally transmitted parasites. Proc R Soc Lond, Ser B: Biol Sci **260:** 321-327.

LIPSITCH, M., S. SILLER and M. A. NOWAK, 1996 The evolution of virulence in pathogens with vertical and horizontal transmission. Evolution **50:** 1729-1741.

LITTLE, T. J., D. M. SHUKER, N. COLEGRAVE, T. DAY and A. L. GRAHAM, 2010 The coevolution of virulence: tolerance in perspective. Plos Pathog **6:** e1001006.doi:1001010.1001371/journal.ppat.1001006.

LONGDON, B., J. D. HADFIELD, C. L. WEBSTER, D. J. OBBARD and F. M. JIGGINS, 2011 Host phylogeny determines viral persistence and replication in novel hosts. Plos Pathogens **7**.

LUSTIG, S., U. OLSHEVSKY, D. BEN-NATHAN, B. E. LACHMI, M. MALKINSON *et al.*, 2000 A live attenuated West Nile virus strain as a potential veterinary vaccine. Viral Immunol **13:** 401-410.

MACKENZIE, J. S., K. B. CHUA, P. W. DANIELS, B. T. EATON, H. E. FIELD *et al.*, 2001 Emerging viral diseases of Southeast Asia and the Western Pacific. Emerging Infect Dis **7:** 497-504.

MACKINNON, M. J., S. GANDON and A. F. READ, 2008 Virulence evolution in response to vaccination: The case of malaria. Vaccine **26:** C42-C52.

MATROSOVICH, M., N. ZHOU, Y. KAWAOKA and R. WEBSTER, 1999 The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. J Virol **73:** 1146-1155.

MAY, R. M., S. GUPTA and A. R. MCLEAN, 2001 Infectious disease dynamics: what characterizes a successful invader? Philos Trans R Soc Lond, Ser B: Biol Sci **356:** 901-910.

MEYERS, L. A., B. R. LEVIN, A. R. RICHARDSON and I. STOJILJKOVIC, 2003 Epidemiology, hypermutation, within-host evolution and the virulence of *Neisseria meningitidis*. Proc R Soc Lond, Ser B: Biol Sci **270:** 1667-1677.

MOERDYK-SCHAUWECKER, M., S. I. HWANG and V. Z. GRDZELISHVILI, 2009 Analysis of virion associated host proteins in vesicular stomatitis virus using a proteomics approach. Virol J **6**.

MORITZ, R. F. A., C. W. W. PIRK, H. R. HEPBURN and P. NEUMANN, 2008 Short-sighted evolution of virulence in parasitic honeybee workers (Apis mellifera capensis Esch.). Naturwissenschaften **95:** 507-513.

MOYA, A., S. F. ELENA, A. BRACHO, R. MIRALLES and E. BARRIO, 2000 The evolution of RNA viruses: A population genetics view. Proc Natl Acad Sci U S A **97:** 6967-6973.

MOYA, A., E. C. HOLMES and F. GONZALEZ-CANDELAS, 2004 The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol **2:** 279-288.

MUELLER, S., D. PAPAMICHAIL, J. R. COLEMAN, S. SKIENA and E. WIMMER, 2006 Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol **80:** 9687-9696.

MULLER, M. P., and A. MCGEER, 2007 Severe acute respiratory syndrome (SARS) coronavirus. Semin Respir Crit Care **28:** 201-212.

NISHIKURA, K., 2010 Functions and regulation of RNA editing by ADAR deaminases, pp. 321-349 in *Annu Rev Biochem*, edited by R. D. KORNBERG, C. R. H. RAETZ, J. E. ROTHMAN and J. W. THORNER.

OHARA, P. J., S. T. NICHOL, F. M. HORODYSKI and J. J. HOLLAND, 1984 Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. Cell **36:** 915-924.

ONDOV, B. D., A. VARADARAJAN, K. D. PASSALACQUA and N. H. BERGMAN, 2008 Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications. Bioinformatics **24:** 2776-2777.

ORR, H. A., 2005 The genetic theory of adaptation: a brief history. Nat Rev Genet **6:** 119-127.

PALLADINO, M. J., L. P. KEEGAN, M. A. O'CONNELL and R. A. REENAN, 2000a dADAR, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. RNA **6:** 1004-1018.

PALLADINO, M. J., L. P. KEEGAN, M. A. O'CONNELL and R. A. REENAN, 2000b A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. Cell **102:** 437-449.

PERLMAN, S. J., and J. JAENIKE, 2003 Infection success in novel hosts: An experimental and phylogenetic study of *Drosophila*-parasitic nematodes. Evolution **57:** 544-557.

PETSCHEK, J. P., M. J. MERMER, M. R. SCHECKELHOFF, A. A. SIMONE and J. C. VAUGHN, 1996 RNA editing in *Drosophila* 4f-rnp gene nuclear transcripts by multiple A-to-G conversions. J Mol Biol **259:** 885-890.

PIALOUX, G., B. A. GAUZERE, S. JAUREGUIBERRY and M. STROBEL, 2007 Chikungunya, an epidemic arbovirosis. Lancet Infect Dis **7:** 319-327.

PIZZATTO, L., C. M. SHILTON and R. SHINE, 2010 Infection dynamics of the lungworm *Rhabdias pseudosphaerocephala* in its natural host, th ecan toad (Bufo marinus), and in novel hosts (Native Australian frogs). J Wild Dis **46:** 1152-1164.

PURCELL, M. K., S. A. HART, G. KURATH and J. R. WINTON, 2006 Strand-specific, real-time RT-PCR assays for quantification of genomic and positive-sense RNAs of the fish rhabdovirus, Infectious hematopoietic necrosis virus. J Virol Methods **132:** 18-24.

RAMBAUT, A., D. POSADA, K. A. CRANDALL and E. C. HOLMES, 2004 The causes and consequences of HIV evolution. Nat Rev Genet **5:** 52-61.

ROBERTS, D. B., 2006 *Drosophila melanogaster*: the model organism. Entomol Exp Appl **121:** 93-103.

ROKYTA, D. R., P. JOYCE, S. B. CAUDLE and H. A. WICHMAN, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat Genet **37:** 441-444.

ROOSSINCK, M. J., 2010 Lifestyles of plant viruses. Philos Trans R Soc Lond, Ser B: Biol Sci **365:** 1899-1905.

ROSE, J. K., and M. A. WHITT, 2001 Rhabdoviridae: the viruses and their replication, pp. 665-688 in *Fundamental virology*, edited by D. M. KNIPE and P. M. HOWLEY. Lippincott Williams & Wilkins, Philadelphia, PA.

SALVAUDON, L., V. HERAUDET and J. A. SHYKOFF, 2007 Genotype-specific interactions and the trade-off between host and parasite fitness. Bmc Evolutionary Biology **7**.

SAMUEL, C. E., 2011 Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. Virology **411:** 180-193.

SCHNEIDER, D., 2000 Using *Drosophila* as a model system. Nat Rev Genet **1:** 218-226.

SCULLY, L. R., and M. J. BIDOCHKA, 2006 Developing insect models for the study of current and emerging human pathogens, pp. 1-9 in *FEMS Microbiol Lett*.

SHARON, L. M., I. J. MOLINEUX and J. J. BULL, 1999 Virulence evolution in a virus obeys a trade-off. Proc R Soc Lond, Ser B: Biol Sci **266:** 397-404.

STURTEVANT, A. H., 1920 Genetic studies on *Drosophila simulans*. I. Introduction. Hybrids with *Drosophila melanogaster*. Genetics **5:** 488-500.

TENINGES, D., F. BRAS and S. DEZÉLÉE, 1993 Genome organization of the sigma rhabdovirus: six genes and a gene overlap. Virology **193:** 1018-1023.

TSUKAMOTO, K., T. IMADA, N. TANIMURA, M. OKAMATSU, M. MASE *et al.*, 2007 Impact of different husbandry conditions on contact and airborne transmission of H5N1 highly pathogenic avian influenza virus to chickens. Avian Dis. **51:** 129-132.

VAN DER MEULEN, K. M., M. B. PENSAERT and H. J. NAUWYNCK, 2005 West Nile virus in the vertebrate world. Arch Virol **150:** 637-657.

VERMEIJ, G. J., 2006 Historical contingency and the purported uniqueness of evolutionary innovations. Proc Natl Acad Sci U S A **103:** 1804-1809.

VODOVAR, N., C. ACOSTA, B. LEMAITRE and F. BOCCARD, 2004 *Drosophila*: a polyvalent model to decipher host-pathogen interactions. Trends Microbiol **12:** 235-242.

WAHL, L. M., and D. C. KRAKAUER, 2000 Models of experimental evolution: the role of genetic chance and selective necessity. Genetics **156:** 1437-1448.

WARD, S. V., C. X. GEORGE, M. J. WELCH, L.-Y. LIOU, B. HAHM *et al.*, 2011 RNA editing enzyme adenosine deaminase is a restriction factor for controlling measles virus replication that also is required for embryogenesis. PNAS **108:** 331-336.

WAYNE, M. L., G. M. BLOHN, M. E. BROOKS, K. L. REGAN, B. Y. BROWN *et al.*, 2011 The prevalence and persistence of sigma virus, a biparentally transmitted parasite of *Drosophila melanogaster*. Evol Ecol Res **13:** tbd.

WICHMAN, H. A., M. R. BADGETT, L. A. SCOTT, C. M. BOULIANNE and J. J. BULL, 1999a Different trajectories of parallel evolution during viral adaptation. Science **285:** 422-424.

WICHMAN, H. A., M. R. BADGETT, L. A. SCOTT, C. M. BOULIANNE and J. J. BULL, 1999b Different trajectories of parallel evolution during viral adaptation. Science **285:** 422-424.

WOLFE, N. D., C. P. DUNAVAN and J. DIAMOND, 2007 Origins of major human infectious diseases. Nature **447:** 279-283.

WOODS, R., D. SCHNEIDER, C. L. WINKWORTH, M. A. RILEY and R. E. LENSKI, 2006 Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci U S A **103:** 9107-9112.

YAMPOLSKY, L. Y., C. T. WEBB, S. A. SHABALINA and A. S. KONDRASHOV, 1999 Rapid accumulation of a vertically transmitted parasite triggered by relaxation of natural selection among hosts. Evol Ecol Res **1:** 581-589.

YANG, H. P., and S. V. NUZHDIN, 2003 Fitness costs of Doc expression are insufficient to stabilize its copy number in *Drosophila melanogaster*. Mol Biol Evol **20:** 800-804.

BIOGRAPHICAL SKETCH

Luis Fernando Matos was born in the Azores Islands (Portugal) and immigrated with his family to California when he was thirteen. Luis attended Chatom Junior High, where his English speaking skills improved significantly, Turlock High School, Modesto Junior College, and the California State University at Stanislaus where he studied biology and in 1999 became the first member of his extended family to earn a Bachelors degree.

Luis entered the masters of entomology program at Washington State University (Pullman, WA) where he examined the interaction between a pestiferous weed and one of its biological control agents. He created an artificial diet for the immature stage of this insect and developed mass rearing protocols using said diet. Luis received his Masters in entomology in May of 2002. During his time in Pullman Luis met Joanna Joyner and they were married in the summer of 2002.

Luis entered the doctoral program in entomology at the University of Florida soon after. His dissertation research focused on understanding the evolution of virulence following host shifts using a new model system that he helped develop. During his program Luis had the opportunity to participate in an NSF-funded GK-12 program for graduate students. This program convinced Luis that he wanted to teach (and conduct research) at a small university where he could apply the inquire-based skills he acquired in during this NSF training. Additionally, Luis became the father of Lucas David Matos in November of 2006 and successfully fought testicular cancer into remission during 2008. He successfully defended his dissertation in November of 2012.