Molecular Ecology (2007) **16**, 3947–3954 doi: 10.1111/j.1365-294X.2007.03460.x

The recent spread of a vertically transmitted virus through populations of *Drosophila melanogaster*

JENNIFER A. CARPENTER,* DARREN J. OBBARD,* XULIO MASIDE† and FRANCIS M. JIGGINS*
**Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK, †Grupo de Medicina Xenómica, Instituto de Medicina Legal, Universidade de Santiago de Compostela, 15782 Galicia, Spain*

Abstract

The sigma virus is a vertically transmitted pathogen that commonly infects natural populations of *Drosophila melanogaster*. This virus is the only known host-specific pathogen of *D. melanogaster*, and so offers a unique opportunity to study the genetics of *Drosophila*–viral interactions in a natural system. To elucidate the population genetic processes that operate in sigma virus populations, we collected *D. melanogaster* from 10 populations across three continents. We found that the sigma virus had a prevalence of 0–15% in these populations. Compared to other RNA viruses, we found that levels of viral genetic diversity are very low across Europe and North America. Based on laboratory measurements of the viral substitution rate, we estimate that most European and North American viral isolates shared a common ancestor approximately 200 years ago. We suggest two explanations for this: the first is that *D. melanogaster* has recently acquired the sigma virus; the second is that a single viral type has recently swept through *D. melanogaster* populations. Furthermore, in contrast to *Drosophila* populations, we find that the sigma viral populations are highly structured. This is surprising for a vertically transmitted pathogen that has a similar migration rate to its host. We suggest that the low structure in the viral populations can be explained by the smaller effective population size of the virus.

Keywords: co-evolution, *Drosophila*, population structure, rhabdovirus, sigma virus

Received 10 January 2007; revision received 15 May 2007; accepted 11 Jun 2007

Introduction

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

The interactions between viruses and their hosts have been largely under-exploited by *Drosophila* biologists studying immune systems (Cherry & Silverman 2006). Seven RNA viruses have been isolated in natural populations of *D. melanogaster* (Berkaloff 1965; Plus & Duthoit 1969; Teninges & Plus 1972; Plus *et al*. 1976; Habayeb *et al*. 2006). These RNA viruses are common in flies — 40% of flies in natural populations and 47% of the laboratory stocks are found to be infected with at least one RNA virus (Brun & Plus 1998).

We are studying the sigma virus, the only known host-specific pathogen of *D. melanogaster*, which offers a unique opportunity to study a natural *Drosophila*–virus interaction. Sigma is a rhabdovirus (negatively sensed single-stranded RNA virus), that is widespread in natural populations and causes death of infected flies after exposure to CO_2

Correspondence: Jennifer A. Carpenter, School of Biological Sciences, Institute of Evolutionary Biology, The University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh EH9 3JT, Scotland, UK. Fax: +44 (0) 131 650 6564;
E-mail: jennifer.carpenter@ed.ac.uk

Table 1 Incidence of CO_2 sensitivity in wild-caught lines of *Drosophila melanogaster*

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

Incidence of CO_2 sensitivity in wild-caught lines of *Drosophila simulans*

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

(L'heritier & Teissier 1937). It is vertically transmitted through both sperm and eggs. It does not integrate into the fly's chromosomes but exists in the cytoplasm of the cell, and, like many selfish genetic elements, is passed from parent to offspring at a greater frequency than expected under Mendelian inheritance (Doolittle *et al*. 1984; Brun & Plus 1998). The costs of infection by the sigma virus include reduced viability of infected eggs (Fleuriet 1981a) and lower survival by infected flies over winter (Fleuriet 1981b).

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

Materials and methods

Collection

Sigma-virus isolates were collected by isolating single female *Drosophila* from wild populations, allowing them to lay in a vial on standard *Drosophila* food, and exposing a proportion (50%) of their offspring to pure CO_2 at 12 °C for 15 minutes. The flies were then allowed to recover at room temperature. Infected flies are sensitive to CO_2 — they died or become severely paralysed in contrast to uninfected flies that fully recover (L'Heritier & Teissier 1937). Only CO_2-sensitive lines were kept and each isofemale line is termed a viral isolate. The flies were collected from vineyards and fruit farms from a number of locations (see Table 1). Flies were collected over 1–7 days from patches of fruit within 1–2 miles of each other. D. Contamine supplied two viral lines (A3 and A3E55) that had been maintained in the laboratory. These viral lines are a single wild-collected isolate which was split between two separate fly lines and maintained at 20 °C. Unfortunately, because of lost records, the precise age at which the lines were split is unknown; however, we know it is more than 10 years and less than 20 years.

Sequencing and sequence analysis

Polymerase chain reaction (PCR) primers were designed from the published sigma sequence from GenBank (×91062) using the program PRIMER 3. We sequenced one large fragment from three wild-collected isolates and two viral lines maintained in the laboratory, and two shorter fragments from a larger sample of viral isolates from five wild-collected populations (see Fig. 1). Both viral and fly RNA were extracted using a Total RNA Isolation System (Promega). The PCR primers were used to reverse-transcribe the genome using M-MLV reverse transcriptase. Fragments were then amplified by PCR. Before sequencing, unused PCR primers and dNTPs were digested with exonuclease 1 and shrimp alkaline phosphatase. The PCR products were then sequenced directly using the PCR

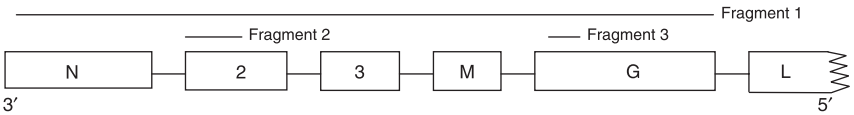

Fig. 1 Schematic of the sigma virus. The six genes shown encode the following proteins: N, nucleocapsid protein; 2, polymerase-associated protein; 3, PP3; M, matrix protein; G, outer-coat protein; and L, polymerase protein. The L gene is incomplete. The first fragment we sequenced covers five genes; the second fragment covers 67% of the polymerase associated gene (636 bp); and the third covers 36% of the outer-coat protein (589 bp).

primers and Big Dye reagents (ABI) on an ABI capillary sequencer. In cases where PCR products could not be sequenced directly, fragments were cloned using the TOPO TA Cloning kit (Invitrogen). Multiple clones were mixed to avoid PCR errors and sequenced. Sequences were initially assembled using SEQUENCHER 4.5 (Gene Codes Corporations) and chromatograms were inspected by eye to confirm the legitimacy of polymorphisms between viral lines. The sequences were then aligned using CLUSTAL W and genes were identified with reference to published sigma sequences.

Estimation of mutation rates and phylogenetic reconstruction

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

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

We performed three tests for recombination within the sigma virus genome. First, we tested for recombination between the sequenced genes by constructing maximum-likelihood trees separately for each gene, and then forcing each gene to take the topology of the other gene. Recombination was indicated if the forced topology had a significantly lowered likelihood relative to the gene's own maximum-likelihood tree using a one-tailed Shimodaira-Hasegawa likelihood-ratio test (Shimodaira & Hasegawa 1999). Second, we also tested for recombination using the maximum chi-squared test (Maynard-Smith 1992) and the reticulate test (Jakobsen & Easteal 1996), which were performed in the program RDP (Martin & Rybicki 2000). The maximum chi-squared test identifies potential recombinant events between two sequences and a putative-derived sequence. We used a sliding-window analysis (50-bp, 1-bp steps). At each step, the number of variable sites was compared in the left and right halves of the window using chi-squared test. Potential breakpoints correspond to peaks in the values of chi-squared. Third, we used the reticulate test, that identifies regions of sequence within an alignment that have phylogenetic relationships that are incompatible with each other. The test then estimates whether these regions are longer than would be expected by chance. The test statistic is the neighbour similarity score (NSS), which is the average proportion of times a region is compatible (shares a phylogenetic history) with a neighbouring region. The null distributions of both the chi-squared and NSS statistics were generated by recalculating them 10^4 times from data sets where the order of sites had been permuted.

We tested for population structure among the viral isolates using Hudson's (2000) nearest-neighbour statistic (S_{nn}) estimated in DNASP 4.10. To assess the significance of observed S_{nn}, sequences were randomly assigned to localities, maintaining the same number of sequences in each locality as in the original sample. The proportion of permutated samples with S_{nn} larger or equal to the observed value is the estimated P value. We also report average values of K_{ST} an analogue of F_{ST}; (Hudson *et al*. 1992) across all the populations and estimate a P value from the proportion of time that the observed K_{ST} value is greater than an estimated value of K_{ST} based on randomly partitioning the data set among localities.

Results

Viral prevalence

We estimated the prevalence of sigma virus in a population by measuring the proportion of infected isofemale lines established from wild-caught females. The number of

infected individuals varied greatly between populations ($\chi^2 = 47.55$, d.f. = 9, $P < 0.001$). We found the highest prevalence of sigma virus in some of the European populations (0–15%), it was lower in North America (0–6%) and we failed to find sigma virus in African populations (Table 1). The prevalence of sigma virus varied widely, even between neighbouring collection sites: for example, viral prevalence differed dramatically between populations collected at Essex and Kent, not more than 100 km apart (Table 1). The virus was not found in other species of *Drosophila* collected alongside *Drosophila melanogaster* (Table 1).

Recombination

We sequenced two regions from a large sample of viral isolates: the first encompasses the polymerase-associated protein; the second includes the G gene, encoding the outer-coat protein (590 bp and 637 bp, respectively; Fig. 1). The sequences come from virus isolated in *D. melanogaster* lines collected in Europe (Greece, UK and Spain) and North America (Georgia and Florida, USA). These sequences are mostly protein coding but include small intergenic regions. We compared the tree topology for the two sequenced regions and found no significant conflict between genealogies using a one-tailed Shimodaira-Hasegawa likelihood-ratio test ($P > 0.05$) (Shimodaira & Hasegawa 1999). In the absence of recombination, these genes should share evolutionary histories, and so the lack of conflict among the genealogies suggests no recombination has occurred. We further tested for recombination between viral lines using a maximum chi-squared test and found no evidence for any significant breakpoints in our sequences ($P = 0.29$). Further support for a lack of recombination in the viral sequences comes from the reticulate test. An NSS of 0.99 ($P = 1.0$) for the sequences indicates that neighbouring sites share similar phylogenetic histories as often as distant sites, as would be expected when there is no recombination. The lack of recombination between sequences allows the concatenation of the two sequenced regions (the polymerase-associated protein and outer-coat protein), affording a larger data set for phylogenetic analysis.

Viral sequence variation and population structure

We sequenced 1224 bp from between two and ten isolates from each of five populations. We analysed sequence data for levels of variability and found that within-population variation is similar in all populations and is extremely low (Essex: $\pi = 0.00062$; Spain: $\pi = 0.00654$; Greece: $\pi = 0.00059$). We found that one more divergent American isolate, collected in Florida, differs considerably from all other European isolates ($K_s = 0.395$) while the other American isolate, collected in Georgia, is more similar ($K_s = 0.0475$) (Fig. 2).

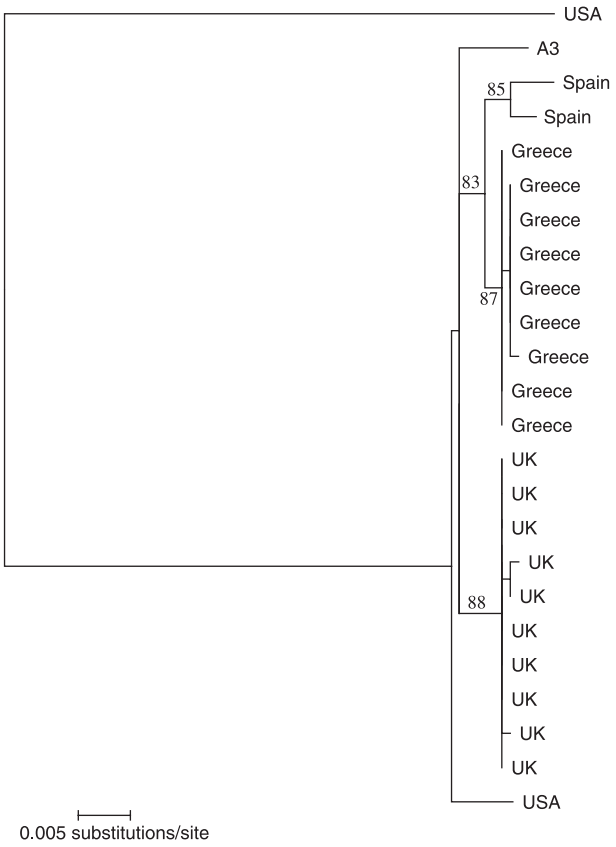

Fig. 2 Maximum-likelihood tree of the concatenated P and G genes of the sigma virus based on all sites. Bootstrap support is shown (1000 reps).

The sigma virus population is highly structured; we found that each European viral population forms a separate monophyletic group (Fig. 2). This is evident in the high value of Hudson's nearest-neighbour statistic (S_{nn}) (Hudson 2000), which measures the proportion of times that nearest neighbours (in terms of genetic distance between sequences) are found in the same population ($S_{nn} = 0.964$; $P < 0.001$). The population structuring is also reflected by high average K_{ST}, which compares genetic diversity within and between populations ($K_{ST}^* = 0.730$, $P < 0.001$) (Hudson *et al*. 1992). American viral isolates were excluded from population structure analysis because of limited sampling.

Estimating substitution rate

We estimated the substitution rate in the laboratory by sequencing 5744 bp of the viral genome from two viral lines (A3 and A3 E55) split from each other and maintained at 20 °C for 10–20 years. We found 25 substitutions between these two viral lines. Using an out-group, we assigned these mutations to either the lineage leading

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

Age of viral spread

The low sequence diversity across all viral populations is consistent with the hypothesis that a single viral type has recently spread through European *D. melanogaster* populations and across the Atlantic to North America. Using the substitution rate determined in the laboratory, we were able to estimate the time since the viral isolates shared a common ancestor. We found that all the viral isolates, except the most divergent Florida isolate, shared a common ancestor 214 years ago (95% CI on the substitution rate estimate: 55 years –1036 years). All sequences collected (including Florida isolate) shared a common ancestor 2106 years ago (95% CI on the substitution rate estimate: 538–10 196 years ago). The genetic distance between sequences was taken from the mean depth of the tree shown in Fig. 2.

Discussion

Low viral diversity

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

The second cause of the low viral diversity is that most of the viral isolates collected shared a common ancestor as recently as 200 years ago. Why do these viral sequences share such a recent ancestor? And why is one of our viral isolates much more divergent than the others? We offer two possible explanations: the first is that *Drosophila melanogaster* has recently acquired the sigma virus; the second is that a single viral type has recently swept through *D. melanogaster* populations.

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

The spread of a recently acquired viral infection through populations of *D. melanogaster* might be analogous with the spread of other parasitic elements through *Drosophila* populations. One example is the recent invasion of P elements — discovered in studies of hybrid dysgenesis — into the *D. melanogaster* genome (Kidwell *et al*. 1977). From

an analysis of P elements in the genus *Drosophila*, it was found that P elements were transmitted from *Drosophila willistoni* to *D. melanogaster*, mostly likely vectored through mites. This transmission could have occurred as recently as *D. melanogaster* (an Old World species) and *D. willistoni* (a New World species) became sympatric only 300 years ago, when *D. melanogaster* was introduced to the New World by humans (Lachaise & Silvain 2004). Another example of horizontal transmission associated with the migration of *D. melanogaster* to the New World is the introduction of *Spiroplasma* — a male-killing bacterium — into *D. melanogaster*. The transmission of the bacterium was mostly likely from *Drosophila nebulosa*, a New World species in the *willistoni* group. In a final example of a parasitic element spreading through *Drosophila* populations, the cytoplasmic incompatibility (CI) causing *Wolbachia* spread through populations of *Drosophila simulans* in western USA over a stretch of 10 years (Turelli & Hoffmann 1991). In this case, long-distance dispersal saw infections spreading at a rate more than 100 km/year to reach levels of 80% infection in areas that previously had none. The emerging picture suggests that host–parasite associations are very dynamic, with parasites constantly gained and lost from host populations.

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

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

Population structure

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

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

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

effective population size is better reflected by the harmonic mean of the prevalence estimates — harmonic means are disproportionately affected by the small values and so better reflect populations that periodically experience very small population sizes. As the harmonic mean is lower than the arithmetic mean, this would tend to further increase levels of structure. Finally, it is worth remembering that we did not sample the virus from the same populations as Caracristi & Schlotterer's (2003) sampled their flies.

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

Sequence availability

The new sequences for the sigma virus have been deposited in GenBank under the following accession nos: AM689308–AM689331.

Acknowledgements

We wish to thank Didier Contamine and Marta Wayne for supplying viral lines and Anastasia Fytrou, Norma Corinne, Sylvain Charlat, Jon Linklater, Edwin Akhusama and Christian Schlötterer for collections. We thank Kelly Dyer and Elie Dolgin for their suggestions and Tom Little and Andrea Betancourt for their valuable critiques during the writing of this article and two anonymous reviewers for their comments. J.A.C. is funded by a Natural Environment Research Council Studentship and a Fieldwork Grant from the Genetics Society, D.J.O. by a Wellcome Trust grant and F.M.J. by a Wellcome Trust Research Career Development Fellowship.

References

Berkaloff A, Bregliano JC, Ohanessian ACR (1965) *Mise en evidence de virions dans des Drosophiles infectees par le virus hereditaires*, pp. 5956–5959. Academic Sciences, Paris.

Brun P, Plus N (1998) The viruses of *Drosophila*, 2d. In: *The Genetics and Biology of Drosophila* (eds Ashburner M, Wright TRF). Academic Press, London.

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

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

Davis PL, Holmes EC, Larrous F *et al.* (2005) Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. *Journal of Virology*, **79**, 10487–10497.

Doolittle WF, Kirkwood TBL, Dempster MAH (1984) Selfish DNAs with self-restraint. *Nature*, **307**, 501–502.

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

Fleuriet A (1981a) Comparison of various physiological traits in flies (*Drosophila melanogaster*) of wild origin, infected or uninfected by the hereditary rhabdovirus sigma. *Archives of Virology*, **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*. *Archives of Virology*, **69**, 253–260.

Fleuriet A (1986) Perpetuation of the hereditary sigma-virus in populations of its host, *Drosophila melanogaster* — geographical analysis of correlated polymorphisms. *Genetica*, **70**, 167–177.

Fleuriet A, Sperlich D (1992) Evolution of The *Drosophila-melanogaster*-sigma virus system in a natural-population from Tubingen. *Theoretical and Applied Genetics*, **85**, 186–189.

Fleuriet A, Periquet G, Anxolabehere D (1990) Evolution of natural-populations in the *Drosophila-melanogaster* sigma virus system. 1. Languedoc (Southern France). *Genetica*, **81**, 21–31.

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

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

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

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

Houck MA, Clark JB, Peterson KR, Kidwell MG (1991) Possible horizontal transfer of *Drosophila* genes by the mite proctolaelaps-regalis. *Science*, **253**, 1125–1129.

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

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, **9**, 138–151.

Hughes GJ, Orciari LA, Rupprecht CE (2005) Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene. *Journal of General Virology*, **86**, 1467–1474.

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

Kidwell MG, Kidwell JF, Sved JA (1977) Hybrid dysgenesis in *Drosophila melanogaster* — syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics*, **86**, 813–833.

L'heritier PH, Teissier G (1937) Une anomolie physiologique hereditaire chez la *Drosophile*. *Comptes Rendus de L'Acedemic Des Sciences*, **205**, 1099–1101.

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

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

Maynard-Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

Plus N, Duthoit JL (1969) Un nouveau virus de *Drosophila melanogaster*, le virus P. *Comptes Rendus de L'acedemic Des Sciences*, **268**, 2313–2315.

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

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences, USA*, **98**, 13757–13762.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Swofford DL (2002) *Phylogenetic Analysis Using Parsimony*. *Version 4.0b10 for Microsoft Windows*. Sinauer & Associates, Sunderland, Massachusetts.

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

Turelli M, Hoffmann AA (1991) Rapid spread of an inherited incompatibility factor in California Drosophila. *Nature*, **353**, 440–442.

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

Jennifer A. Carpenter is interested in host-parasite coevolution, using *Drosophila* and their natural parasites (viruses, fungi and parasitic wasps) to investigate the genetic determination of resistance and virulence. Darren J. Obbard is interested in the molecular basis of adaptation, and the population genetics of host-parasite interaction. Francis M. Jiggins studies the evolutionary genetics of host-parasite interactions in insects. Xulio Maside is interested in molecular evolution, focusing on the study of the evolutionary forces that shape natural genetic variation and the population genetics of transposable elements.
