View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Open Research Exeter

1 Flies on the move: an inherited virus mirrors *Drosophila melanogaster*'s elusive ecology and
2 demography

3

4 Lena Wilfert[1]

5 Francis M. Jiggins[2]

6

7 [1)] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, TR10
8 9FE, United Kingdom

9

10 [2)] Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, United Kingdom

11

12 keywords: migration, overwintering, biogeography, co-evolution

13 Corresponding Author: Lena Wilfert, Tel. +44 (0) 1326370723, Fax ++44 (0) 1326253638:
14 lena.wilfert@ex.ac.uk

15 Running title: Flies on the move

16

17

18

19 Abstract

20 Vertically transmitted parasites rely on their host's reproduction for their transmission,
21 leading to the evolutionary histories of both parties being intimately entwined. Parasites can
22 thus serve as a population genetic magnifying glass for their host's demographic history.
23 Here, we study the fruitfly *Drosophila melanogaster's* vertically transmitted sigma virus
24 DMelSV. The virus has a high mutation rate and low effective population size, allowing us to
25 reconstruct at a fine scale how the combined forces of the movement of flies and selection on
26 the virus have shaped its migration patterns. We found that the virus is likely to have spread
27 to Europe from Africa, mirroring the colonization route of *Drosophila*. The North American
28 DMelSV population appears to be the result of a recent single immigration from Europe,
29 invading together with its host in the late 19th century. Across Europe, DMelSV migration
30 rates are low and populations are highly genetically structured, likely reflecting limited fly
31 movement. Despite being intolerant of extreme cold, viral diversity suggests that fly
32 populations can persist in harsh continental climates and that recolonisation from the warmer
33 south plays a minor role. In conclusion, studying DMelSV can provide insights into the
34 poorly understood ecology of *D. melanogaster,* one of the best-studied organisms in biology.

35 Introduction

36 The genetic structure of parasite populations provides a record of their past transmission
37 routes. This is shaped by the migration and contact patterns of their hosts, together with
38 selection on the parasite genome and other epidemiological processes. Therefore, with care,
39 parasites that are closely associated with their host can be used as proxies for understanding
40 the phylogeographic or demographic history of their hosts (Wirth *et al*. 2005; Nieberding &
41 Olivieri 2007). For example, Biek *et al*. (2006) used feline immune deficiency virus as a
42 marker to infer the recent demographic history of its cougar host. Human history too has been
43 uncovered by parasites – for example, the divergence time of head and body lice has shown
44 that the use of clothing evolved only in the last 100,000 years (72'000 ± 42'000 years, (Kittler
45 *et al*. 2003) whereas the bacterium *Helicobacter pylori* reveals details of recent human
46 migrations, such as during the slave trade or the settling of Polynesia (Falush *et al*. 2003).

47 Key population genetic parameters shaping the genetic structure of parasite populations can
48 differ dramatically from the host, meaning that the spatial and temporal scale at which
49 demographic processes can be detected in the host and parasite populations can be very
50 different (Nieberding & Olivieri 2007). For example, RNA virus genomes evolve faster than
51 their hosts, allowing the reconstruction of very recent events during their evolution
52 (Nieberding & Olivieri 2007). Some parasites also have smaller effective population sizes
53 (N_e), which has the effect of increasing genetic structure within populations (Nieberding &
54 Olivieri 2007). If parasites are to be used to elucidate their host's evolutionary history, they
55 are most useful when they have transmission routes that link them closely to the movement or
56 reproduction of their hosts, such as vertical or sexual transmission (Nieberding & Olivieri
57 2007).

58 The sigma virus DMelSV is a naturally occurring and globally distributed parasite that is
59 specific to *Drosophila melanogaster* (Carpenter *et al*. 2007). This virus is exclusively
60 vertically transmitted to the offspring by both the mother and the father, so its demographic
61 history is intimately linked with – and limited by – that of its host. As a vertically transmitted
62 parasite where there is little genetic variation within hosts and multiple infections are rare (F.
63 Jiggins unpublished data; (Carpenter *et al*. 2007), its effective population size can be
64 estimated in terms of numbers of infected hosts, in a similar way to mitochondrial genes. At
65 its simplest the virus can be treated as a haploid host gene, except that its N_e is likely to be
66 lower because only a minority of flies are infected (prevalences of 0-15% (Carpenter *et al*.
67 2007) and 7.2 -29.7 % (Fleuriet 1976) have been reported from natural populations).
68 Furthermore, as a typical RNA virus, DMelSV's mutation rate is several orders of magnitude
69 higher than that of *Drosophila* (Carpenter *et al*. 2007; Keightley *et al*. 2009). As is common
70 for negative-sense RNA viruses, there is no evidence for genetic recombination between
71 sigma viruses in nature (Carpenter *et al.* 2007), so phylogenetic approaches can be used to
72 investigate its evolution. However, as a pathogen it can be subject to strong selection and has
73 variable prevalence, which will decouple patterns of genetic variation in the host and virus
74 (Fleuriet & Sperlich 1992; Fleuriet & Periquet 1993; Carpenter *et al.* 2007; Wilfert & Jiggins
75 2013).

76 Few organisms have been as well studied as the fruitfly *Drosophila melanogaster*. From the
77 earliest days of the field of genetics, *D. melanogaster* has allowed researchers to gain insights
78 into fields ranging from development to aging, physiology to behaviour, and immunology to
79 sexual selection to name but a few. However, to this day, we know surprisingly little about
80 this species' ecology and natural history in the wild (Keller 2007). Today, *D. melanogaster* is
81 a human commensal with a cosmopolitan distribution. Based on genetic variation at all levels

82 of the genome, it is clear that its origins lie in sub-Saharan Africa (reviewed in David and
83 Capy (1988) and Stephan & Li (2007)). It has been suggested that its dependency on humans
84 dates back as far as 18,000 years (Lachaise & Silvain 2004), predating its move into Europe
85 sometime after the last glaciation which ended about 16,000 years ago. It then spread across
86 Eurasia, but was only able to reach the Americas, Australia and remote islands with the help
87 of man over the last few centuries, with North America in particular being colonised only in
88 the late 19th century by cold-adapted flies from Europe (David & Capy 1988; Keller 2007;
89 Stephan & Li 2007).

90 The contemporary migration patterns of *D. melanogaster* are much less understood. The
91 constant global movement of fruitflies with the food trade and the rapid spread of P-elements
92 in the middle of the last century suggest constant gene flow (David & Capy 1988), yet there
93 also is ample evidence of stable latitudinal and longitudinal genetic variation (Rako *et al.*
94 2009; Fabian *et al.* 2012). Appreciable levels of migration are known to occur between D.
95 melanogaster populations. While direct observations of fruitflies suggest that movement may
96 be limited to just a few hundreds of meters (reviewed by Dobzhansky 1973), Coyne &
97 Milstead (1987) demonstrated that released alleles of *D. melanogaster* spread ~10 km in 3
98 months through a continuously habitable and pre-colonised area, potentially by a combination
99 of active dispersal and passive wind transport. High rates of dispersal also occur in the closely
100 related and ecologically similar species *D. simulans.* Here, the vertically transmitted bacterial
101 symbiont *Wolbachia* spread at a rate of more than 100 km per year across California, which
102 allowed Turelli & Hoffmann (1991) to estimate that migration rates in nature are far greater
103 than typically observed in behavioural experiments. One particular aspect of migration that is
104 not fully resolved is the extent to which populations in temperate and cold regions are
105 resident, overwintering in local refugia such as human households, or whether these

106 populations are partly reseeded annually by immigrants from warmer regions (Keller 2007),
107 generating biased migration from south to north in the Northern hemisphere.

108 Here, we study the population genetics of DMelSV and demonstrate a link between the
109 demographic histories of this obligate parasite and its host. On the one hand, this close
110 association may be shaped by coevolution between host and parasite. At the same time, the
111 rapidly evolving virus can give insights into its host's ecology that are hard to discern by
112 studying the fly's genetics.

113

114 Materials and Methods

115

116 Virus isolates

117 As DMelSV is vertically transmitted, it can be isolated by collecting infected flies and
118 establishing infected fly lines in the laboratory. After allowing flies isolated from the field to
119 reproduce in the lab, we diagnosed DMelSV infections through exposure to CO_2, a treatment
120 from which uninfected flies fully recover but which leaves infected flies paralysed (Brun &
121 Plus 1980; Wilfert & Jiggins 2010). Details of virus isolate collections can be found in table
122 A1 (supplementary material). Wild isolates of DMelSV were collected from 15 populations
123 of *D. melanogaster* across Europe (Germany – Southwest and Southeast, Greece – Athens,
124 Spain – Galicia, Sweden – Uppsala, UK – Cambridgeshire, Derby, Essex and Kent) the
125 United States (USA – California, Florida, Georgia and North Carolina) and Africa (Ghana –
126 Accra and Kenya – Thika (n = 2))between 2005 and 2010 (table A1) (Carpenter *et al*. 2007;
127 Wilfert & Jiggins 2010), including 8 samples for which we retrieved sequences from

128 Genbank (3 samples from Athens, Greece: AM689323, AM689330, AM689331; 4 from
129 Essex, UK: AM689310, AM689314, AM689317, AM689318; and 1 from Georgia, USA:
130 AM689320). The populations in Essex, UK and Athens, Greece, were repeatedly sampled
131 across these years. We also included Hap23, a well-studied lab isolate of DMelSV. Infected
132 fly lines collected up to the end of 2007 were maintained in the laboratory for 0.5 to 2.5 years
133 in isofemale lines (with the exception of the 8 lines for which we retrieved sequences from
134 Genbank, see above). During this time, we substituted the 2nd and 3rd wild-type chromosomes
135 with the isogenic line w^{1118} (described by Parks *et al*. (2004)) using the double balancer line
136 *SM2/Pm; TM3/Sb;* spa^{Pol} (Wilfert & Jiggins 2010) to allow their use in other experiments. We
137 also obtained eight samples from California (collected by Darren Obbard in 2008), two
138 samples from Kenya (collected by John Pool in 2009), one sample from Ghana (collected by
139 Claire Webster in 2010) and 56 samples from Cambridgeshire and Essex, UK (collected by
140 Heather Cagney in 2009), which were maintained in the lab for 1 to 3 months. These samples
141 were sequenced from the wild-type host.

142

143 Sequencing viral isolates

144 For each virus-infected fly line, we extracted RNA from 10 flies using 300 μl Trizol
145 (Invitrogen, Carlsbad, Ca.) following the manufacturer's instruction. The RNA was
146 resuspended in 50 μl RNA storage solution (Ambion, Austin, Tx.) and reverse transcribed
147 into cDNA using random hexamers and M-MLV reverse transcriptase (Invitrogen, Carlsbad,
148 Ca.). The viral isolates were sequenced using direct sequencing of PCR products using Sanger
149 sequencing as described by Carpenter *et al*. (2007) (see Table A2 for primer sequences). The
150 sequences were assembled into contigs and aligned using Sequencher 4.2 (Gene Codes

151 Corporations, Ann Arbor, Mi.). In total, we sequenced 5,732 bp of 111 viral isolates spanning
152 the *n, p, x, m* and *g* genes in the '3 prime half of the DMelSV genome and found no evidence
153 for heterozygotes in this data. All sequences have been submitted to Genbank under the
154 numbers GQ451693-GQ451695 and HQ655003-HQ655110. Additionally, we sequenced
155 1015 bp of the *p*- and *g*-genes of an additional 57 isolates from the UK and Ghana (559 bp
156 and 456 bp of each gene respectively; Genbank accession numbers JN167249-JN167305).

157

158 Estimating the viral mutation rate

159 To refine the estimated mutation rate reported in Carpenter *et al*. (2007), we directly
160 measured the mutation rate of DMelSV by re-sequencing virus isolates maintained in live fly
161 stocks. Virus isolates were initially collected and sequenced by Carpenter *et al*. (2007) in
162 2005. The viral isolates were then maintained in live fly stocks and transferred to new vials
163 every 3 – 4 weeks for 2.4 years (November 2005 – April 2008), at which point a new sample
164 of viral RNA was isolated for sequencing. We compared a total of 26,857 bp over 15 lines
165 (Supplementary material, table A3), covering partial sequences of the *p*- and *g*-genes for 13
166 lines (1208 – 1227 bp) and the 3' half of the genome (5626 bp and 5729 bp respectively) for
167 two isolates, through an automated search for mismatches in pair-wise alignments of the
168 corresponding sequence pairs. This resulted in a total of 6 substitutions. Additionally, we used
169 the data reported in Carpenter *et al*. (2007): 3 mutations in 5744 bp for line A3 (excluding 2
170 deletions in non-coding regions) and none in 5144 bp for line A3-E55 (excluding a 600 bp
171 region of the genome that was modified by ADAR-induced hypermutation (Carpenter *et al*.
172 2009)) over a period of 10 – 20 years. The substitution rate was calculated by a maximum
173 likelihood approach, assuming a Poisson process. The confidence intervals are based on a 2-

174 unit reduction in the log-likelihood, taking account of the uncertainty in the divergence time
175 of lines A3 and A3-E55 by basing the lower and upper value of the confidence intervals on
176 the values for the longer and shorter divergence times respectively. As there was no
177 significant difference between mutations at the 1st and 2nd codon position as opposed to the 3rd
178 position (Fisher's Exact test, $p = 0.453$), we interpret this rate as the mutation rate, rather than
179 the evolutionary rate, assuming that divergence times and selection strength for these lab-
180 maintained lines may not have been sufficient to purge slightly deleterious mutations (Rocha
181 *et al.* 2006). We have implemented this by using the mutation rate as a prior for the 3rd-codon
182 substitution rate for phylogenetic analyses.

183

184 Phylodynamic reconstruction

185 All phylogenetic analyses were carried out on two data sets: 1) 'short': a dataset consisting of
186 1015 bp of the partial *p*- and *g*-genes from 176 viral isolates, which includes the 8 previously
187 isolated sequences retrieved from Genbank (see under 'viral isolates') (58 segregating sites at
188 codon positions 1&2, 120 at codon position 3); 2) 'long': a subset of 111 viral isolates for
189 which we obtained the 5' half of the DMelSV genome, covering 5735 bp of the *n*-, *p*-, *x*-, *m*-
190 and *g*-genes (332 segregating sites at codon position 1&2, 424 at codon position 3). Within
191 the long dataset, the sequences of all isolates were sampled within a 6-month period. In the
192 short dataset, the earliest samples in Genbank were sampled in November 2005, whereas the
193 last sample from Ghana was added in February 2010. We have not used tip date information
194 for phylodynamic reconstruction as tip age provides no information in this data set (an
195 expected 0.2 mutations between the two sequences with the largest time difference).

196 We used the coalescent sampling algorithm BEAST v.1.7.5 (Drummond *et al.* 2012) to
197 reconstruct the phylogeny and demographic history of DMelSV. We ran 3 types of analysis a)
198 a basic analysis of the short data set to determine the time to most recent ancestor (tmrca) for
199 all DMelSV isolates – this analysis revealed that there are a 'common' and a 'rare' clade of
200 DMelSV (see results) b) discrete phylogeographic analyses within the long data set of the
201 'common' clade to address population structure and migration (Lemey *et al.* 2009) and c)
202 models with exponential growth rate in the 'common' clade both within and across
203 populations using the long data set.

204 All models use the SRD06 model of sequence evolution (Shapiro *et al.* 2006) for consistence,
205 as the more complex General Time Reversible model was not supported in analyses with
206 smaller data subsets. We partitioned sites into two categories according to position (1&2 and
207 3) and separately estimated substitution and rate heterogeneity. Except for the models
208 exploring exponential growth, the phylogenetic tree was reconstructed using a Jeffreys prior
209 distribution of node heights and tree topology under a model of constant population size
210 (Drummond *et al.* 2002), as this was preferred over a GMRF skyride model (Minin *et al.*
211 2008) based on a Monte Carlo Marcov Chain equivalent of Aikaike's information criterion
212 (AICM) on the marginal likelihoods as implemented in Tracer v.1.6 (Baele *et al.* 2012)
213 (http://tree.bio.ed.ac.uk/software/tracer/; difference in AICM 411.93).

214 To test whether molecular clock rates are uniform across the tree, we first ran models with a
215 relaxed uncorrelated lognormal molecular clock (Drummond *et al.* 2006). Based on the shape
216 of the marginal posterior distribution of the clock parameter, a strict molecular clock could
217 not be excluded. In accordance with these tests, the models assume a strict molecular clock
218 for sequence evolution. We used the experimentally determined substitution rate as a fully

219 informative prior for the third codon position, accounting for uncertainty in this parameter by
220 approximating with a lognormal distribution (mean in real space = 3.95×10^{-5}, log(standard
221 deviation) = 0.5 substitutions/year/site). For all other priors (except location rate prior for
222 phylogeograpic models, see below), we used the default priors implemented in BEAUTI
223 1.7.5. We ran models with 2 runs each of 10 million MCMC generations, sampling the chain
224 every 1000 generations with a burn-in of 1 million generations, to obtain effective sample
225 sizes >200 for all parameters, except for the model exploring exponential growth within the
226 'common' clade, which we ran 3x as long to obtain effective sample sizes > 200. We
227 combined the two runs using Tracer v.1.5 and examined them for convergence. We used
228 TreeAnnotator v. 1.7.5 to produce Maximum Clade Credibility (MCC) trees.

229 To infer migration rates and geographic origins of branches, we fitted a discrete
230 phylogeographic model with an assymmetric substitution model as described in Lemey *et al.*
231 (2009) using Ferreira & Suchard (2008) conditional reference prior for the change in location
232 rate. For this analysis, we used the 'long' dataset for the 'common' clade to maximize
233 information content. We pooled samples per continent as well as into 6 geographic regions
234 (Germany, Greece, Spain, Sweden, UK and US) and removed the lab-derived strain Hap23
235 (310 segregating sites at codon position 1&2, 364 at codon position 3). Well-supported rates
236 were identified using a Bayes factors analysis based on the Bayesian stochastic search
237 variable selection (Lemey *et al.* 2009), using a ln Bayes Factor of 3 as a cut-off. To
238 investigate demographic history, we tested whether models with exponential growth rate were
239 supported. As population structure may violate the assumptions of demographic
240 reconstruction and may introduce biases (Ho & Shapiro 2011; Heller *et al.* 2013), we also
241 performed these analyses in population subsets.

242 We calculated indicators of population structure and differentiation using DnaSP v5 (Librado
243 & Rozas 2009): π, the mean pair-wise difference between sequences; K_{ST}, a measure of
244 population differentiation based on the proportion of between-population nucleotide
245 differences (Hudson *et al.* 1992); and the nearest-neighbor statistic S_{NN}, which calculates the
246 proportion at which genetic nearest neighbors are found within the same population (Hudson
247 2000).

248

249 Results

250

251 The DMelSV phylogeny

252 The phylogenetic tree of 176 DMelSV isolates isolated from geographically diverse *D.*
253 *melanogaster* populations reveals two clades, which we refer to as the 'common' and the
254 'rare' clades (Figure 1). The large majority of isolates cluster together in the common clade
255 (fig. 1), and the 'rare' clade represents only 1.7 % of the DMelSV isolates (AP30
256 NC_013135, Eir3 JN167273 and Ghana JN167305). During our collections in
257 Cambridgeshire and Essex in the South of England during 2009, we tested 1,610 wild-caught
258 *D. melanogaster* individuals for DMelSV through the CO_2-assay (Brun & Plus 1980; Wilfert
259 & Jiggins 2010) and subsequent sequencing, and found that the prevalence for the 'common'
260 clade was 4.9 % while that of the 'rare' clade was 0.06 %, with only one isolate found in the
261 entire sample.

262 To allow us to date events on the DMelSV tree, we first estimated the rate of sequence
263 evolution of the virus. To do this, we counted the number of sequence changes that had

264 occurred over several years in viruses that we maintained in *Drosophila* lines (see methods
265 for details). We estimated that there are 3.95 x 10^{-5} (C.I. 1.5 x 10^{-5} – 9.5 x 10^{-5}) substitutions
266 per site per year. This supports and greatly increases the precision of a previous estimate
267 using some of the same data by Carpenter *et al.* (2007) of 4.6 x 10^{-5} substitutions per site per
268 year (C.I. 1.8 x 10^{-4} – 9.5 x 10^{-6}), and is similar to the rates found in other rhabodviruses
269 (Furio *et al.* 2005; Sanjuan *et al.* 2010). We found no evidence of heterogeneity in the
270 evolutionary rates across branches on the phylogenetic tree (see methods for details), which
271 together with Carpenter *et al.*'s analyses (2007) suggests that mutation rates estimated in the
272 lab can be extended to field isolates of DMelSV.

273 Based on an analysis of sequences of all 176 isolates, the most recent common ancestor of all
274 the DMelSV isolates dates to ~ 3900 years ago ('short alignment'; posterior mean: 3873
275 years, 95 % HPD: 778 – 8198 years). The most recent common ancestor of the 'rare' clade
276 dates back ~ 3300 years (posterior mean: 3262 years, 95 % HPD 577 – 6940), while the most
277 recent common ancestor of the 'common' clade occurred only ~ 1100 years ago (posterior
278 mean: 1079 years, 95% HPD 177– 2330), with all the isolates except one in this clade sharing
279 an ancestor 517 years ago (95 % HPD 98 – 1114 years). In interpreting these results, it should
280 be noted that using contemporary evolutionary rates to infer the divergence of viral clades in
281 the distant past may lead to underestimates of the true age of these events (Sharp &
282 Simmonds 2011).

283

284 Structure and migration between continents

285 There is a striking difference in the distribution across our phylogeny of isolates from Africa
286 and the other continents (Figure 1). We only have three isolates from Africa, but even this

287 sample size shows that diversity in Africa is greater than elsewhere. The three African
288 sequences are each found on distant lineages, with two of them having no close relatives
289 among the isolates from other continents. This is reflected in a higher genetic diversity of the
290 African sequences compared to the pooled sample of non-African sequences (short alignment:
291 $\pi_{\text{Africa}} = 0.06$, $\pi_{\text{non-Africa}} = 0.01$).

292 The European and North American isolates are also genetically distinct. These isolates
293 are mostly clustered together in the 'common' clade, with single sequences from each
294 continent in the 'rare' clade (Figure 1). Within the common clade there is substantial
295 genetic structure between Europe and the US, with sequences that are genetic nearest
296 neighbours overwhelmingly being within the same continent (Hudson's nearest neighbour
297 statistic, short alignment: $S_{nn} = 0.95, p < 0.001$) (Hudson 2000). Genetic diversity is higher in
298 Europe than North America (long alignment, common clade: $\pi_{\text{Europe}} = 0.008$, $\pi_{\text{US}} = 0.003$).
299 This does not appear to be accounted for by the larger number of populations sampled in
300 Europe, as the genetic diversity within the US, comprising samples from San Diego
301 (California), Raleigh (North Carolina) and, for the short dataset, Athens (Georgia), is less than
302 that of any of the European populations (Table 1). All 'common' samples from the US are
303 found on a single branch of the phylogenetic tree, forming a monophyletic clade with one
304 sample from the East of England (figs. 1 & 3). These samples span a geographic distance of
305 2000 – 2200 miles, whereas the maximum distance between any of the European populations
306 is 1700 miles (Greece and Spain).

307 To test whether these patterns of migration between continents are supported statistically, we
308 analysed migration between Kenya, Europe and the US using the asymmetric
309 phylogeographic model in BEAST v. 1.7.5 (Drummond *et al*. 2012). Using only sequences in

310 the common clade, we found evidence for migration between Europe and Kenya (Bayes
311 Factors: $BF_{Europe\text{->}Kenya}$ = 31.4), and Europe and the US ($BF_{Europe\text{->}US}$ = 31.4, $BF_{US\text{->}Europe}$ =
312 1003.6). In the common clade, the most recent ancestor of European samples is 385 years (95
313 % HPD 92 – 786 years), 114 years for the US population (95 % HPD 26 -239 years) and 1082
314 years for the two samples from Kenya (95 % HPD 253 -2219 years).

315

316 Structure and migration between populations

317 There is also genetic structure between populations on the same continent (Figures 1 and 2).
318 To investigate migration at this finer scale, we analysed isolates in the common clade from
319 Germany, Greece, Spain, Sweden, UK and the US (we excluded Africa due to low sampling).
320 These populations are strongly genetically structured as indicated by a high value of Hudson's
321 nearest neighbour statistic ($S_{nn} = 0.835$, $p < 0.001$). Some populations, such as the UK and
322 US, are largely monophyletic, whilst others, such as Germany, are more scattered across the
323 tree. This is reflected in our estimates of genetic diversity, which are higher in continental
324 European populations than in the UK and US (Table 1).

325 We used a Bayesian phylogeographic approach to reconstruct routes of migration between
326 populations. As shown in Fig. 2, we found strong support for migration from Greece to
327 Germany, with the German population serving as a hub with migration supported to all other
328 populations, including the US. It should be noted that immigration from other unsampled
329 populations around the world could inflate diversity in Germany and generate this pattern, so
330 this conclusion should be treated with caution. Additionally, we found support for migration
331 from the US to the UK, due to a single recent back-migration event (Fig. 1 and 3). In
332 accordance with these discrete migration routes, we find little evidence of isolation by

333 distance: geographic distance explains only 8 % of the pairwise genetic distance between
334 samples (Mantel-Test, 10000 replicates, p <0.001).

335

336 Effective population size

337 Sequence variation can also allow us to estimate the effective population size of the virus (N_e)
338 and reconstruct how it has changed through time. Using the alignment of all 176 sequences,
339 we estimate that the global effective population size of DMelSV is 1.7 x 10^4 (95 % HPD 0.4
340 x 10^4 – 3.7 x 10^4). Population structure means that this estimate depends on our choice of
341 sampling scheme, so we also estimated N_e in the monophyletic branches in the largely closed
342 UK and US populations (UK: N_e = 783, 95 % HPD 97 – 1825; US: N_e = 3023, 95 % HPD 455
343 – 7310). For these estimates, we assume a constant N_e and 10 generations per year as for the
344 host (e.g. Richardson *et al*. 2012)), as this virus is exclusively vertically transmitted and
345 shows no quasi-species formation (pers. comm. M.L. Wayne).

346 To investigate how N_e may have changed through time, we attempted to minimise the
347 confounding effects of population structure (Ho & Shapiro 2011; Heller *et al*. 2013) by
348 analysing each population separately as well as all the non-African isolates combined. First,
349 we compared a model of exponential growth to a model of constant population size within the
350 common clade. While exponential growth is supported overall (exponential growth rate
351 0.024, 95 % HPD 0.005 – 0.05, doubling time = 29.4 years), at a population level, this pattern
352 is only supported in the Greek and US populations and the rate of growth is very uncertain
353 (common clade only; Greece: growth rate = 0.018, 95 % HPD 0.003 – 0.040, doubling time =
354 37.7 years; US: growth rate = 0.17, 95% HPD 0.023 – 0.404, doubling time = 4.1 years).

355 Population structure tends to give a false signature of a declining N_e (Heller *et al*. 2013), so
356 this is unlikely to confound this analysis.

357

358 Local and temporal patterns

359 We found evidence for changes in the genetic composition of the viral population in Greece
360 through time. The Greek isolates are found on two branches of the phylogenetic tree (Fig. 1 &
361 3). As shown in Fig. 1, the isolates from 2005 (n = 2) and 2006 (n = 9) are exclusively found
362 on one branch, with the 2007 isolates distributed over both branches (n = 24 and n = 31
363 respectively; Fisher's exact test comparing prevalence in 2005, 2006 and 2007: $p = 0.002$).
364 The change in genotype frequencies is reflected in significant genetic structure between the
365 samples from the different years ($K_{ST\mathrm{short}} = 0.13173$, $p < 0.001$).

366 Populations in the East of England (Cambridgeshire and Essex) were also sampled repeatedly,
367 but here there were no changes through time. Flies were collected from Essex in 2005, 2007
368 and 2009 (n = 10, n = 10 and n = 5, respectively), and in 2009 from Cambridgeshire (n = 50,
369 ~45km from Essex). Here all the years were represented on the same major branch and there
370 was little change through time ($K_{ST\mathrm{short}} = 0.04667$, $p < 0.05$). There only limited population
371 structure within the Cambridgeshire sample, comprising of samples from Cambridge town
372 and the surrounding villages Girton, Fulborn, St. Ives and Waterbeach, with only a minority
373 of genetic nearest neighbours coming from the same population ($Snn = 0.36$, $p < 0.05$).

374

375 Discussion

376 DMelSV is a vertically transmitted pathogen that can only be transmitted by the eggs or
377 sperm of *D. melanogaster.* Despite being passed between generations along with the host
378 genome, critical population genetic parameters affecting patterns of genetic variation within
379 the viral populations are very different from its host. We estimate that the effective population
380 size of DMelSV is roughly 10,000, which is about two orders of magnitude smaller than *D.*
381 *melanogaster* (Andolfatto & Przeworski 2000). This low N_e is expected given that the virus
382 only infects a few percent of flies, so all else being equal its effective population size will be
383 proportionately less than that of fly genes (Carpenter *et al.* 2007). The viral mutation rate is
384 much greater — we see about 3 x 10^{-6} mutations per site per fly generation in viruses
385 maintained in lab fly stocks (this may be an underestimate of the mutation rate if some
386 deleterious mutations are purged by selection). In contrast, the mutation rate of *D.*
387 *melanogaster* is three orders of magnitude lower (3.5×10^{-9} per site per generation for
388 nuclear genes (Keightley *et al.* 2009)). Together these factors mean that patterns of population
389 structure and migration can be seen at a temporal resolution in the DMelSV population that
390 are difficult to detect when studying *D. melanogaster* directly without taking recourse to
391 whole-genome population sequencing (Campo *et al.* 2013).

392 As expected given its low effective population size, we see far greater population structure in
393 DMelSV than one would expect to find in the *Drosophila* genome. This variation can
394 therefore potentially be used to provide high resolution insights into the past demography and
395 migration of the flies themselves. This could be valuable, as despite being one of biology's
396 best-studied organisms, many aspects of *D. melanogaster*'s field ecology remain obscure.
397 However, this must also be done with caution, as natural selection within the viral population,
398 fluctuations in prevalence and the spread of the virus though uninfected populations may have
399 overwritten the signature of host demography. In particular, there is evidence DMelSV

400 populations have recently evolved to overcome host resistance due to a polymorphism in the
401 gene *ref(2)P* (Fleuriet & Sperlich 1992; Fleuriet & Periquet 1993; Wayne *et al.* 1996;
402 Bangham *et al.* 2007; Wilfert & Jiggins 2013). Some of the patterns that we see may therefore
403 reflect the spread of these virulent viral genotypes through populations.

404 At a continental scale, the evolutionary history of the virus mirrors that of the host, with
405 patterns of genetic variation suggesting that the virus originated in Africa, migrated to
406 Europe, and then on to the US. Does this mean that we are seeing the traces of the out of
407 Africa migration of flies in the viral population? At first sight our time-calibrated phylogeny
408 would suggest not, as while the most recent common ancestor of all our isolates dates back
409 ~3900 years, 98% of the isolates in Europe shared a common ancestor just ~520 years ago. In
410 contrast, even the most recent estimates of the genetic bottleneck that flies experienced during
411 the initial migration of *D. melanogaster* out of Africa date back several thousand years, and
412 other estimates are substantially older (Lachaise *et al.* 1988; Ometto *et al.* 2005; Thornton &
413 Andolfatto 2006). However, it is thought to be common for long-term rates of sequence
414 evolution in viruses to be much slower than are observed over short time periods, and this
415 may frequently result in massive underestimates of the age of distant nodes on viral
416 phylogenies (Sharp & Simmonds 2011). With this bias in mind, along with the substantial
417 errors attached to all the estimates, we would tentatively suggest that the most parsimonious
418 explanation of our data is that we are seeing the traces of the global fly migration (although
419 selection in Europe may also be important; see below). In North America, we estimate that
420 the most recent common ancestor of the 'common' viral clade dates back ~114 years to the
421 turn of the last century. This suggests that DMelSV arrived in the United States at the
422 beginning of the invasion of this continent by *D. melanogaster* in the late 19th century (Keller

423 2007). A recent invasion is also supported by a local pattern of population growth and the
424 lack of genetic structure across the US (albeit with limited sampling).

425 Within Europe, DMelSV populations are highly genetically structured, with most sequences
426 in some populations forming almost monophyletic clades. This would suggest that there are
427 low levels of migration across Europe, despite there being limited genetic structure at this
428 geographical scale in the *Drosophila* genome, which has low but significant F_{ST} values across
429 European populations for *D. melanogaster* (Caracristi & Schlotterer 2003; Verspoor &
430 Haddrill 2011). There is, however, still some ongoing migration, with recent immigrants
431 clearly visible on the phylogeny, and this allows us to reconstruct the movement of the flies
432 carrying DMelSV between populations. Across the phylogeny, in a given year a DMelSV
433 lineage has approximately a 1 in 700 chance of switching to a different population (μ=1.39 x
434 10^{-3} per year; 95% HPD 2.47 x 10^{-4} – 3.12 x 10^{-3}; see Lemey *et al.* (2009) for details). The
435 fruit trade may mediate this migration. For example, drosophilids are frequently found in
436 shipments of imported fruit in the UK (Reid & Malumphy 2009), a country that in 2011
437 imported 33 % of its fresh fruit from the EU and 55 % from outside the EU (DEFRA 2012),
438 including close neighbors of the southern US (Costa Rica: 8%, Colombia: 7 %, Dominican
439 Republic: 6 %, Ecuador: 4%) (DEFRA 2012).

440 Overwintering and the persistence of populations in cold climates is a particular aspect on
441 which studying DMelSV population genetics can begin to shed some light. *D. melanogaster*
442 is not tolerant of extreme cold (Izquierdo 1991) and can only overwinter as adults, with
443 females entering reproductive diapause (Saunders *et al.* 1989). Bouletreau-Merle *et al.* (2003)
444 have shown that in the Rhone Valley in Southern France, *D. melanogaster* is able to
445 overwinter in refugia inside human dwellings, being tolerant to underfeeding at cool

446 temperatures, thus forming a permanent population in this area. Its sister species *D. simulans*,
447 which does not use such shelters, however is re-invading this area from southern refugia after
448 cold winters (Bouletreau-Merle *et al.* 2003). *D. simulans* does not occur in cold climates north
449 of the Alps in central and northern Europe (Babaaissa *et al.* 1988). The existence of stable
450 latitudinal clines in allele frequencies in *D. melanogaster* indicates that there is some genetic
451 continuity between generations in populations in cold climates. However, the relative
452 importance of recolonisation from the warmer south versus local persistence in these
453 populations is unknown. If a significant proportion of the population is recolonized from the
454 south every generation, then migration should predominantly occur from the south to north.

455 Our analysis of DMelSV migration patterns suggests that populations persist locally in cold
456 winters and recolonisation from southern Europe is not important. The isolates from the two
457 populations with the harshest winters, Sweden and Germany, have the most phylogenetically
458 diverse DMelSV strains in Europe (Figure 3). Our analysis of migration patterns suggests that
459 central German populations, which experience extreme cold in the winter, show no evidence
460 of immigration from the south. Therefore, rather than being recolonized every generation, this
461 population appears to have persisted over the long term and has probably exported migrants
462 elsewhere. This is despite evidence that DMelSV infection may reduce overwintering success
463 in *D. melanogaster* (Fleuriet 1981b), which further increases the chances of recolonisation
464 from the south. While our data indicates that northern populations are not recolonized from
465 the Iberian or Balkan peninsulas every spring, an alternative explanation of our data is that
466 northern European populations could be recolonised from un-sampled populations in Africa,
467 Asia or South-America through the global fruit-trade, or from Italy. While sampling of these
468 populations would be required to formally reject this hypothesis, it seems improbable to us

469 that immigration from these areas would be greater than from nearby populations within
470 Europe.

471 Within the South-East of England we carried out a detailed large-scale study of the local
472 population structure of DMelSV. This revealed very little population structure and variation
473 over years, and in particular showed no evidence of migration other than the long-term
474 seeding from Northern America and central Europe. This suggests that the populations in the
475 South-East of England are long-term resident populations. In Athens, Greece on the other
476 hand, we found a shift in genetic diversity between 2006 and 2007: the 11 samples from 2005
477 and 2006 were found on only one major branch of the phylogeny, whereas 31 of the 55
478 samples from 2007 fall onto another major branch. Assuming that this pattern is not a quirk
479 of our limited sampling in the earlier years, this change could be due to seasonal fluctuations,
480 migration from other populations, perhaps reflecting local population structure, or selection
481 favoring viruses from the new clade in 2007. Such rapid changes in viral types are possible in
482 this system (Wilfert & Jiggins 2013).

483 In this host-parasite system, natural selection due to coevolution with the host may be an
484 important factor shaping patterns of genetic variation (Wilfert & Jiggins 2013). DMelSV
485 infected flies suffer a reduction in egg viability (Fleuriet 1981a) and an increase in larval
486 development time (Seecof 1964). Additionally, infected flies may have reduced survival
487 under stressful conditions such as limited resources (Yampolsky *et al.* 1999) and
488 overwintering (Fleuriet 1981b). Several polymorphic genes in *D. melanogaster* populations
489 convey resistance against DMelSV (Gay 1978; Bangham *et al.* 2008; Magwire *et al.* 2011;
490 Magwire *et al.* 2012). As mentioned above, an allele of the *ref(2)P* gene that can block
491 maternal transmission of DMelSV, has undergone a partial selective sweep in *D.*

492 *melanogaster* (Wayne *et al.* 1996; Bangham *et al.* 2007). However, only certain avirulent
493 DMelSV genotypes are affected by *ref(2)P,* and in Languedoc, France (Fleuriet & Periquet
494 1993) and Tubingen, Germany (Fleuriet & Sperlich 1992) (also sampled in this study), the
495 avirulent virus type was mostly replaced by virulent viruses, which are largely unaffected by
496 *ref(2)P*, in the 1980s and early 1990s. We have previously characterized the effect of the
497 resistant allele of *ref(2)P* on five of the isolates in our phylogeny (Carpenter *et al.*
498 2007)(marked on Figure 1). One isolate in the 'rare' clade was sensitive to the *ref(2)P*
499 resistance allele, while four isolates scattered across the common clade were all virulent and
500 are not affected (Carpenter *et al.* 2007). As it is not known whether DMelSV evolved to
501 evade the effects of *ref(2)P* once or many times, it is not clear exactly how selection on this
502 trait may have affected the patterns we see. However, it is possible that the common clade is
503 largely virulent, and the migration patterns may be partly driven by selection on this trait.

504 In conclusion, we show that the vertically transmitted *D. melanogaster* pathogen DMelSV has
505 been associated with its host for several thousand years. It has a much smaller effective
506 population size and higher mutation rate than its host, resulting in high levels of genetic
507 structure and a detailed record of recent migration events. The virus has spread around the
508 world following the same migration route is its host, and there is a low rate of ongoing
509 migration within and between continents. The degree of genetic structure and direction of
510 migration suggest that *D. melanogaster* can persist in temperate and cold climates.

511

512

513

514 Acknowledgements

515 We would like to thank Florian Bayer, Mitchel Chewi and Heather Cagney for assistance in
516 the lab. Trevor Bedford, Sam Lycett and John Welch provided invaluable advice on statistics
517 and phylogenetic reconstruction. We are very grateful to the people who helped collecting
518 viral isolates in the field: Natasa Fytrou & family, Jennifer Carpenter, Sandra South, Rebecca
519 Schulte, Florian Bayer & family, Severin Weybora (Bayrische Landesjagdschule), Ben
520 Longdon, Claire Webster, Heather Cagney, Darren Obbard and John Pool. This work was
521 financially supported by a Leverhulme Trust grant and a Royal Society University Research
522 Fellowship to F.J.

523

524 References

525 Andolfatto P , Przeworski M (2000) A genome-wide departure from the standard neutral
526 model in natural populations of *Drosophila*. *Genetics*, **156**, 257-268.
527 Babaaissa F, Solignac M, Dennebouy N *et al.* (1988) Mitochondrial DNA variability in
528 *Drosophila simulans* - quasi absence of polymorphism within each of the 3 cytoplasmic races.
529 *Heredity*, **61**, 419-426.
530 Baele G, Lemey P, Bedford T *et al.* (2012) Improving the accuracy of demographic and
531 molecular clock model comparison while accommodating phylogenetic uncertainty.
532 *Molecular Biology and Evolution*, **29**, 2157-2167.
533 Bangham J, Kim KW, Webster CL *et al.* (2008) Genetic variation affecting host-parasite
534 interactions: Different genes affect different aspects of sigma virus replication and
535 transmission in *Drosophila melanogaster*. *Genetics*, **178**, 2191-2199.
536 Bangham J, Obbard DJ, Kim KW *et al.* (2007) The age and evolution of an antiviral
537 resistance mutation in *Drosophila melanogaster*. *Proceedings of the Royal Society B-*
538 *Biological Sciences*, **274**, 2027-2034.
539 Biek R, Drummond AJ , Poss M (2006) A virus reveals population structure and recent
540 demographic history of its carnivore host. *Science*, **311**, 538-541.
541 Bouletreau-Merle J, Fouillet P , Varaldi J (2003) Divergent strategies in low temperature
542 environment for the sibling species *Drosophila melanogaster* and *D-simulans*: overwintering
543 in extension border areas of France and comparison with African populations. *Evolutionary*
544 *Ecology*, **17**, 523-548.
545 Brun P , Plus N. 1980. The viruses of *Drosophila*. In: Ashburner M, Wright TRF, editors.
546 The genetics and biology of *Drosophila*. London: Academic Press. p. 625-702.

547 Campo D, Lehmann K, Fjeldsted C *et al.* (2013) Whole-genome sequencing of two North
548 American *Drosophila melanogaster* populations reveals genetic differentiation and positive
549 selection. *Molecular Ecology*, **22**, 5084-5097.
550 Caracristi G , Schlotterer C (2003) Genetic differentiation between American and European
551 *Drosophila melanogaster* populations could be attributed to admixture of African alleles.
552 *Molecular Biology and Evolution*, **20**, 792-799.
553 Carpenter JA, Hadfield JD, Bangham J *et al.* (2012) Specific interactions between host and
554 parasite genotypes do not act as a constraint on the evolution of antiviral resistance in
555 *Drosophila*. *Evolution*, **66**, 1114-1125.
556 Carpenter JA, Keegan L, Wilfert L *et al.* (2009) Evidence for ADAR-induced hypermutation
557 of the Drosophila sigma virus (Rhabdoviridae). *BMC Genetics*, **10**, 75.
558 Carpenter JA, Obbard DJ, Maside X *et al.* (2007) The recent spread of a vertically transmitted
559 virus through populations of *Drosophila melanogaster*. *Molecular Ecology*, **16**, 3947-3954.
560 Coyne JA , Milstead B (1987) Long-distance migration of *Drosophila.* 3. Dispersal of
561 *Drosophila melanogaster* alleles from a Maryland orchard. *American Naturalist*, **130**, 70-82.
562 David JR , Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations.
563 *Trends in Genetics*, **4**, 106-111.
564 DEFRA. 2012. Agriculture in the United Kingdom - 2012. In.
565 Dobzhansky T (1973) Active Dispersal and passive transport in *Drosophila*. *Evolution*, **27**,
566 565-575.
567 Drummond AJ, Ho SYW, Phillips MJ *et al.* (2006) Relaxed phylogenetics and dating with
568 confidence. *PLoS Biology*, **4**, 699-710.

569 Drummond AJ, Nicholls GK, Rodrigo AG *et al.* (2002) Estimating mutation parameters,
570 population history and genealogy simultaneously from temporally spaced sequence data.
571 *Genetics*, **161**, 1307-1320.
572 Drummond AJ, Suchard MA, Xie D *et al.* (2012) Bayesian phylogenetics with BEAUti and
573 the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969-1973.
574 Fabian DK, Kapun M, Nolte V *et al.* (2012) Genome-wide patterns of latitudinal
575 differentiation among populations of *Drosophila melanogaster* from North America.
576 *Molecular Ecology*, **21**, 4748-4769.
577 Falush D, Wirth T, Linz B *et al.* (2003) Traces of human migrations in *Helicobacter pylori*
578 populations. *Science*, **299**, 1582-1585.
579 Ferreira MAR , Suchard MA (2008) Bayesian analysis of elapsed times in continuous-time
580 Markov chains. *Canadian Journal of Statistics-Revue Canadienne De Statistique*, **36**, 355-
581 368.
582 Fleuriet A (1976) Presence of hereditary rhabdovirus sigma and polymorphism for a gene for
583 resistance to this virus in natural populations of *Drosophila melanogaster*. *Evolution*, **30**, 735-
584 739.
585 Fleuriet A (1981a) Comparison of various physiological traits in flies (*Drosophila*
586 *melanogaster*) of wild origin, infected or uninfected by the hereditary rhabdovirus sigma.
587 *Archives of Virology*, **69**, 261-272.
588 Fleuriet A (1981b) Effect of overwintering on the frequency of flies infected by the
589 rhabdovirus sigma in experimental populations of *Drosophila melanogaster*. *Archives of*
590 *Virology*, **69**, 253-260.

591 Fleuriet A , Periquet G (1993) Evolution of the *Drosophila melanogaster* sigma virus system
592 in natural populations from Languedoc (Southern France). *Archives of Virology*, **129**, 131-
593 143.
594 Fleuriet A , Sperlich D (1992) Evolution of the *Drosophila melanogaster* - sigma virus
595 system in a natural population from Tubingen. *Theoretical and Applied Genetics*, **85**, 186-
596 189.
597 Furio V, Moya A , Sanjuan R (2005) The cost of replication fidelity in an RNA virus.
598 *Proceedings of the National Academy of Sciences of the United States of America*, **102**,
599 10233-10237.
600 Gay P (1978) *Drosophila* genes which intervene in multiplication of sigma virus. *Molecular*
601 *& General Genetics*, **159**, 269-283.
602 Heller R, Chikhi L , Siegismund HR (2013) The confounding effect of population structure
603 on bayesian skyline plot inferences of demographic history. *PLoS ONE*, **8**, e62992.
604 Ho SYW , Shapiro B (2011) Skyline-plot methods for estimating demographic history from
605 nucleotide sequences. *Molecular Ecology Resources*, **11**, 423-434.
606 Hudson RR (2000) A new statistic for detecting genetic differentiation. *Genetics*, **155**, 2011-
607 2014.
608 Hudson RR, Boos DD , Kaplan NL (1992) A statistical test for detecting geographic
609 subdivision. *Molecular Biology and Evolution*, **9**, 138-151.
610 Izquierdo JI (1991) How does *Drosophila melanogaster* overwinter. *Entomologia*
611 *Experimentalis Et Applicata*, **59**, 51-58.
612 Keightley PD, Trivedi U, Thomson M *et al.* (2009) Analysis of the genome sequences of
613 three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Research*,
614 **19**, 1195-1201.

615 Keller A (2007) *Drosophila melanogaster*'s history as a human commensal. *Current Biology*,
616 **17**, R77-R81.
617 Kittler R, Kayser M , Stoneking M (2003) Molecular evolution of *Pediculus humanus* and
618 the origin of clothing. *Current Biology*, **13**, 1414-1417.
619 Lachaise D, Cariou M-L, David J *et al.* (1988) Historical biogeography of the *Drosophila*
620 *melanogaster* species subgroup. *Evolutionary Biology*, **22**, 159-225.
621 Lachaise D , Silvain JF (2004) How two Afrotropical endemics made two cosmopolitan
622 human commensals: the *Drosophila melanogaster-D.simulans* palaeogeographic riddle.
623 *Genetica*, **120**, 17-39.
624 Lemey P, Rambaut A, Drummond AJ *et al.* (2009) Bayesian phylogeography finds its roots.
625 *PLoS Computational Biology*, **5**, e1000520.
626 Librado P , Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA
627 polymorphism data. *Bioinformatics*, **25**, 1451-1452.
628 Magwire MM, Bayer F, Webster CL *et al.* (2011) Successive increases in the resistance of
629 *Drosophila* to viral infection through a transposon insertion followed by a duplication. *PLoS*
630 *Genetics*, **7**, e1002337.
631 Magwire MM, Fabian DK, Schweyen H *et al.* (2012) Genome-wide association studies reveal
632 a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila*
633 *melanogaster*. *PLoS Genetics*, **8**, e1003057.
634 Minin VN, Bloomquist EW , Suchard MA (2008) Smooth skyride through a rough skyline:
635 Bayesian coalescent-based inference of population dynamics. *Molecular Biology and*
636 *Evolution*, **25**, 1459-1471.
637 Nieberding CM , Olivieri I (2007) Parasites: proxies for host genealogy and ecology? *Trends*
638 *in Ecology & Evolution*, **22**, 156-165.

639 Ometto L, Glinka S, De Lorenzo D *et al.* (2005) Inferring the effects of demography and
640 selection on Drosophila melanogaster populations from a chromosome-wide scan of DNA
641 variation. *Molecular Biology and Evolution*, **22**, 2119-2130.
642 Parks AL, Cook KR, Belvin M *et al.* (2004) Systematic generation of high-resolution deletion
643 coverage of the *Drosophila melanogaster* genome. *Nature Genetics*, **36**, 288-292.
644 Rako L, Poulsen NA, Shirriffs J *et al.* (2009) Clinal variation in post-winter male fertility
645 retention; an adaptive overwintering strategy in *Drosophila melanogaster*. *Journal of*
646 *Evolutionary Biology*, **22**, 2438-2444.
647 Reid S , Malumphy C (2009) Fruit flies (Diptera: Tephritidae) intercepted on plant produce
648 imported into England and Wales. *Entomologist's Monthly Magazine*, **145**, 213-226.
649 Richardson MF, Weinert LA, Welch JJ *et al.* (2012) Population genomics of the Wolbachia
650 endosymbiont in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1003129.
651 Rocha EPC, Smith JM, Hurst LD *et al.* (2006) Comparisons of dN/dS are time dependent for
652 closely related bacterial genomes. *Journal of Theoretical Biology*, **239**, 226-235.
653 Sanjuan R, Nebot MR, Chirico N *et al.* (2010) Viral mutation rates. *Journal of Virology*, **84**,
654 9733-9748.
655 Saunders DS, Henrich VC , Gilbert LI (1989) Induction of diapause in *Drosophila*
656 *melanogaster* - photoperiodic regulation and the impact of arrhythmic clock mutations on
657 time measurement. *Proceedings of the National Academy of Sciences of the United States of*
658 *America*, **86**, 3748-3752.
659 Seecof RL (1964) Deleterious effects on *Drosophila* development associated with the sigma
660 virus infection. *Virology*, **22**, 142-148.

661 Shapiro B, Rambaut A , Drummond AJ (2006) Choosing appropriate substitution models for
662 the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*, **23**,
663 7-9.
664 Sharp PM , Simmonds P (2011) Evaluating the evidence for virus/host co-evolution. *Current*
665 *Opinion inVirology*, **1**, 436-441.
666 Stephan W , Li H (2007) The recent demographic and adaptive history of *Drosophila*
667 *melanogaster*. *Heredity*, **98**, 65-68.
668 Thornton K , Andolfatto P (2006) Approximate Bayesian inference reveals evidence for a
669 recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics*,
670 **172**, 1607-1619.
671 Turelli M , Hoffmann AA (1991) Rapid spread of an inherited incompatibility factor in
672 California *Drosophila*. *Nature*, **353**, 440-442.
673 Verspoor RL , Haddrill PR (2011) Genetic dversity, population structure and Wolbachia
674 infection status in a worldwide sample of *Drosophila melanogaster* and *D. simulans*
675 populations. *PLoS ONE*, **6**, e26318.
676 Wayne ML, Contamine D , Kreitman M (1996) Molecular population genetics of *ref(2)P*, a
677 locus which confers viral resistance in *Drosophila*. *Molecular Biology and Evolution*, **13**,
678 191-199.
679 Wilfert L , Jiggins FM (2010) Host-parasite coevolution: genetic variation in a virus
680 population and the interaction with a host gene. *Journal of Evolutionary Biology*, **23**, 1447-
681 1455.
682 Wilfert L , Jiggins FM (2013) The dynamics of reciprocal selective sweeps of host resistance
683 and a parasite counter-adaptation in *Drosophila*. *Evolution*, **67**, 761-773.

684 Wirth T, Meyer A , Achtman M (2005) Deciphering host migrations and origins by means of
685 their microbes. *Molecular Ecology*, **14**, 3289-3306.
686 Yampolsky LY, Webb CT, Shabalina SA *et al.* (1999) Rapid accumulation of a vertically
687 transmitted parasite triggered by relaxation of natural selection among hosts. *Evolutionary*
688 *Ecology Research*, **1**, 581-589.
689

690 Data Accesibility

691 DNA sequences with all accompanying information: Genbank accession numbers GQ451693-
692 GQ451695, HQ655003-HQ655110 and JN167249-JN167305. Alignments for the 'short' and
693 'long' datasets, the BEAST xml files for fig. 1, 2&3 as well as the tree files for fig.1 and 3 are
694 archived on dryad (doi:10.5061/dryad.7n696).

695

696 Author contributions

697 LW collected samples, performed analyses and wrote the paper. FJ collected samples and
698 wrote the paper.

699

700

701

702

703

704

705

706

707

708

709

710 Figure Legends

711 Figure 1: Maximum clade credibility (MCC) tree of 176 DMelSV isolates reconstructed from
712 the sequence of 1015bp of the *p*- and *g*-genes (the ‘short’ dataset). The scale shows estimated
713 divergence dates based on laboratory estimates of the rate of viral evolution. The coloured bar
714 indicates the population of origin of the branches. The grey bars indicate the years in which
715 samples from Greece were collected and arrows indicate 5 isolates of DMelSV that are
716 virulent (black arrow), i.e. can overcome the host’s *ref(2)P* host resistance mutation, or
717 avirulent (red arrow) (Carpenter *et al.* 2012). Posterior support >0.6 is indicated for nodes of
718 at least 3rd order.

719

720 Figure 2: Migration patterns within the ‘common’ clade in Europe and North America. The
721 weight of the line indicates the Bayes Factor support (from thin to thick arrows: BF = 3 – 10,
722 10 – 30, >30)

723

724 Figure 3. MCC phylogeny of the ‘common’ DMelSV populations in Europe and North
725 America, showing the reconstructed ancestral population states. The tree was reconstructed
726 from 5732bp of sequence from a subset of the isolates shown in Figure 1. The scale shows
727 estimated divergence dates based on laboratory estimates of the rate of viral evolution. The

728 branches are colored according to reconstructions of the population in which the lineages
729 were found. Posterior support >0.6 is indicated for nodes of at least 3rd order.

730

731 Tables

732 Table 1: Genetic diversity in DMelSV populations. Estimates are based on either 1015bp of
733 sequence from 176 isolates (π_{short}), or 5732bp of sequence from a subset of 111 isolates (π_{long}).

	π_{long}	π_{short}
Germany	0.00767	0.00789
Greece	0.00623	0.00625
Spain	0.0031	0.00804
Sweden	0.00927	0.00888
UK	0.00463	0.00261
US	0.00266	0.00197
Total/Europe	0.00822	0.00795

734

735

736