Between virus correlations in the outcome of infection across host species: evidence of virus genotype by host species interactions

Ryan M. Imrie*, Katherine E. Roberts, Ben Longdon

Centre for Ecology & Conservation, Biosciences, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Penryn, Cornwall

*corresponding author: ri237@exeter.ac.uk

Abstract

Virus host shifts are a major source of outbreaks and emerging infectious diseases, and predicting the outcome of novel host and virus interactions remains a key challenge for virus research. The evolutionary relationships between host species can explain variation in transmission rates, virulence, and virus community composition between hosts, but the potential for different viruses to interact with host species effects has yet to be established. Here, we measure correlations in viral load of four *Cripavirus* isolates across experimental infections of 45 *Drosophilidae* host species. We find positive correlations between every pair of viruses tested, suggesting that broadly susceptible host clades could act as reservoirs and donors for certain types of viruses. Additionally, we find evidence of genotype-by-genotype interactions between viruses and host species, highlighting the importance of both host and virus traits in determining the outcome of virus host shifts. More closely related viruses tended to be more strongly correlated, providing tentative evidence that virus evolutionary relatedness may be a useful proxy for determining the likelihood of novel virus emergence, which warrants further research.

Impact Summary

Many new infectious diseases are caused by viruses jumping into novel host species. Estimating the probability that jumps will occur, what the characteristics of new viruses will be, and how they are likely to evolve after jumping to new host species are major challenges. To solve these challenges, we require a detailed understanding of the interactions between different viruses and hosts, or metrics that can capture some of the variation in these interactions. Previous studies have shown that the evolutionary relationships between host species can be used to predict traits of infections in different hosts, including transmission rates and the damage caused by infection. However, the potential for different viruses to influence the patterns of these host species effects has yet to be determined. Here, we use four viruses of insects in experimental infections across 45 different fruit fly host species to begin to answer this question. We find similarities in the patterns of replication and persistence between all four viruses, suggesting susceptible groups of related hosts could act as reservoirs and donors for certain types of virus. However, we also find evidence that different virus genotypes interact in different ways with some host species. Viruses that were more closely related tended to behave in similar ways, and so we suggest that virus evolutionary relatedness may prove to be a useful metric for predicting the traits of novel infections and should be explored further in future studies.

49 ## Introduction

50 Virus host shifts, where viruses jump to and establish onward transmission in novel host species, are
51 a major source of outbreaks and emerging infectious diseases [1–3]. Many human infections,
52 including HIV, Ebola, and recently SARS-CoV-2, have shifted into humans from other species, and
53 continue to cause significant damage to public health, society, and the global economy [4–7].
54 Predicting and preventing virus host shifts have consequently become major goals of virus research
55 [8]. Many challenges remain in achieving these goals, including improving our understanding of the
56 host, virus, and ecological factors that influence the outcome of initial cross-species transmission
57 [9,10], and the evolutionary and epidemiological factors that determine which pathogens become
58 established in novel hosts [11].
59
60 Several studies have investigated how host evolutionary relatedness can explain variation in the
61 outcome of infection across host species. Greater phylogenetic distance between the natural
62 (donor) and recipient hosts is associated with decreased likelihood of cross-species transmission
63 [12,13] and reduced onward transmission within the novel host species [14]. Phylogenetic distance
64 between hosts also explains variation in virulence after cross-species transmission, which increases
65 when viruses jump between more distantly related hosts [14–16]. Groups of closely related hosts
66 have also been shown to share similar levels of susceptibility to novel viruses, independent of the
67 distance to the natural host [17,18], and harbour similar virus communities [19–21].
68
69 Less is known about the interactions that exist between virus genotype and host species in
70 determining the outcome of infection [18]. Within species, genotype-by-genotype interactions
71 between host and virus can be important determinants of the outcome of infection. These
72 interactions alter the rank-order of host susceptibility and so reduce the strength of correlations
73 between viruses [22]. Comparative analyses of fungal pathogens of plants [23] and ectoparasites of
74 mammals [24] have revealed parasite phylogenetic effects, and some evidence exists to suggest

similar effects may be found in viruses. Closely related viruses tend to infect the same broad host taxa [25], despite high levels of geographic range overlap between potential hosts [26], suggesting they share similar constraints on their host ranges. Both co-speciation and the preferential host switching of viruses can support this assumption, given that viruses are overwhelmingly likely to encounter other host taxa over the timescales required for speciation. That said, shifts between divergent host species are also common across every virus family [27] and these exceptions include several human zoonoses of major concern [14].

The extent to which closely related viruses share traits has yet to be established. However, inferring the characteristics of viruses from better studied relatives is common and sometimes necessary. This is frequently the case during the early stages of outbreaks, where primary research on new viruses or variants is not available. When SARS-CoV-2 first emerged, its characteristics and epidemiological trajectory were inferred from closely related zoonotic and endemic coronaviruses [28], and from other pandemic respiratory viruses such as influenza A [29]. Comparisons to previous outbreaks were used to parameterise disease models in the 2009 H1N1 pandemic [30,31], the 2014 Ebolavirus outbreak [32], and in forecast models of seasonal influenza [33]. Even for viruses that are not newly emerged, many experimental models of infection rely on surrogates when the virus of interest is unavailable, non-permissive in cell culture or animal models, or requires considerable adaptation to experimental hosts [34,35].

These comparisons assume that the traits of one virus are similar to other, related viruses. However, comparisons between more distantly related viruses, such as bat and canine rabies viruses [44] and diverged lineages of influenza viruses [45–47], found stark differences across larger evolutionary scales. Many examples also exist of small genetic changes having large phenotypic effects in viruses, including single SNP changes altering the host range of canine parvoviruses [36], the vector specificity of Chikungunya virus [37], and the infectivity of naturally occurring Ebolaviruses [38]. Only

101 three amino acid substitutions are required to switch receptor specificity of avian H7N9 influenza
102 from poultry to human cell receptors [39]. Virus evolution is often characterised by high mutation
103 rates and frequent reassortment and recombination [40–42]. This, alongside an incomplete sampling
104 of extant viruses [43], has left many poorly resolved evolutionary relationships between and within
105 existing virus lineages [44]. Given these complications, it remains an open question whether
106 comparisons between related viruses can produce accurate inferences of infection traits.
107
108 In this study, we have investigated how patterns of host susceptibility (the ability of a virus to persist
109 and replicate) are correlated between viruses, using experimental infections with four *Cripavirus*
110 isolates (family *Dicistroviridae*) across a panel of 45 host species of *Drosophilidae*. Three of the
111 viruses are isolates of Drosophila C virus (DCV-C, DCV-EB and DCV-M), a well-studied virus isolated
112 from *Drosophila melanogaster* [45]. The fourth virus is the closely related Cricket Paralysis virus
113 (CrPV) which was isolated from Australian field crickets (*Teleogryllus commodus*) and is a widely
114 used model insect pathogen [46–48]. Both DCV and CrPV infect a broad range of insect taxa [49],
115 cause virulent infections in adult flies [17,50], and share similar mechanisms for co-opting the host
116 translation machinery [51]. A major-effect resistance gene called *pastrel* increases resistance to DCV
117 in *D. melanogaster* [52–54] and has also been shown to provide cross-resistance to CrPV along with
118 another gene, *Ubc-E2H* [55]. Both DCV and CrPV are targeted by the host antiviral RNAi pathway and
119 each encodes a potent suppressor of antiviral RNAi. However, these suppressors have different
120 functions and target different components of the RNAi pathway [56,57]. DCV and CrPV also differ in
121 their tissue pathology; DCV has been shown to infect gut tissues, causing intestinal obstruction
122 following septic inoculation in *D. melanogaster*, which was not observed in CrPV infection [58].
123 Although little is known about the differences between DCV isolates, they have been shown to cause
124 similar levels of virulence in *D. melanogaster* and interact differently with the endosymbiont
125 *Wolbachia* [59].
126

127 Previous work in this host system has shown that susceptibility to DCV-C varies across host species,
128 with the host phylogeny explaining a large proportion of the variance in both viral load and virulence
129 [17]. The host phylogeny is also an important determinant of the evolution of DCV-C in novel hosts,
130 with evidence that mutations that adapt the virus to one host may also adapt it to closely related
131 host species. This suggests virus genotype could alter the likelihood of host shifts in *Drosophila* [60].
132 Here, we measure correlations in the ability of four viruses to replicate and persist across host
133 species and provide evidence of both broad similarities in infection outcome and differences
134 consistent with virus by host species interactions.
135

136 Materials & Methods

137 Fly Stocks

138 Flies were taken from laboratory stocks of 45 different species of *Drosophilidae* (S1 Text Table A).
139 Before experiments began all included stocks were confirmed to be negative for infection with DCV
140 and CrPV by quantitative reverse transcription PCR (qRT-PCR, described below). Stocks were
141 maintained in multi-generation *Drosophila* stock bottles (Fisherbrand) at 22˚C, in a 12-hour light-
142 dark cycle. Each bottle contained 50ml of one of four varieties of food media (S1 Text
143 Supplementary Methods).
144

145 Host Phylogeny

146 The method used to infer the host phylogeny has been described in detail elsewhere [17]. Briefly,
147 publicly available sequences of the *28S*, *Adh*, *Amyrel*, *COI*, *COII*, *RpL32*, and *SOD* genes were collected
148 from Genbank (see https://doi.org/10.6084/m9.figshare.13079366.v1 for a full breakdown of genes
149 and accessions by species). Gene sequences were aligned in Geneious v9.1.8
150 (https://www.geneious.com) using a progressive pairwise global alignment algorithm with free end
151 gaps and a 70% similarity IUB cost matrix. Gap open penalties, gap extension penalties, and
152 refinement iterations were kept as default.

153

154 Phylogenetic reconstruction was performed using BEAST v1.10.4 [61] as the subsequent
155 phylogenetic mixed model (see below) requires a tree with the same root-tip distances for all taxa.
156 Genes were partitioned into separate ribosomal (*28S*), mitochondrial (*COI*, *COII*), and nuclear (*Adh*,
157 *Amyrel*, *RpL32*, *SOD*) groups. The mitochondrial and nuclear groups were further partitioned into
158 groups for codon position 1+2 and codon position 3, with unlinked substitution rates and base
159 frequencies across codon positions. Each group was fitted to separate relaxed uncorrelated
160 lognormal molecular clock models using random starting trees and 4-category gamma-distributed
161 HKY substitution models. The BEAST analysis was run twice, with 1 billion MCMC generations
162 sampled every 100,000 iterations, using a birth-death process tree-shape prior. Model trace files
163 were evaluated for chain convergence, sampling, and autocorrelation using Tracer v1.7.1 [62]. A
164 maximum clade credibility tree was inferred from the posterior sample with a 10% burn-in. The
165 reconstructed tree was visualised using ggtree v2.0.4 [63].

166

167 **Virus Isolates**

168 Virus stocks were kindly provided by Julien Martinez (DCV isolates) [59], and Valérie Dorey and
169 Maria Carla Saleh (CrPV) [57]. DCV-C, DCV-EB and DCV-M were originally isolated from fly stocks
170 with origins in three separate continents; DCV-C and DCV-EB were isolated from lab stocks
171 established by wild capture in Charolles, France and Ellis Beach, Australia respectively, while DCV-M
172 was isolated directly from wild flies in Marrakesh, Morocco [45]. The CrPV isolate was collected from
173 *Teleogryllus commodus* in Victoria, Australia [64]. Virus stocks were diluted in Ringers solution [65]
174 to equalise the relative amounts of viral RNA and checked for contamination with CrPV (DCV
175 isolates) and DCV (CrPV isolate) by qRT-PCR as described below.

176

177

178 **Virus phylogeny**

179 Full genome sequences for DCV-C (*MK645242*), DCV-EB (*MK645239*), DCV-M (*MK645243*), and CrPV
180 (*NC_003924*) were retrieved from the NCBI Nucleotide database. Annotations of ORFs for the
181 replicase polyprotein (CrPV: *Q9IJX4*, DCV: *O36966*) and structural polyprotein (CrPV: *P13418*, DCV:
182 *O36967*) were collected from the UniProtKB database and used to separate the coding and non-
183 coding regions of each virus. ORF sequences were concatenated and aligned using the Geneious
184 progressive pairwise translation alignment algorithm with a Blosum50 cost matrix and default
185 parameters. Alignments were manually checked for quality and sequences aligning to CrPV ORF1
186 nucleotides 1-387 and 2704-2728 were removed due to the presence of large indels.
187 Phylogenetic reconstruction was performed using BEAST v1.10.4 with translated ORF sequences
188 fitted to an uncorrelated relaxed lognormal molecular clock model using a speciation birth-death
189 process tree-shape prior. A Blosum62 substitution model [66] with a gamma distribution of rate
190 variation with four categories and a proportion of invariable sites was used. The model was run for
191 10 million MCMC generations sampled every 1,000 iterations and evaluated in Tracer v1.7.1 as
192 above, and a maximum clade credibility tree inferred with a 10% burn-in.
193

194 **Inoculation**

195 Before inoculation, 0-1 day old male flies were kept in vials containing cornmeal media (S1 Text
196 Supplementary Methods), and were transferred to fresh media every 2 days for one week. Vials
197 contained between 5 and 20 flies (mean = 14.5) and were kept at 22˚C at 70% relative humidity in a
198 12-hour light-dark cycle. Flies were inoculated at 7-8 days old under CO_2 anaesthesia via septic pin
199 prick with 12.5µm diameter stainless steel needles (Fine Science Tools, CA, USA). These needles
200 were bent approximately 250µm from the end to provide a depth stop and dipped in virus solution
201 before being pricked into the pleural suture of each fly. Inoculation by this method has been shown
202 to follow the same course as oral infection but is less stochastic [67]. Inoculated flies were then snap
203 frozen immediately in liquid nitrogen, providing a 0 days-post-infection (dpi) timepoint, or

204 maintained in cornmeal vials for a further 2 days ± 3 hours before freezing, providing a 2 dpi time
205 point. Within replicate blocks the 0 and 2 dpi vials for each virus were inoculated on the same day,
206 and together constituted one biological replicate. We aimed to collect three biological replicates for
207 each species and virus combination, with the order of species, vial (0 or 2 dpi), and virus randomised
208 for each replicate block.
209

210 **Measuring Change in Viral Load**

211 To measure the change in viral load between 0 and 2 dpi, total RNA was extracted from flies
212 homogenised in Trizol (Invitrogen, supplied by ThermoFisher) using chloroform-isoproponyl
213 extraction, and reverse transcribed using Promega GoScript reverse transcriptase (Sigma) with
214 random hexamer primers. qRT-PCR was carried out on 1:10 diluted cDNA on an Applied Biosystems
215 StepOnePlus system using Sensifast Hi-Rox Sybr kit (Bioline). Cycle conditions were as follows: initial
216 denaturation at 95˚C for 120 seconds, then 40 cycles of 95˚C for 5 seconds, and 60˚C for 30 seconds.
217

218 DCV isolates were measured using the same primer pair (Forward: 5'-GACACTGCCTTTGATTAG-3',
219 Reverse: 5'-CCCTCTGGGAACTAAATG-3') which targeted a conserved location and had similarly high
220 efficiencies across all isolates. For CrPV the following primers were used: Forward: 5'-
221 TTGGCGTGGTAGTATGCGTAT-3', Reverse: 5'-TGTTCCGTCCTGCGTCTC-3'. *RpL32* housekeeping gene
222 primers varied by species (S1 Text Tables B & C). For each sample, two technical replicates were
223 performed for each amplicon (viral and *RpL32*).
224

225 Between-plate variation in C_T values was estimated and corrected for using a linear model with plate
226 ID and biological replicate ID as parameters, as described elsewhere [68,69]. Mean viral C_T values
227 from technical replicate pairs were normalised to *RpL32* and converted to fold-change in viral load
228 using the $2^{-\Delta\Delta CT}$ method, where $\Delta C_T = C_{T:Virus} - C_{T:Rpl32}$, and $\Delta\Delta C_T = \Delta C_{T:day0} - \Delta C_{T:day2}$.

229 Amplification of the correct products was verified by melt curve analysis. Repeated failure to amplify
230 product, the presence of melt curve contaminants, or departures from the melt curve peaks of
231 positive samples (± 1.5˚C for viral amplicons, ± 3˚C for *Rpl32*) in either the 0 or 2 dpi samples were
232 used as exclusion criteria for biological replicates. In total, of the 180 unique combinations of host
233 species and virus measured, 3 biological replicates were obtained for 161 combinations, 2 replicates
234 for 18 combinations, and 1 replicate for 1 combination (*Drosophila virilis,* CrPV).
235

236 **Statistical Analysis**

237 Phylogenetic generalised linear mixed models were used to investigate the effects of host
238 relatedness on viral load, and to examine correlations between the different virus isolates.
239 Multivariate models were fitted using the R package MCMCglmm [70] with the viral load of each
240 virus isolate as the response variable. The structures of the models were as follows:

241 $$y_{hiv} = \beta_{1:v} + \mu_{p:hv} + \mu_{s:hv} + e_{hiv} \quad (1)$$

242 $$y_{hiv} = \beta_{1:v} + \mu_{p:hv} + e_{hiv} \quad (2)$$

243 In these models, y_{hiv} is the change in viral load for virus v in the ith biological replicate of host species
244 h. The fixed effect β_1 represents the intercepts for each virus isolate, the random effect μ_p represents
245 the effects of the host phylogeny assuming a Brownian motion model of evolution, and e represents
246 the model residuals. Model (1) also includes a species-specific random effect that is independent of
247 the host phylogeny ($\mu_{s:hv}$). This explicitly estimates the non-phylogenetic component of between-
248 species variance and allows the proportion of variance explained by the host phylogeny to be
249 calculated. $\mu_{s:hv}$ was removed from model (2) as model (1) struggled to separate the phylogenetic
250 and species-specific traits. Wing size, measured as the length of the IV longitudinal vein from the tip
251 of the proximal segment to the join of the distal segment with vein V [71], provided a proxy for body
252 size [72] and was included in a further model as a fixed effect ($wingsize_h.\beta_2$).
253

254 Within each of these models the random effects and residuals were assumed to follow a
255 multivariate normal distribution with a centred mean of zero and a covariance structure of $V_p \otimes A$ for
256 the phylogenetic effects, $V_s \otimes I$ for species-specific effects, and $V_e \otimes I$ residuals, where $\otimes$ represents
257 the Kronecker product. *A* represents the host phylogenetic relatedness matrix, *I* an identity matrix
258 and *V* represents 4x4 covariance matrices describing the between-species variances and covariances
259 of changes in viral load for the different viruses. Specifically, the matrices V_p and V_s describe the
260 phylogenetic and non-phylogenetic between-species variances in viral load for each virus and the
261 covariances between them, while the residual covariance matrix V_e describes within-species variance
262 that includes both true within-species effects and measurement errors. Since each biological
263 replicate was tested with a single virus isolate, the covariances of V_e cannot be estimated and were
264 set to zero.
265
266 Models were run for 13 million MCMC generations, sampled every 5,000 iterations with a burn-in of
267 3 million generations. Parameter expanded priors were placed on the covariance matrices, resulting
268 in multivariate F distributions with marginal variance distributions scaled by 1,000. Inverse-gamma
269 priors were placed on the residual variances, with a shape and scale equal to 0.002. To ensure the
270 model outputs were robust to changes in prior distribution, models were also fitted with flat and
271 inverse-Wishart priors, which gave qualitatively similar results.
272
273 The proportion of the between species variance that can be explained by the phylogeny was
274 calculated from model (1) using the equation $v_p /(v_p + v_s)$, where v_p and v_s represent the phylogenetic
275 and species-specific components of between-species variance [73], and is equivalent to phylogenetic
276 heritability or Pagel's lambda [74,75]. The repeatability of viral load measurements was calculated
277 from model (2) as $v_p/(v_p + v_e)$, where V_e is the residual variance of the model [76]. Inter-specific
278 correlations in viral load were calculated from model (2) V_p matrix as $cov_{x,y}/\sqrt{(var_x + var_y)}$. If
279 correlations between viruses are close to one (with no change in the variance whilst the means

280 remain constant) it would suggest there are no host species-by-virus interactions [22]. Parameter
281 estimates reported are means of the posterior density, and 95% credible intervals (95% CI) were
282 taken to be the 95% highest posterior density intervals.
283
284 The data files and R scripts used in this study are available in an online repository:
285 https://doi.org/10.6084/m9.figshare.13750711.v1.
286

287 ## Results

288 **Change in viral load is a repeatable trait among host species**

289 To investigate similarities between related viruses in the outcome of infection across host species, as
290 well as the potential for virus genotype to interact with host species effects, we experimentally
291 infected 45 species of *Drosophilidae* with four virus isolates: DCV-C, DCV-EB, DCV-M, and CrPV. The
292 DCV isolates formed a distinct clade (>93% genome and ORF amino acid identity, with 265-556 SNPs
293 between isolates), with the closest relationship between DCV-C and DCV-EB. CrPV formed an
294 outgroup to the DCV isolates (57-59% identity, with over 4000 SNPs between CrPV and each DCV
295 isolate, Fig. 1, Table 1). In total, 15,657 flies were inoculated, and the change in viral load after two
296 days of infection was determined by qRT-PCR (Fig. 2). The mean viral load within host species ranged
297 from an approximately 2.7-billion-fold increase in *Drosophila persimilis* infected with DCV-M to a 2.5-
298 fold decrease in *Zaprionus tuberculatus* infected with DCV-C. Viral loads across host species tended
299 to be higher for the DCV isolates, with a mean fold-increase of roughly 11,000 - 19,000, and lower
300 for CrPV, with a mean fold-increase of roughly 1,600.
301

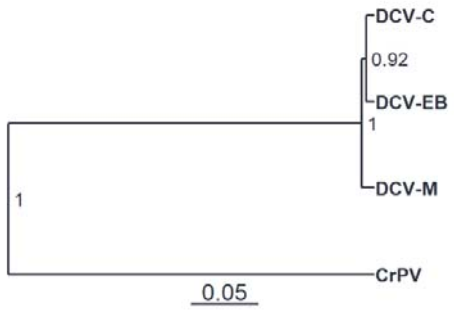

Figure 1: Phylogeny of virus isolates. Evolutionary relationships estimated from open reading frame (ORF) amino acid sequences presented in a midpoint-rooted tree. Node labels represent the posterior probabilities of each clade, and the scale bar represents amino acid substitutions per site.

	DCV-C	DCV-EB	DCV-M	CrPV
DCV-C		97.10%	94.00%	57.40%
DCV-EB	98.81%		93.90%	57.30%
DCV-M	98.30%	98.20%		57.40%
CrPV	59.00%	58.70%	58.70%	

Table 1: Virus isolate sequence similarity. Percentage sequence identity was calculated from multiple-alignment of whole genome nucleotides (white) or concatenated amino acid sequences of ORFs 1 & 2 (grey). Approximately 92 SNPs and 28 amino acid substitutions exist for every 1% of sequence divergence.

Phylogenetic generalised linear mixed models were fitted to the data to determine the proportion of variation in viral load explained by the host phylogeny (Table 2). The phylogeny explained 79% of the variation in viral load for CrPV but only 9-21% of the variation for the DCV isolates, with wide credible intervals on all the DCV estimates. This was due to the model struggling to separate phylogenetic and species-specific effects for these viruses. The repeatability of viral load across host species was high for both CrPV (0.66) and the DCV isolates (0.92-0.96), with the between-species phylogenetic component (v_p) explaining a high proportion of the variation in viral load with little within-species variation or measurement error (v_r). We found no significant effect of wing length (a proxy for host body size) on viral load for any of the included viruses, with all estimates having credible intervals overlapping zero.

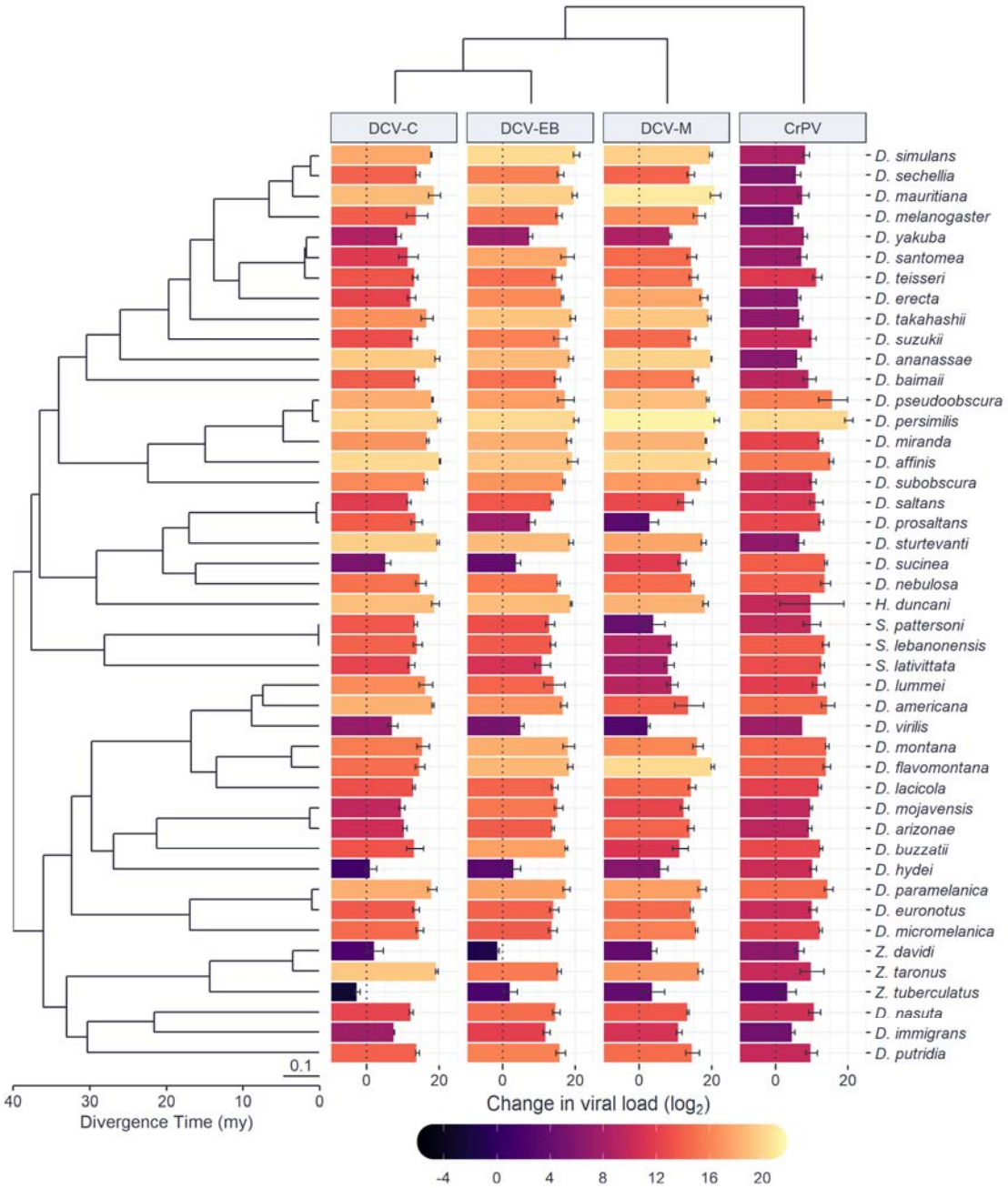

Figure 2: Change in viral load across a diverse panel of *Drosophilidae* host species for different virus isolates. Bar height and colour show the mean change in viral load at 2 dpi on a $\log_2$ scale, with error bars representing the standard error of the mean. The phylogeny of *Drosophilidae* hosts is presented on the left, with the scale bar representing the number of nucleotide substitutions per site and scale axis representing the approximate age since divergence in millions of years (my) based on estimates from [77,78]. The virus cladogram, presented at the top, is based on the evolutionary relationships shown in Fig. 1.

Virus	Mean change in viral load (log_2)	Repeatability	Variance explained by phylogeny
DCV-C	13.50 (95% CI: 11.17, 15.89)	0.96 (95% CI: 0.93, 0.98)	0.11 (95% CI: 0, 0.35)
DCV-EB	14.22 (95% CI: 11.42, 16.76)	0.96 (95% CI: 0.93, 0.98)	0.09 (95% CI: 0, 0.32)
DCV-M	13.63 (95% CI: 10.52, 16.59)	0.92 (95% CI: 0.87, 0.96)	0.23 (95% CI: 0, 0.51)
CrPV	10.66 (95% CI: 8.59, 12.66)	0.66 (95% CI: 0.46, 0.83)	0.79 (95% CI: 0.50, 1.00)

Table 2: Estimates of mean change in viral load, repeatability, and the proportion of variation explained by the host phylogeny. Estimates of the mean change in viral load and repeatability are taken from model (2), while estimates of the variation explained by the host phylogeny are taken from model (1).

Correlations between viruses are consistent with virus by host species interactions

Inter-specific correlations in viral load between viruses were then estimated from the variance-covariance matrices of model (2) (Fig. 3A). We found strong positive correlations between the DCV isolates ($r > 0.93$), with the strongest correlation between DCV-C and DCV-EB ($r = 0.97$). Correlations between DCV isolates and the more distantly related CrPV were positive ($r = 0.52$-0.59) but weaker than the correlations between the DCV isolates. The fact the DCV:CrPV correlations (and their 95% CI's) are not close to one is consistent with virus genotype by host species interactions on viral load [22]. This is further demonstrated by the notable differences in the rank order of host species susceptibility for each virus (Fig. 3B), equivalent to a crossing over of reaction norms for the susceptibility of host species between virus genotypes [79].

DCV-C was more strongly correlated to DCV-EB than to DCV-M ($\Delta r = 0.04$, 95% CI = 0.00, 0.09), and more strongly correlated to DCV-M than to CrPV ($\Delta r = 0.40$, 95% CI = 0.18, 0.82), consistent with an increase in the strength of correlation between viruses with closer evolutionary relatedness. Point estimates imply a similar pattern for DCV-EB, but the evidence for a stronger correlation with DCV-C than DCV-M was not well supported (S1 Text Table D).

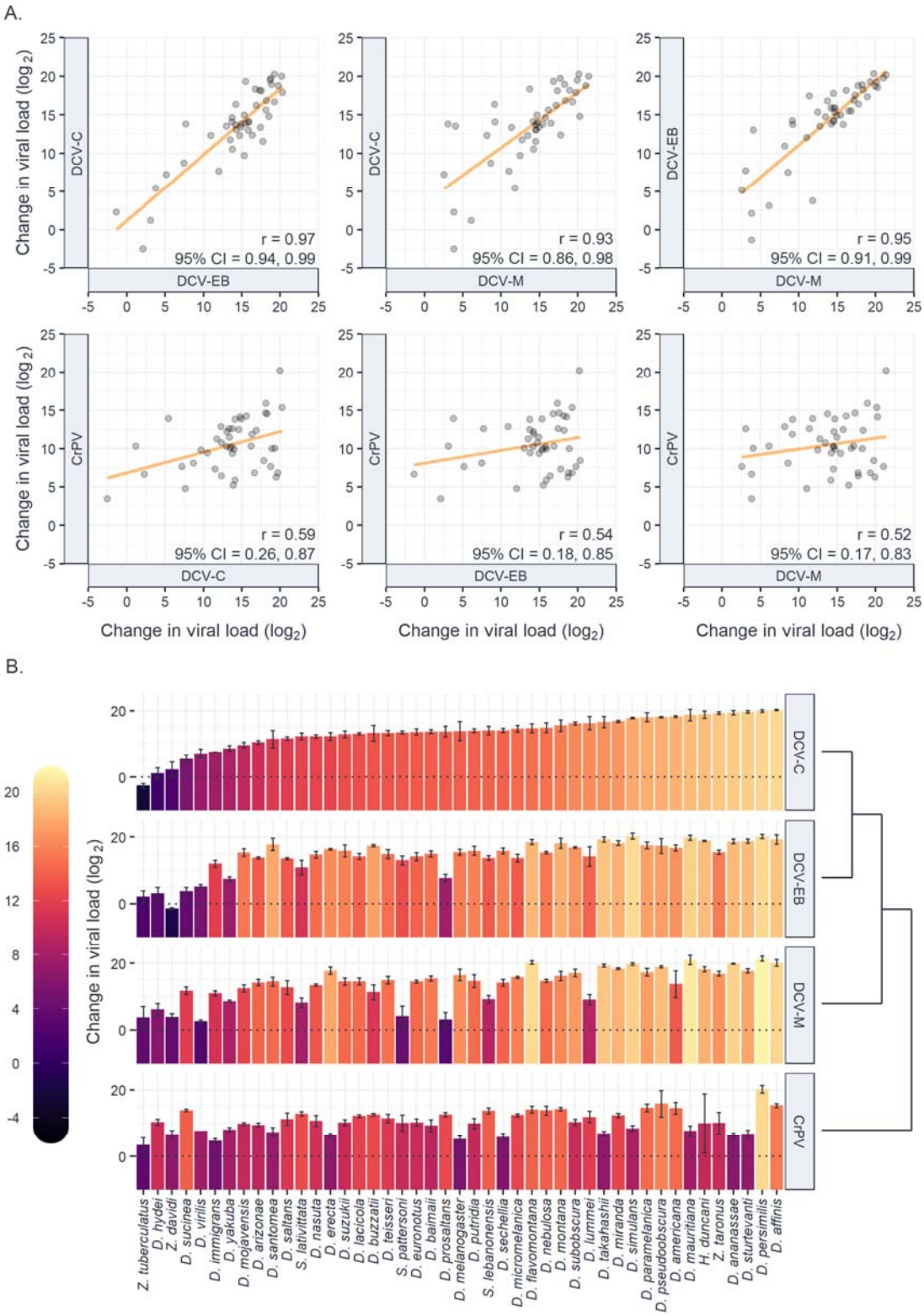

Figure 3: Similarities in infection outcome across host species and interactions between virus genotype and host species. A) Correlations in viral load between virus isolates. Individual points represent the mean change in viral load after 2 dpi for each host species on a $\log_2$ scale, and trendlines have been added from a univariate least-squares linear model for illustrative purposes. Correlations (r) are the total inter-specific correlations and

95% CIs from the output of model (2). B) Differences in the rank-order of host species susceptibility between virus isolates. Bar height and colour show the mean change in viral load at 2 dpi on a $\log_2$ scale, with error bars representing the standard error of the mean. The order of species along the x-axis has been sorted in ascending order of viral load during infection with DCV-C. Deviations from this rank-order of host species susceptibility for other viruses is indicative of crossing reaction norms and interactions between virus genotype and host species. The virus cladogram is based on the evolutionary relationships shown in Fig. 1.

Discussion

Closely related host species present similar environments to novel viruses [73,80], and so tend to share similar levels of susceptibility to a given virus [12–18]. Likewise, closely related viruses are often assumed to share characteristics that make their host interactions, transmission, and evolutionary trajectories comparable [28–35]. Here, we measured the strength of correlations in viral load between four *Cripavirus* isolates across 45 host species of *Drosophilidae*, to look for similarities between related viruses as well as evidence of virus genotype by host species interactions on the outcome of infection. We found positive correlations between every pair of viruses tested, indicating broad similarities in the outcome of infection across host species, but also evidence for interactions between virus genotype and host species with changes in the rank order of host species susceptibility between the different viruses (Fig. 3). This highlights the importance of considering both host and virus traits in understanding the outcomes of virus host shifts.

The strong positive correlations between DCV isolates are likely due to relatively high levels of sequence conservation resulting in only small differences in their ability to infect different host species. However, in other viruses a small number of mutations have been shown to allow successful infections in novel hosts [36,39]. We find a few instances of such effects here. For example, in *Zaprionus davidi*, DCV-EB shows a decline in viral load, suggesting it is failing to replicate and persist in this host species, whereas the other isolates show an increase in viral load in the same host. Similarly, *Scaptodrosophila pattersoni* is amongst the least susceptible to DCV-M but has relatively high viral loads for the other virus isolates. A greater number of these effects can be seen when

385 comparing hosts infected with DCV isolates to those infected with CrPV, where multiple species have
386 markedly different susceptibilities depending on the virus infecting them. For example, both
387 *Drosophila ananassae* and *Drosophila strutevanti* are within the five most susceptible species to
388 DCV-C, but also the eight least susceptible to CrPV. The weaker correlations that exist between DCV
389 and CrPV may be due to interactions with different host traits that vary in their patterns across the
390 host phylogeny. CrPV and DCV are known to have distinct methods of suppression of the host
391 antiviral RNAi pathway [56,57] and cause pathology in different tissues [62]. Additionally, their
392 relatively high levels of sequence divergence (57-59% identity) may have resulted in changes in the
393 ability of each virus to bind to host cell receptors, utilise host replication machinery, or avoid host
394 immune defences [81].
395
396 The existence of correlations between viruses suggests that host susceptibility is not specific to
397 individual viruses and that certain host clades may be broadly susceptible to infection. Host species
398 that are permissive to multiple virus genotypes may allow for the persistence of increased genetic
399 diversity in the virus population, allowing viruses to generate and maintain mutations that make
400 them more likely to emerge in novel host species [82,83]. They also have the potential to act as
401 "mixing vessels", providing increased opportunities for virus reassortment and recombination [84],
402 which has been proposed as a possible route for several viruses to acquire pandemic potential
403 [85,86]. Broadly susceptible host clades may therefore act as common reservoirs and donors of
404 emerging infectious diseases and identifying them in relevant systems could inform control and
405 prevention strategies [87].
406
407 The differences in correlation strength between pairs of viruses tended to follow differences in their
408 evolutionary divergence, such that more closely related pairs of viruses were more strongly
409 correlated in the outcome of infection across host species. This provides some tentative evidence
410 that the ability of a virus to infect a novel host may be inferred based on its evolutionary relatedness

411 to other viruses, although we lack the sample size or diversity of viruses needed to test this
412 conclusively. Numerous examples exist where a small number of genetic changes in viruses cause
413 large phenotypic differences [36–39], which would be exceptions to any link between correlation
414 strength and evolutionary relatedness [87]. Nevertheless, virus phylogenetic effects may still prove
415 to be a useful proxy for determining the likelihood of novel virus emergence. Further work is now
416 needed to expand the findings of this study to broader groups of viruses, and to test the importance
417 of the virus phylogeny in determining the potential outcomes of virus host shifts.
418

419 ## Acknowledgements

420 We would like to thank Julien Martinez, Valérie Dorey, and Maria Carla Saleh for kindly providing us
421 with DCV and CrPV virus isolates, Jarrod Hadfield for his advice on MCMCglmm model fitting, and
422 Joanne Lello for useful discussions. Ryan M. Imrie is supported by a studentship funded by the
423 Natural Environment Research Council (NERC) GW4+ Doctoral Training Partnership and the College
424 of Life and Environmental Sciences, University of Exeter. Ben Longdon and Katherine E. Roberts are
425 supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society
426 (Grant Number 109356/Z/15/Z). For the purpose of Open Access, the author has applied a CC BY
427 public copyright licence to any Author Accepted Manuscript version arising from this submission.
428

429 ## References

430 1. Woolhouse MEJ, Gowtage-Sequeria S. Host Range and Emerging and Reemerging Pathogens. Emerg Infect Dis.
431 2005;11(12):1842–7. doi: 10.3201/eid1112.050997

432 2. Cleaveland S, Laurenson MK, Taylor LH. Diseases of humans and their domestic mammals: pathogen characteristics, host
433 range and the risk of emergence. Philosophical Transactions Royal Soc Lond Ser B Biological Sci. 2001;356(1411):991–9.
434 doi: 10.1098/rstb.2001.0889

435 3. Taylor LH, Latham SM, woolhouse MEJ. Risk factors for human disease emergence. Philosophical Transactions Royal Soc
436 Lond Ser B Biological Sci. 2001;356(1411):983–9. doi: 10.1098/rstb.2001.0888

437 4. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of
438 human coronaviruses. J Adv Res. 2020;24:91–8. doi: 10.1016/j.jare.2020.03.005

439 5. Miller IF, Becker AD, Grenfell BT, Metcalf CJE. Disease and healthcare burden of COVID-19 in the United States. Nat Med.
440 2020;26(8):1212–7. doi: 10.1038/s41591-020-0952-y

441 6. Frank TD, Carter A, Jahagirdar D, Biehl MH, Douwes-Schultz D, Larson SL et al. Global, regional, and national incidence,
442 prevalence, and mortality of HIV, 1980–2017, and forecasts to 2030, for 195 countries and territories: a systematic analysis
443 for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. Lancet Hiv. 2019;6(12):e831–59. doi:
444 10.1016/s2352-3018(19)30196-1

445 7. Huber C, Finelli L, Stevens W. The Economic and Social Burden of the 2014 Ebola Outbreak in West Africa. J Infect Dis.
446 2018;218(Supplement_5):S698–704. doi: 10.1093/infdis/jiy213

447 8. Holmes EC. What can we predict about viral evolution and emergence? Curr Opin Virol. 2013;3(2):180–4. doi:
448 10.1016/j.coviro.2012.12.003

449 9. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL et al. Pathways to zoonotic spillover. Nat Rev
450 Microbiol. 2017;15(8):502–10. doi: 10.1038/nrmicro.2017.45

451 10. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover
452 from mammals. Nature. 2017;546(7660):646–50. doi: 10.1038/nature22975

453 11. Geoghegan JL, Holmes EC. Predicting virus emergence amid evolutionary noise. Open Biol. 2017;7(10):170189. doi:
454 10.1098/rsob.170189

455 12. Gilbert GS, Webb CO. Phylogenetic signal in plant pathogen–host range. Proc National Acad Sci. 2007;104(12):4979–83.
456 doi: 10.1073/pnas.0607968104

457 13. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host Phylogeny Constrains Cross-
458 Species Emergence and Establishment of Rabies Virus in Bats. Science. 2010;329(5992):676–9. doi:
459 10.1126/science.1188836

460 14. Guth S, Visher E, Boots M, Brook CE. Host phylogenetic distance drives trends in virus virulence and transmissibility
461 across the animal–human interface. Philosophical Transactions Royal Soc B Biological Sci. 2019;374(1782):20190296. doi:
462 10.1098/rstb.2019.0296

463 15. Farrell MJ, Davies TJ. Disease mortality in domesticated animals is predicted by host evolutionary relationships. Proc
464 National Acad Sci. 2019;116(16):201817323. doi: 10.1073/pnas.1817323116

465 16. Mollentze N, Streicker DG, Murcia PR, Hampson K, Biek R. Virulence mismatches in index hosts shape the outcomes of
466 cross-species transmission. Proc National Acad Sci. 2020;117(46):28859–66. doi: 10.1073/pnas.2006778117

467 17. Longdon B, Hadfield JD, Day JP, Smith SCL, McGonigle JE, Cogni R et al. The Causes and Consequences of Changes in
468 Virulence following Pathogen Host Shifts. Plos Pathog. 2015;11(3):e1004728. doi: 10.1371/journal.ppat.1004728

469 18. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM. Host Phylogeny Determines Viral Persistence and
470 Replication in Novel Hosts. Plos Pathog. 2011;7(9):e1002260. doi: 10.1371/journal.ppat.1002260

471 19. Davies TJ, Pedersen AB. Phylogeny and geography predict pathogen community similarity in wild primates and humans.
472 Proc Royal Soc B Biological Sci. 2008;275(1643):1695–701. doi: 10.1098/rspb.2008.0284

473 20. Shaw LP, Wang AD, Dylus D, Meier M, Pogacnik G, Dessimoz C et al. The phylogenetic range of bacterial and viral
474 pathogens of vertebrates. Mol Ecol. 2020;29(17):3361–79. doi: 10.1111/mec.15463

475 21. Albery GF, Eskew EA, Ross N, Olival KJ. Predicting the global mammalian viral sharing network using phylogeography.
476 Nat Commun. 2020;11(1):2260. doi: 10.1038/s41467-020-16153-4

477 22. Hudson AI, Fleming-Davies AE, Páez DJ, Dwyer G. Genotype-by-genotype interactions between an insect and its
478 pathogen. J Evolution Biol. 2016;29(12):2480–90. doi: 10.1111/jeb.12977

479 23. Vienne DMD, Hood ME, Giraud T. Phylogenetic determinants of potential host shifts in fungal pathogens. J Evolution
480 Biol. 2009;22(12):2532–41. doi: 10.1111/j.1420-9101.2009.01878.x

481 24. Hadfield JD, Krasnov BR, Poulin R, Nakagawa S. A Tale of Two Phylogenies: Comparative Analyses of Ecological
482 Interactions. Am Nat. 2014;183(2):174–87. doi: 10.1086/674445

483 25. Kitchen A, Shackelton LA, Holmes EC. Family level phylogenies reveal modes of macroevolution in RNA viruses. Proc
484 National Acad Sci. 2011;108(1):238–43. doi: 10.1073/pnas.1011090108

485 26. Jenkins CN, Pimm SL, Joppa LN. Global patterns of terrestrial vertebrate diversity and conservation. Proc National Acad
486 Sci. 2013;110(28):E2602–10. doi: 10.1073/pnas.1302251110

487 27. Geoghegan JL, Duchêne S, Holmes EC. Comparative analysis estimates the relative frequencies of co-divergence and
488 cross-species transmission within viral families. Plos Pathog. 2017;13(2):e1006215. doi: 10.1371/journal.ppat.1006215

489 28. Zhu Z, Lian X, Su X, Wu W, Marraro GA, Zeng Y. From SARS and MERS to COVID-19: a brief summary and comparison of
490 severe acute respiratory infections caused by three highly pathogenic human coronaviruses. Respir Res. 2020;21(1):224.
491 doi: 10.1186/s12931-020-01479-w

492 29. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F et al. Comparing SARS-CoV-2 with SARS-CoV and
493 influenza pandemics. Lancet Infect Dis. 2020;20(9):e238–44. doi: 10.1016/s1473-3099(20)30484-9

494 30. Hsieh Y-H. Age groups and spread of influenza: implications for vaccination strategy. Bmc Infect Dis. 2010;10(1):106.
495 doi: 10.1186/1471-2334-10-106

496 31. Chao DL, Halloran ME, Obenchain VJ, Longini IM. FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation
497 Model. Plos Comput Biol. 2010;6(1):e1000656. doi: 10.1371/journal.pcbi.1000656

498 32. Gomes MFC, Piontti AP y, Rossi L, Chao D, Longini I, Halloran ME et al. Assessing the International Spreading Risk
499 Associated with the 2014 West African Ebola Outbreak. Plos Curr Outbreaks. 2014;6. doi:
500 10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5

501 33. Du X, King AA, Woods RJ, Pascual M. Evolution-informed forecasting of seasonal influenza A (H3N2). Sci Transl Med.
502 2017;9(413):eaan5325. doi: 10.1126/scitranslmed.aan5325

503 34. Ruiz SI, Zumbrun EE, Nalca A. Animal Models for the Study of Human Disease (Second Edition). Part K Viral Dis.
504 2017;853–901. doi: 10.1016/b978-0-12-809468-6.00033-4

505 35. Cann JA, Jahrling PB, Hensley LE, Wahl-Jensen V. Comparative Pathology of Smallpox and Monkeypox in Man and
506 Macaques. J Comp Pathol. 2013;148(1):6–21. doi: 10.1016/j.jcpa.2012.06.007

507 36. Allison AB, Organtini LJ, Zhang S, Hafenstein SL, Holmes EC, Parrish CR. Single Mutations in the VP2 300 Loop Region of
508 the Three-Fold Spike of the Carnivore Parvovirus Capsid Can Determine Host Range. J Virol. 2016;90(2):753–67. doi:
509 10.1128/jvi.02636-15

510 37. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A Single Mutation in Chikungunya Virus Affects Vector Specificity
511 and Epidemic Potential. Plos Pathog. 2007;3(12):e201. doi: 10.1371/journal.ppat.0030201

512 38. Wong G, He S, Leung A, Cao W, Bi Y, Zhang Z et al. Naturally-occurring single mutations in Ebola observably impact
513 infectivity. J Virol. 2018;93(1):JVI.01098-18. doi: 10.1128/jvi.01098-18

514 39. Vries RP de, Peng W, Grant OC, Thompson AJ, Zhu X, Bouwman KM et al. Three mutations switch H7N9 influenza to
515 human-type receptor specificity. Plos Pathog. 2017;13(6):e1006390. doi: 10.1371/journal.ppat.1006390

516 40. Holmes EC. The Evolutionary Genetics of Emerging Viruses. Annu Rev Ecol Evol Syst. 2009;40(1):353–72. doi:
517 10.1146/annurev.ecolsys.110308.120248

518 41. Müller NF, Stolz U, Dudas G, Stadler T, Vaughan TG. Bayesian inference of reassortment networks reveals fitness
519 benefits of reassortment in human influenza viruses. Proc National Acad Sci. 2020;117(29):17104–11. doi:
520 10.1073/pnas.1918304117

521 42. Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. Recombination in viruses: Mechanisms, methods
522 of study, and evolutionary consequences. Infect Genetics Evol. 2015;30:296–307. doi: 10.1016/j.meegid.2014.12.022

523 43. Zhang Y-Z, Chen Y-M, Wang W, Qin X-C, Holmes EC. Expanding the RNA Virosphere by Unbiased Metagenomics. Ann
524 Rev Virol. 2019;6(1):1–21. doi: 10.1146/annurev-virology-092818-015851

525 44. Zanotto PM, Gibbs MJ, Gould EA, Holmes EC. A reevaluation of the higher taxonomy of viruses based on RNA
526 polymerases. J Virol. 1996;70(9):6083–96. doi: 10.1128/jvi.70.9.6083-6096.1996

527 45. Johnson KN, Christian PD. Molecular Characterization of Drosophila C Virus Isolates. J Invertebr Pathol. 1999;73(3):248–
528 54. doi: 10.1006/jipa.1998.4830

529 46. Cherry S, Silverman N. Host-pathogen interactions in drosophila: new tricks from an old friend. Nat Immunol.
530 2006;7(9):911–7. doi: 10.1038/ni1388

531 47. Bonning BC, Miller WA. Dicistroviruses. Entomology. 2010;55(1):129–50. doi: 10.1146/annurev-ento-112408-085457

532 48. Reinganum C, O'Loughlin GT, Hogan TW. A nonoccluded virus of the field crickets Teleogryllus oceanicus and T.
533 commodus (Orthoptera: Gryllidae). J Invertebr Pathol. 1970;16(2):214–20. doi: 10.1016/0022-2011(70)90062-5

534 49. Scotti PD, Longworth JF, Plus N, Croizier G, Reinganum C. The Biology and Ecology of Strains of an Insect Small RNA
535 Virus Complex. Adv Virus Res. 1981;26:117–43. doi: 10.1016/s0065-3527(08)60422-4

536 50. Nayak A, Kim DY, Trnka MJ, Kerr CH, Lidsky PV, Stanley DJ et al. A Viral Protein Restricts Drosophila RNAi Immunity by
537 Regulating Argonaute Activity and Stability. Cell Host Microbe. 2018;24(4):542-557.e9. doi: 10.1016/j.chom.2018.09.006

538 51. Majzoub K, Hafirassou ML, Meignin C, Goto A, Marzi S, Fedorova A et al. RACK1 Controls IRES-Mediated Translation of
539 Viruses. Cell. 2014;159(5):1086–95. doi: 10.1016/j.cell.2014.10.041

540 52. Magwire MM, Fabian DK, Schweyen H, Cao C, Longdon B, Bayer F et al. Genome-Wide Association Studies Reveal a
541 Simple Genetic Basis of Resistance to Naturally Coevolving Viruses in Drosophila melanogaster. Plos Genet.
542 2012;8(11):e1003057. doi: 10.1371/journal.pgen.1003057

543 53. Cao C, Cogni R, Barbier V, Jiggins FM. Complex Coding and Regulatory Polymorphisms in a Restriction Factor Determine
544 the Susceptibility of Drosophila to Viral Infection. Genetics. 2017;206(4):2159–73. doi: 10.1534/genetics.117.201970

545 54. Cogni R, Cao C, Day JP, Bridson C, Jiggins FM. The genetic architecture of resistance to virus infection in Drosophila. Mol
546 Ecol. 2016;25(20):5228–41. doi: 10.1111/mec.13769

547 55. Martins NE, Faria VG, Nolte V, Schlötterer C, Teixeira L, Sucena É et al. Host adaptation to viruses relies on few genes
548 with different cross-resistance properties. Proc National Acad Sci. 2014;111(16):5938–43. doi: 10.1073/pnas.1400378111

549 56. Nayak A, Berry B, Tassetto M, Kunitomi M, Acevedo A, Deng C et al. Cricket paralysis virus antagonizes Argonaute 2 to
550 modulate antiviral defense in Drosophila. Nat Struct Mol Biol. 2010;17(5):547–54. doi: 10.1038/nsmb.1810

551 57. Rij RP van, Saleh M-C, Berry B, Foo C, Houk A, Antoniewski C et al. The RNA silencing endonuclease Argonaute 2
552 mediates specific antiviral immunity in Drosophila melanogaster. Gene Dev. 2006;20(21):2985–95. doi:
553 10.1101/gad.1482006

554 58. Chtarbanova S, Lamiable O, Lee K-Z, Galiana D, Troxler L, Meignin C et al. Drosophila C Virus Systemic Infection Leads to
555 Intestinal Obstruction. J Virol. 2014;88(24):14057–69. doi: 10.1128/jvi.02320-14

556 59. Martinez J, Bruner-Montero G, Arunkumar R, Smith SCL, Day JP, Longdon B et al. Virus evolution in Wolbachia-infected
557 Drosophila. Proc Royal Soc B. 2019;286(1914):20192117. doi: 10.1098/rspb.2019.2117

558 60. Longdon B, Day JP, Alves JM, Smith SCL, Houslay TM, McGonigle JE et al. Host shifts result in parallel genetic changes
559 when viruses evolve in closely related species. Plos Pathog. 2018;14(4):e1006951. doi: 10.1371/journal.ppat.1006951

560 61. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol.
561 2012;29(8):1969–73. doi: 10.1093/molbev/mss075

562 62. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using
563 Tracer 1.7. Systematic Biol. 2018;67(5):901–4. doi: 10.1093/sysbio/syy032

564 63. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinform. 2020;69(1):e96. doi:
565 10.1002/cpbi.96

566 64. Johnson KN, Christian PD. A molecular taxonomy for cricket paralysis virus including two new isolates from Australian
567 populations ofDrosophila (Diptera: Drosophilidae). Arch Virol. 1996;141(8):1509–22. doi: 10.1007/bf01718251

568 65. Drosophila Ringer's solution. Cold Spring Harbor Protocols. 2007; doi: 10.1101/pdb.rec10919

569 66. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc National Acad Sci.
570 1992;89(22):10915–9. doi: 10.1073/pnas.89.22.10915

571 67. Landum M, Silva MS, Martins N, Teixeira L. Viral route of infection determines the effect of Drosophila
572 melanogaster gut bacteria on host resistance and tolerance to disease. Biorxiv [Internet]. 2021 Jan
573 1;2021.02.18.431843. doi: 10.1101/2021.02.18.431843 Available from:
574 http://biorxiv.org/content/early/2021/02/18/2021.02.18.431843.abstract

575 68. Ruijter JM, Thygesen HH, Schoneveld OJ, Das AT, Berkhout B, Lamers WH. Factor correction as a tool to eliminate
576 between-session variation in replicate experiments: application to molecular biology and retrovirology. Retrovirology.
577 2006;3(1):2. doi: 10.1186/1742-4690-3-2

578 69. Ruijter JM, Villalba AR, Hellemans J, Untergasser A, Hoff MJB van den. Removal of between-run variation in a multi-
579 plate qPCR experiment. Biomol Detect Quantification. 2015;5:10–4. doi: 10.1016/j.bdq.2015.07.001

580 70. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package.
581 Journal of Statistical Software [Internet]. 2010;33(2):1–22. Available from: http://www.jstatsoft.org/v33/i02/

582 71. Gilchrist GW, Huey RB, Serra L. Rapid evolution of wing size clines in Drosophila subobscura. Genetica. 2001;112–
583 113(1):273–86. doi: 10.1023/a:1013358931816

584 72. Huey RB, Moreteau B, Moreteau J-C, Gibert P, Gilchrist GW, Ives AR et al. Sexual size dimorphism in a Drosophila clade,
585 the D. obscura group. Zoology. 2006;109(4):318–30. doi: 10.1016/j.zool.2006.04.003

586 73. Freckleton RP, Harvey PH, Pagel M. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. Am
587 Nat. 2002;160(6):712–26. doi: 10.1086/343873

588 74. Pagel M. Inferring the historical patterns of biological evolution. Nature. 1999;401(6756):877–84. doi: 10.1038/44766

589 75. Housworth EA, Martins EP, Lynch M. The Phylogenetic Mixed Model. Am Nat. 2004;163(1):84–96. doi: 10.1086/380570

590 76. Falconer D. Introduction to quantitative genetics. 4th ed. Pearson Education India; 1996.

591 77. Russo CA, Takezaki N, Nei M. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol.
592 1995;12(3):391–404. doi: 10.1093/oxfordjournals.molbev.a040214

593 78. Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. Estimating Divergence Dates and Substitution
594 Rates in the Drosophila Phylogeny. Mol Biol Evol. 2012;29(11):3459–73. doi: 10.1093/molbev/mss150

595 79. INGLEBY FC, HUNT J, HOSKEN DJ. The role of genotype-by-environment interactions in sexual selection. J Evolution Biol.
596 2010;23(10):2031–45. doi: 10.1111/j.1420-9101.2010.02080.x

597 80. Poulin R, Krasnov BR, Mouillot D, Thieltges DW. The comparative ecology and biogeography of parasites. Philosophical
598 Transactions Royal Soc B Biological Sci. 2011;366(1576):2379–90. doi: 10.1098/rstb.2011.0048

599 81. Rothenburg S, Brennan G. Species-Specific Host–Virus Interactions: Implications for Viral Host Range and Virulence.
600 Trends Microbiol. 2019;28(1):46–56. doi: 10.1016/j.tim.2019.08.007

601 82. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. Philosophical
602 Transactions Royal Soc B Biological Sci. 2012;367(1604):2864–71. doi: 10.1098/rstb.2011.0354

603 83. Woolhouse MEJ, Adair K, Brierley L. RNA Viruses: A Case Study of the Biology of Emerging Infectious Diseases. Microbiol
604 Spectr. 2014;1(1):83–97. doi: 10.1128/microbiolspec.oh-0001-2012

605 84. Zhang H, Li H, Wang W, Wang Y, Han G-Z, Chen H et al. A unique feature of swine ANP32A provides susceptibility to
606 avian influenza virus infection in pigs. Plos Pathog. 2020;16(2):e1008330. doi: 10.1371/journal.ppat.1008330

607 85. Silva UC de, Tanaka H, Nakamura S, Goto N, Yasunaga T. A comprehensive analysis of reassortment in influenza A virus.
608 Biol Open. 2012;1(4):385–90. doi: 10.1242/bio.2012281

609 86. Goldstein SA, Brown J, Pedersen BS, Quinlan AR, Elde NC. Extensive recombination-driven coronavirus diversification
610 expands the pool of potential pandemic pathogens. Biorxiv. 2021;2021.02.03.429646. doi: 10.1101/2021.02.03.429646

611 87. Streicker DG, Gilbert AT. Contextualizing bats as viral reservoirs. Science. 2020;370(6513):172–3. doi:
612 10.1126/science.abd4559

613