Virus relatedness predicts susceptibility in novel host species

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

As a major source of outbreaks and emerging infectious diseases, virus host shifts cause significant health, social and economic damage. Predicting the outcome of infection with novel combinations of virus and host remains a key challenge in virus research. Host evolutionary relatedness can explain variation in transmission rates, virulence, and virus community composition between host species, but there is much to learn about the potential for virus evolutionary relatedness to explain variation in the ability of viruses to infect novel hosts. Here, we measure correlations in the outcomes of infection across 45 *Drosophilidae* host species with four *Cripavirus* isolates that vary in their evolutionary relatedness. We found positive correlations between every pair of viruses tested, with the strength of correlation tending to decrease with greater evolutionary distance between viruses. These results suggest that virus evolutionary relatedness can explain variation in the outcome of host shifts and may be a useful proxy for determining the likelihood of novel virus emergence.

Author Summary

Many outbreaks and new infectious diseases are caused by viruses jumping into new host species. Predicting when jumps will occur, what the characteristics of viruses will be, and how they are likely to evolve in their new hosts are major challenges, requiring an understanding of the interactions between virus and host that is not often available for new viruses or variants. Instead, while studies of the new virus are ongoing, we rely on comparisons to closely related viruses to fill the gaps in our knowledge. How much we can trust these comparisons to be informative, and how their usefulness decreases when comparing viruses that are more distantly related, has not been well studied. Here we used four viruses of fruit flies with different levels of relatedness to begin to answer these questions. We found that, across 45 different fly species, infections were more similar in closely related viruses. The similarity of infections between any two viruses closely matched how similar the genomes of each virus were, suggesting this may be a good rule of thumb to assess how reliable comparisons between viruses are, and how well we can estimate the likelihood of new viruses jumping between host species.

51 ## Introduction

52 Virus host shifts, where viruses jump to and establish onward transmission in novel host species, are
53 a major source of outbreaks and emerging infectious diseases [1–3]. Many human infections,
54 including HIV, Ebola, and recently SARS-CoV-2, have shifted into humans from other species, and
55 continue to cause significant health, social and economic damage [4–7]. Predicting and preventing
56 virus host shifts have consequently become major goals of virus research [8]. Many challenges
57 remain in achieving these goals, including improving our understanding of the host, virus, and
58 ecological factors that influence the outcome of host shifts [9,10], and deciding on the best strategy
59 for surveillance [11,12]. A better grasp of the evolutionary and epidemiological trajectories of viruses
60 after host shifts is also required to infer which pathogens may become established in novel host
61 species [13].
62
63 Several studies have investigated how host evolutionary relatedness can explain variation in the
64 outcome of infection across host species. Greater phylogenetic distance between the natural
65 (donor) and recipient hosts is associated with decreased likelihood of cross-species transmission
66 [14,15] and reduced onward transmission within the novel host species [16]. Phylogenetic distance
67 between hosts also explains variation in virulence after cross-species transmission, which increases
68 when viruses jump between more distantly related hosts [16–18]. Groups of closely related hosts
69 have also been shown to share similar levels of susceptibility to novel viruses, independent of the
70 distance to the natural host [19,20], and harbour similar virus communities [21–23].
71
72 Fewer studies have investigated how the evolutionary relatedness and genetics of viruses influence
73 the outcome of host shifts. Within species, genotype-by-genotype interactions between hosts and
74 viruses can be important in determining the outcomes of infection [24–26], but little is known about
75 the importance of genetic differences between viruses in determining infection across host species
76 [20]. Following host shifts, virus genomes accumulate mutations as they adapt to the novel host

77 [27,28], and higher pre-standing genetic variation may make some viruses more likely to emerge
78 than others [29,30]. Despite high rates of geographical overlap between potential hosts [31], closely
79 related viruses tend to infect the same broad host taxa [32], suggesting they share similar
80 constraints on their host ranges. However, shifts between divergent host species are also common
81 across every virus family [33]. A study of fungal pathogens of plants found that, in addition to the
82 host phylogeny being an important determinant of a pathogen's ability to infect novel host species,
83 related pathogens share similar host ranges [34]. It has yet to be determined if a similar pathogen
84 phylogenetic effect exists in viruses.
85
86 Although we do not know the extent to which closely related viruses share traits, inferring the
87 characteristics of viruses from better studied relatives is sometimes necessary. This is often the case
88 during the early stages of outbreaks, where primary research on new viruses or variants is not yet
89 available. When SARS-CoV-2 first emerged, its characteristics and epidemiological trajectory were
90 inferred from the closely related SARS-CoV, MERS-CoV, endemic coronaviruses responsible for
91 common colds [35], and from other pandemic respiratory viruses such as influenza [36].
92 Comparisons to previous outbreaks were also used to parameterise disease models in the 2009
93 H1N1 pandemic [37,38], the 2014 Ebolavirus outbreak [39], and in forecast models of seasonal
94 influenza [40]. Even for viruses that are not newly emerged, many experimental models of infection
95 rely on surrogates when the virus of interest is unavailable, non-permissive in cell culture or animal
96 models, or requires considerable adaptation to experimental hosts [41,42].
97
98 In these approaches, the evolutionary scale at which comparisons are being made is typically not
99 quantified, relying instead on the classification of different virus strains and species [43]. As such, it
100 is difficult to assess their ability to predict the characteristics of newly emerging viruses.
101 Comparisons between more distantly related viruses, such as bat and canine rabies viruses [44] and
102 diverged lineages of influenza viruses [45–47], found stark differences across larger evolutionary

103 scales. Given the huge diversity of viruses in wildlife [48], and our inability to exhaustively test every
104 feasible virus variant [8], a greater understanding of the correlations that exist between viruses, and
105 the evolutionary distances at which these correlations break down, is necessary to understand the
106 likelihood of novel virus emergence.
107
108 In this study, we have investigated how patterns of host susceptibility are correlated between
109 closely related viruses, using experimental infections with four *Cripavirus* isolates (family
110 *Dicistroviridae*) across a panel of 45 host species of *Drosophilidae*. Three of the viruses are isolates of
111 Drosophila C virus (DCV-C, DCV-EB and DCV-M), a well-studied virus isolated from *Drosophila*
112 *melanogaster* [49]. The fourth virus is the closely related Cricket Paralysis virus (CrPV) which was
113 isolated from Australian field crickets (*Teleogryllus commodus*) and is a widely used model insect
114 pathogen [50–52]. Both DCV and CrPV infect a broad range of host species [53], cause virulent
115 infections in adult flies [19,54], and share similar mechanisms for co-opting the host translation
116 machinery [55]. A major-effect resistance gene called *pastrel* increases resistance to DCV in *D.*
117 *melanogaster* [56–58] and has also been shown to provide cross-resistance to CrPV along with
118 another gene, *Ubc-E2H* [59]. 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 [60,61]. 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 [62].
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* [63].
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 [19]. 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* [28].
132 Here we measured the ability of four viruses to replicate and persist across host species and
133 examined correlations between these viruses in the context of their evolutionary relatedness.
134

135 Materials & Methods

136 **Fly Stocks**

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

144 **Host Phylogeny**

145 The method used to infer the host phylogeny has been described in detail elsewhere [19]. Briefly,
146 publicly available sequences of the *28S*, *Adh*, *Amyrel*, *COI*, *COII*, *RpL32*, and *SOD* genes were collected
147 from Genbank (see https://doi.org/10.6084/m9.figshare.13079366.v1 for a full breakdown of genes
148 and accessions by species). Gene sequences were aligned in Geneious v9.1.8
149 (https://www.geneious.com) using a progressive pairwise global alignment algorithm with free end
150 gaps and a 70% similarity IUB cost matrix. Gap open penalties, gap extension penalties, and
151 refinement iterations were kept as default.
152

153 Phylogenetic reconstruction was performed using BEAST v1.10.4 [64] with genes partitioned into
154 separate ribosomal (*28S*), mitochondrial (*COI*, *COII*), and nuclear (*Adh*, *Amyrel*, *RpL32*, *SOD*) groups.

155 The mitochondrial and nuclear groups were further partitioned into groups for codon position 1+2
156 and codon position 3, with unlinked substitution rates and base frequencies across codon positions.
157 Each group was fitted to separate relaxed uncorrelated lognormal molecular clock models using
158 random starting trees and 4-category gamma-distributed HKY substitution models. The BEAST
159 analysis was run twice, with 1 billion MCMCglmm generations sampled every 100,000 iterations,
160 using a birth-death process tree-shape prior. Model trace files were evaluated for chain
161 convergence, sampling, and autocorrelation using Tracer v1.7.1 [65]. The reconstructed tree was
162 visualised using ggtree v2.0.4 [66].
163

164 **Virus Isolates**

165 Virus stocks were kindly provided by Julien Martinez (DCV isolates) [63], and Valérie Dorey and
166 Maria Carla Saleh (CrPV) [61]. DCV-C, DCV-EB and DCV-M were originally isolated from fly stocks
167 with origins in three separate continents; DCV-C and DCV-EB were isolated from lab stocks
168 established by wild capture in Charolles, France and Ellis Beach, Australia respectively, while DCV-M
169 was isolated directly from wild flies in Marrakesh, Morocco [49]. The CrPV isolate was collected from
170 *Teleogryllus commodus* in Victoria, Australia [67]. Virus stocks were diluted in Ringers solution [68]
171 to equalise the relative amounts of viral RNA and checked for contamination with CrPV (DCV
172 isolates) and DCV (CrPV isolate) by qRT-PCR as described below.
173

174 **Virus phylogeny**

175 Full genome sequences for DCV-C (*MK645242*), DCV-EB (*MK645239*), DCV-M (*MK645243*), and CrPV
176 (*NC_003924*) were retrieved from the NCBI Nucleotide database. Annotations of ORFs for the
177 replicase polyprotein (CrPV: *Q9IJX4*, DCV: *O36966*) and structural polyprotein (CrPV: *P13418*, DCV:
178 *O36967*) were collected from the UniProtKB database and used to separate the coding and non-
179 coding regions of each virus. ORF sequences were concatenated and aligned using the Geneious

180 progressive pairwise translation alignment algorithm with a Blosum50 cost matrix and default
181 parameters. Alignments were manually checked for quality and sequences aligning to CrPV ORF1
182 nucleotides 1-387 and 2704-2728 were removed due to the presence of large indels.
183 Phylogenetic reconstruction was performed using BEAST v1.10.4 with translated ORF sequences
184 fitted to an uncorrelated relaxed lognormal molecular clock model using a speciation birth-death
185 process tree-shape prior. A Blosum62 substitution model [69] with a gamma distribution of rate
186 variation with four categories and a proportion of invariable sites was used. The model was run for
187 10 million MCMC generations sampled every 1,000 iterations and evaluated in Tracer v1.7.1 as
188 above.
189

190 **Inoculation**

191 Before inoculation, 0-1 day old male flies were kept in vials containing cornmeal media (S1 Text
192 Supplementary Methods), and were transferred to fresh media every 2 days for one week. Vials
193 contained between 5 and 20 flies (mean = 14.5) and were kept at 22˚C at 70% relative humidity in a
194 12-hour light-dark cycle. Flies were inoculated at 7-8 days old under CO_2 anaesthesia via septic pin
195 prick with 12.5µm diameter stainless steel needles (Fine Science Tools, CA, USA). These needles
196 were bent approximately 250µm from the end to provide a depth stop and dipped in virus solution
197 before being pricked into the pleural suture of each fly. Inoculated flies were then snap frozen
198 immediately in liquid nitrogen, providing a 0 days-post-infection (dpi) timepoint, or maintained in
199 cornmeal vials for a further 2 days ± 3 hours before freezing, providing a 2 dpi time point. Within
200 replicate blocks the 0 and 2 dpi vials for each virus were inoculated on the same day, and together
201 constituted one biological replicate. We aimed to collect three biological replicates for each species
202 and virus combination, with the order of species, vial (0 or 2 dpi), and virus randomised for each
203 replicate block.
204

205

206 **Measuring Change in Viral Load**

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

213

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

220

221 Between-plate variation in C_T values was estimated and corrected for using a linear model with plate
222 ID and biological replicate ID as parameters, as described elsewhere [70,71]. Mean viral C_T values
223 from technical replicate pairs were normalised to *RpL32* and converted to fold-change in viral load
224 using the $2^{-\Delta\Delta CT}$ method, where $\Delta C_T = C_{T:Virus} - C_{T:Rpl32}$, and $\Delta\Delta C_T = \Delta C_{T:day0} - \Delta C_{T:day2}$.
225 Amplification of the correct products was verified by melt curve analysis. Repeated failure to amplify
226 product, the presence of melt curve contaminants, or departures from the melt curve peaks of
227 positive samples (± 1.5˚C for viral amplicons, ± 3˚C for *Rpl32*) in either the 0 or 2 dpi samples were
228 used as exclusion criteria for biological replicates. In total, of the 180 unique combinations of host
229 species and virus measured, 3 biological replicates were obtained for 161 combinations, 2 replicates
230 for 18 combinations, and 1 replicate for 1 combination (*Drosophila virilis,* CrPV).

231

232 **Statistical Analysis**

233 Phylogenetic generalised linear mixed models were used to investigate the effects of host and virus
234 relatedness on viral load. Multivariate models were fitted using the R package MCMCglmm [72] with
235 the viral load of each of the four virus isolates as the response variable. The structures of the models
236 were as follows:

237 $$y_{hiv} = \beta_{1:v} + \mu_{p:hv} + \mu_{s:hv} + e_{hiv} \quad (1)$$

238 $$y_{hiv} = \beta_{1:v} + \mu_{p:hv} + e_{hiv} \quad (2)$$

239 In these models, y_{hiv} is the change in viral load for virus *v* in the ith biological replicate of host species
240 *h*. The fixed effect β_1 represents the intercepts for each virus isolate, the random effect μ_p represents
241 the effects of the host phylogeny assuming a Brownian motion model of evolution, and *e* represents
242 the model residuals. Model (1) also includes a random species-specific effect that is independent of
243 the host phylogeny ($\mu_{s:hv}$). This explicitly estimates the non-phylogenetic component of between-
244 species variance and allows the proportion of variance explained by the host phylogeny to be
245 calculated. $\mu_{s:hv}$ was removed from model (2) as model (1) struggled to separate the phylogenetic
246 and species-specific traits. Wing size, measured as the length of the IV longitudinal vein from the tip
247 of the proximal segment to the join of the distal segment with vein V [73], provided a proxy for body
248 size [74] and was included in a further model as a fixed effect ($wingsize_h.\beta_2$).

249

250 Within each of these models the random effects and residuals were assumed to follow a
251 multivariate normal distribution with a centred mean of zero and a covariance structure of $V_p \otimes A$ for
252 the phylogenetic effects, $V_s \otimes I$ for species-specific effects, and $V_e \otimes I$ residuals, where $\otimes$ represents
253 the Kronecker product. *A* represents the host phylogenetic relatedness matrix, *I* an identity matrix
254 and *V* represents 4x4 covariance matrices describing the between-species variances and covariances
255 of changes in viral load for the different viruses. Specifically, the matrices V_p and V_s describe the
256 phylogenetic and non-phylogenetic between-species variances in viral load for each virus and the

covariances between them, while the residual covariance matrix V_e describes within-species variance that includes both true within-species effects and measurement errors. Since each biological replicate was tested with a single virus isolate, the covariances of V_e cannot be estimated and were set to zero.

Models were run for 13 million MCMC generations, sampled every 5,000 iterations with a burn-in of 3 million generations. Parameter expanded priors were placed on the covariance matrices, resulting in multivariate F distributions with marginal variance distributions scaled by 1,000. Inverse-gamma priors were placed on the residual variances, with a shape and scale equal to 0.002. To ensure the model outputs were robust to changes in prior distribution, models were also fitted with flat and inverse-Wishart priors, which gave qualitatively similar results.

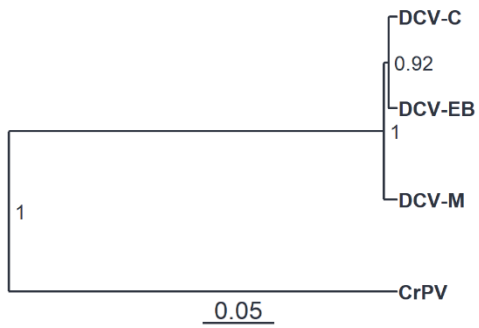
The proportion of the between species variance that can be explained by the phylogeny was calculated from model (1) using the equation $v_p /(v_p + v_s)$, where v_p and v_s represent the phylogenetic and species-specific components of between-species variance [75], and is equivalent to phylogenetic heritability or Pagel's lambda [76,77]. The repeatability of viral load measurements was calculated from model (2) as $v_p/(v_p + v_e)$, where V_e is the residual variance of the model [78]. Inter-specific correlations in viral load were calculated from model (2) V_p matrix as $cov_{x,y}/\sqrt{(var_x + var_y)}$. Parameter estimates reported are means of the posterior density, and 95% credible intervals (95% CI) were taken to be the 95% highest posterior density intervals.

The data files and R scripts used in this study are available in an online repository: https://doi.org/10.6084/m9.figshare.13750711.v1.

Results

Change in viral load is a repeatable trait among host species

282 To investigate how patterns of host susceptibility differ between related viruses, we experimentally
283 infected 45 species of *Drosophila* with four virus isolates: DCV-C, DCV-EB, DCV-M, and CrPV. The DCV
284 isolates were all highly similar (>93% genome and ORF amino acid identity), with the closest
285 relationship between DCV-C and DCV-EB. CrPV formed an outgroup to the DCV isolates (57-59%
286 identity, Fig. 1, Table 1).
287
288 In total, 15,657 flies were inoculated, and the change in viral load after two days of infection was
289 determined by qRT-PCR (Fig. 2). The mean viral load within host species ranged from an
290 approximately 2.7-billion-fold increase in *D. persimilis* infected with DCV-M to a 2.5-fold decrease in
291 *Z. tuberculatus* infected with DCV-C. Viral loads across host species tended to be higher for the DCV
292 isolates, with a mean fold-increase of roughly 11,000 - 19,000, and lower for CrPV, with a mean fold-
293 increase of roughly 1,600.
294



295
296 **Figure 1: Phylogeny of virus isolates.** Evolutionary relationships estimated from open reading frame (ORF)
297 amino acid sequences presented in a midpoint-rooted tree. Node labels represent the posterior probabilities
298 of each clade, and the scale bar represents amino acid substitutions per site.

299

	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%	

300 **Table 1: Virus isolate sequence similarity.** Percentage sequence identity was calculated from multiple-
301 alignment of whole genome nucleotides (white) or concatenated amino acid sequences of ORFs 1 & 2 (grey).
302

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

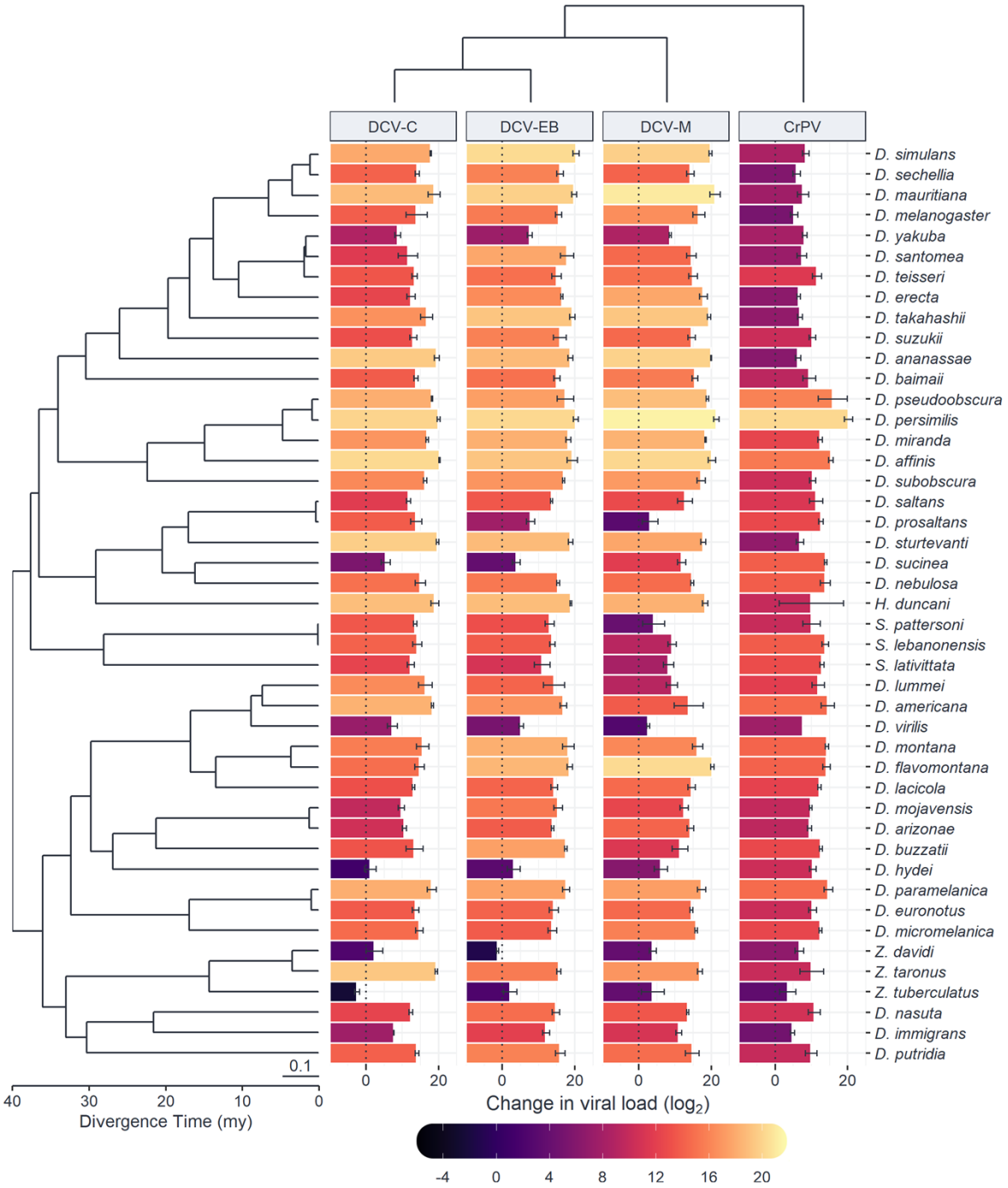


Figure 2: Change in viral load across a diverse panel of *Drosophila* host species for different virus isolates. Bar height and colour show the mean change in RNA viral load by 2 dpi on a log2 scale, with error bars representing the standard error of the mean. The phylogeny of *Drosophila* hosts is presented on the left. The scale bar represents number of nucleotide substitutions per site and the scale axis represents the approximate age since divergence in millions of years (my) based on estimates from [79,80]. The virus cladogram is based on the evolutionary relationships shown in Figure 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 deteriorate with evolutionary distance

We next estimated the inter-specific correlations in patterns of susceptibility between the different viruses (Fig. 3). 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 ($r = 0.52$ to 0.59) but weaker than the correlations between the DCV isolates. 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), reflecting a deterioration in correlation strength with evolutionary distance. 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). Together, this suggests that correlations in patterns of susceptibility across hosts deteriorate with virus genome divergence, although differences between the closely related DCV isolates were not always detectable.

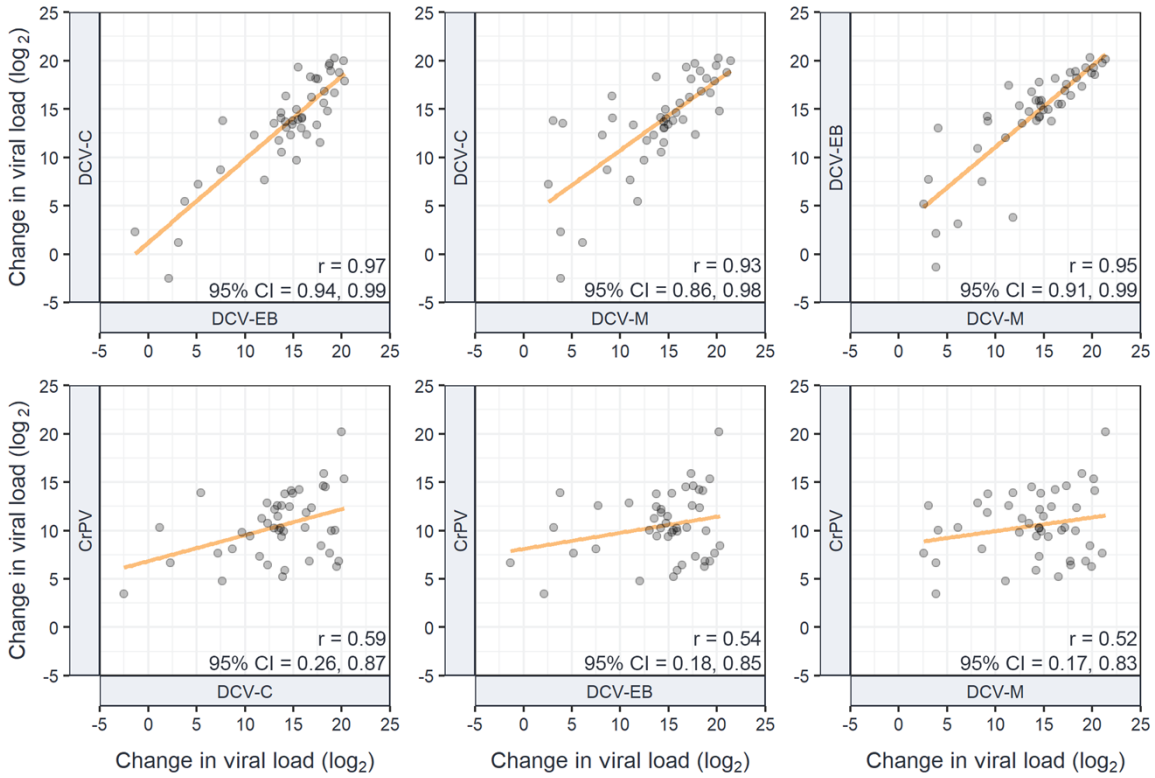


Fig 3: Correlation in viral load between virus isolates. Individual points represent the mean change in viral load after 2 dpi for each host species on a log2 scale, and trendlines have been added from a univariate least-squares linear model for illustrative purposes. Correlations (r) are the total inter-specific correlations and 95% CIs from the output of model (2).

Discussion

Closely related host species present similar environments to novel viruses [75,81], and so often share similar levels of susceptibility to a given virus [14–20]. Likewise, closely related viruses are often assumed to share characteristics that make their host interactions, transmission, and evolutionary trajectories comparable [35–42]. Here, we measured the strength of correlations in the outcomes of infection with four Cripavirus isolates across 45 host species of *Drosophilidae*, to investigate whether the susceptibility of novel hosts is similar for closely related viruses. We found positive correlations between every pair of viruses tested, and the strength of correlation tended to decrease with greater evolutionary distance between viruses. Between DCV isolates, differences in correlation strength were detected but not always well supported. However, DCV isolates were consistently more strongly correlated to each other than to the more distantly related CrPV. This

359 suggests that between virus variation in the outcome of host shifts may be explained by the
360 evolutionary relatedness of the viruses.
361
362 The strong positive correlations between DCV isolates are likely due to high levels of sequence
363 conservation resulting in only small differences in their ability to infect different host species.
364 However, in other instances a small number of mutations have been shown to allow some viruses to
365 successfully infect a novel host [82–85]. We find some evidence of such effects here. For example, in
366 *Zaprionus davidi* DCV-EB shows a decline in viral load, suggesting it is failing to replicate and persist
367 in this host species, whereas the other DCV isolates and CrPV show an increase in viral load in the
368 same host. The weaker correlations between DCV and CrPV may be due to interactions with
369 different host traits that vary in their patterns across the host phylogeny. CrPV and DCV are known
370 to have distinct methods of suppression of the host antiviral RNAi pathway [60,61] and cause
371 pathology in different tissues [62]. Additionally, their relatively high levels of sequence divergence
372 (57-59% identity) may have resulted in changes in the ability of each virus to bind to host cell
373 receptors, utilise host replication machinery, or subvert host immune defences [86].
374
375 Our findings suggest a general rule of thumb: that the ability of a virus to infect a novel host can be
376 inferred based on its relatedness to other viruses. Many examples exist of one or a few virus
377 mutations causing large phenotypic changes that would be exceptions to this rule [87–90]. We do
378 not know the rate at which exceptions like these occur in natural systems and predicting them would
379 require a mechanistic understanding of the underlying host-virus interactions. Our data also suggest
380 that host susceptibility is not specific to individual viruses and that certain host clades may be
381 broadly susceptible to infection. Susceptible host clades may act as reservoirs and donors of
382 emerging infectious diseases and identifying them in relevant systems could inform control and
383 prevention strategies [91]. Further work is now needed to expand these findings to broader groups

384 of viruses, to provide a true comparative framework for understanding virus traits, and to test the
385 importance of the virus phylogeny in understanding the potential outcomes of virus host shifts.
386

387 ## Acknowledgements

388 We would like to thank Julien Martinez, Valérie Dorey, and Maria Carla Saleh for kindly providing us
389 with DCV and CrPV virus isolates, and Jarrod Hadfield for his advice on MCMCglmm model fitting.
390 Ben Longdon and Katherine E. Roberts are supported by a Sir Henry Dale Fellowship jointly funded
391 by the Wellcome Trust and the Royal Society (Grant Number 109356/Z/15/Z). Ryan M. Imrie is
392 supported by a studentship funded by the Natural Environment Research Council (NERC) GW4+
393 Doctoral Training Partnership and the College of Life and Environmental Sciences, University of
394 Exeter.
395

396 ## References

397

398 1. Woolhouse MEJ, Gowtage-Sequeria S. Host Range and Emerging and Reemerging Pathogens. Emerg Infect Dis.
399 2005;11(12):1842–7. doi: 10.3201/eid1112.050997

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

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

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

407 5. Miller IF, Becker AD, Grenfell BT, Metcalf CJE. Disease and healthcare burden of COVID-19 in the United States. Nat Med.
408 2020;26(8):1212–7. doi: 10.1038/s41591-020-0952-y

409 6. Frank TD, Carter A, Jahagirdar D, Biehl MH, Douwes-Schultz D, Larson SL et al. Global, regional, and national incidence,
410 prevalence, and mortality of HIV, 1980–2017, and forecasts to 2030, for 195 countries and territories: a systematic analysis
411 for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. Lancet Hiv. 2019;6(12):e831–59. doi:
412 10.1016/s2352-3018(19)30196-1

413 7. Huber C, Finelli L, Stevens W. The Economic and Social Burden of the 2014 Ebola Outbreak in West Africa. J Infect Dis.
414 2018;218(Supplement_5):S698–704. doi: 10.1093/infdis/jiy213

415 8. Holmes EC. What can we predict about viral evolution and emergence? Curr Opin Virol. 2013;3(2):180–4. doi:
416 10.1016/j.coviro.2012.12.003

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

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

421 11. Holmes EC, Rambaut A, Andersen KG. Pandemics: spend on surveillance, not prediction. Nature. 2018;558(7709):180–
422 2. doi: 10.1038/d41586-018-05373-w

423 12. Levinson J, Bogich TL, Olival KJ, Epstein JH, Johnson CK, Karesh W et al. Targeting Surveillance for Zoonotic Virus
424 Discovery - Volume 19, Number 5—May 2013 - Emerging Infectious Diseases journal - CDC. Emerg Infect Dis.
425 2013;19(5):743–7. doi: 10.3201/eid1905.121042

426 13. Geoghegan JL, Holmes EC. Predicting virus emergence amid evolutionary noise. Open Biol. 2017;7(10):170189. doi:
427 10.1098/rsob.170189

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

430 15. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host Phylogeny Constrains Cross-
431 Species Emergence and Establishment of Rabies Virus in Bats. Science. 2010;329(5992):676–9. doi:
432 10.1126/science.1188836

433 16. Guth S, Visher E, Boots M, Brook CE. Host phylogenetic distance drives trends in virus virulence and transmissibility
434 across the animal–human interface. Philosophical Transactions Royal Soc B Biological Sci. 2019;374(1782):20190296. doi:
435 10.1098/rstb.2019.0296

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

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

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

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

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

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

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

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

452 25. Lambrechts L, Fellous S, Koella JC. Coevolutionary interactions between host and parasite genotypes. Trends Parasitol.
453 2006;22(1):12–6. doi: 10.1016/j.pt.2005.11.008

454 26. Carpenter JA, Hadfield JD, Bangham J, Jiggins FM. Specific interactions between host and parasite genotypes do not act
455 as a constraint on the evolution of antiviral resistance in Drosophila. Evolution. 2012;66(4):1114–25. doi: 10.1111/j.1558-
456 5646.2011.01501.x

457 27. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The Evolution and Genetics of Virus Host Shifts. Plos
458 Pathog. 2014;10(11):e1004395. doi: 10.1371/journal.ppat.1004395

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

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

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

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

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

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

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

473 35. 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
474 severe acute respiratory infections caused by three highly pathogenic human coronaviruses. Respir Res. 2020;21(1):224.
475 doi: 10.1186/s12931-020-01479-w

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

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

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

482 39. Gomes MFC, Piontti AP y, Rossi L, Chao D, Longini I, Halloran ME et al. Assessing the International Spreading Risk
483 Associated with the 2014 West African Ebola Outbreak. Plos Curr Outbreaks. 2014;6. doi:
484 10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5

485 40. Du X, King AA, Woods RJ, Pascual M. Evolution-informed forecasting of seasonal influenza A (H3N2). Sci Transl Med.
486 2017;9(413):eaan5325. doi: 10.1126/scitranslmed.aan5325

487 41. Ruiz SI, Zumbrun EE, Nalca A. Animal Models for the Study of Human Disease (Second Edition). Part K Viral Dis.
488 2017;853–901. doi: 10.1016/b978-0-12-809468-6.00033-4

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

491 43. Simmonds P, Aiewsakun P. Virus classification – where do you draw the line? Arch Virol. 2018;163(8):2037–46. doi:
492 10.1007/s00705-018-3938-z

493 44. Troupin C, Dacheux L, Tanguy M, Sabeta C, Blanc H, Bouchier C et al. Large-Scale Phylogenomic Analysis Reveals the
494 Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. Plos Pathog. 2016;12(12):e1006041. doi:
495 10.1371/journal.ppat.1006041

496 45. Tan Y, Guan W, Lam TT-Y, Pan S, Wu S, Zhan Y et al. Differing Epidemiological Dynamics of Influenza B Virus Lineages in
497 Guangzhou, Southern China, 2009-2010. J Virol. 2013;87(22):12447–56. doi: 10.1128/jvi.01039-13

498 46. Vijaykrishna D, Holmes EC, Joseph U, Fourment M, Su YC, Halpin R et al. The contrasting phylodynamics of human
499 influenza B viruses. Elife. 2015;4:e05055. doi: 10.7554/elife.05055

500 47. Chen R, Holmes EC. The Evolutionary Dynamics of Human Influenza B Virus. J Mol Evol. 2008;66(6):655. doi:
501 10.1007/s00239-008-9119-z

502 48. Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K et al. The evolutionary history of vertebrate RNA viruses. Nature.
503 2018;556(7700):197–202. doi: 10.1038/s41586-018-0012-7

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

506 50. Cherry S, Silverman N. Host-pathogen interactions in drosophila: new tricks from an old friend. Nat Immunol.
507 2006;7(9):911–7. doi: 10.1038/ni1388

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

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

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

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

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

517 56. Magwire MM, Fabian DK, Schweyen H, Cao C, Longdon B, Bayer F et al. Genome-Wide Association Studies Reveal a
518 Simple Genetic Basis of Resistance to Naturally Coevolving Viruses in Drosophila melanogaster. Plos Genet.
519 2012;8(11):e1003057. doi: 10.1371/journal.pgen.1003057

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

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

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

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

528 61. Rij RP van, Saleh M-C, Berry B, Foo C, Houk A, Antoniewski C et al. The RNA silencing endonuclease Argonaute 2
529 mediates specific antiviral immunity in Drosophila melanogaster. Gene Dev. 2006;20(21):2985–95. doi:
530 10.1101/gad.1482006

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

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

535 64. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol.
536 2012;29(8):1969–73. doi: 10.1093/molbev/mss075

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

539 66. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinform. 2020;69(1):e96. doi:
540 10.1002/cpbi.96

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

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

544 69. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc National Acad Sci.
545 1992;89(22):10915–9. doi: 10.1073/pnas.89.22.10915

546 70. Ruijter JM, Thygesen HH, Schoneveld OJ, Das AT, Berkhout B, Lamers WH. Factor correction as a tool to eliminate
547 between-session variation in replicate experiments: application to molecular biology and retrovirology. Retrovirology.
548 2006;3(1):2. doi: 10.1186/1742-4690-3-2

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

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

553 73. Gilchrist GW, Huey RB, Serra L. Rapid evolution of wing size clines in Drosophila subobscura. Genetica. 2001;112–
554 113(1):273–86. doi: 10.1023/a:1013358931816

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

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

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

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

561 78. Falconer D. Introduction to quantitative genetics. 4th ed. Pearson Education India; 1996.

562 79. Russo CA, Takezaki N, Nei M. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol.
563 1995;12(3):391–404. doi: 10.1093/oxfordjournals.molbev.a040214

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

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

568 82. Boon M, Holtappels D, Lood C, Noort V van, Lavigne R. Host Range Expansion of Pseudomonas Virus LUZ7 Is Driven by a
569 Conserved Tail Fiber Mutation. Phage. 2020;1(2):87–90. doi: 10.1089/phage.2020.0006

570 83. Lounková A, Kosla J, Přikryl D, Štafl K, Kučerová D, Svoboda J. Retroviral host range extension is coupled with Env-
571 activating mutations resulting in receptor-independent entry. Proc National Acad Sci. 2017;114(26):E5148–57. doi:
572 10.1073/pnas.1704750114

573 84. Allison AB, Organtini LJ, Zhang S, Hafenstein SL, Holmes EC, Parrish CR. Single Mutations in the VP2 300 Loop Region of
574 the Three-Fold Spike of the Carnivore Parvovirus Capsid Can Determine Host Range. J Virol. 2016;90(2):753–67. doi:
575 10.1128/jvi.02636-15

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

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

580 87. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature.
581 2020;1–6. doi: 10.1038/s41586-020-2895-3

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

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

586 90. Shan C, Xia H, Haller SL, Azar SR, Liu Y, Liu J et al. A Zika virus envelope mutation preceding the 2015 epidemic enhances
587 virulence and fitness for transmission. Proc National Acad Sci. 2020;117(33):20190–7. doi: 10.1073/pnas.2005722117

588 91. Streicker DG, Gilbert AT. Contextualizing bats as viral reservoirs. Science. 2020;370(6513):172–3. doi:
589 10.1126/science.abd4559

590

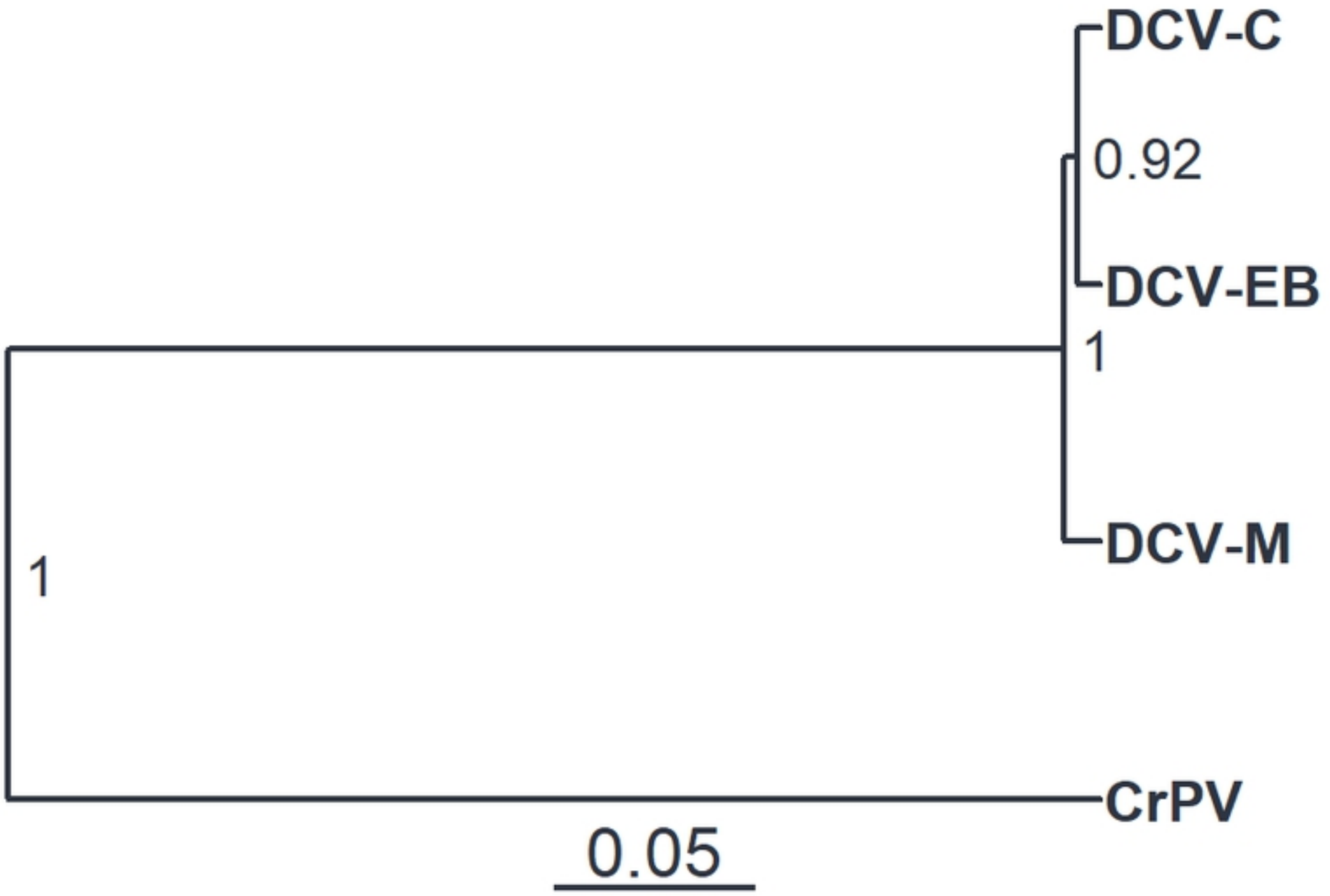


Figure 1

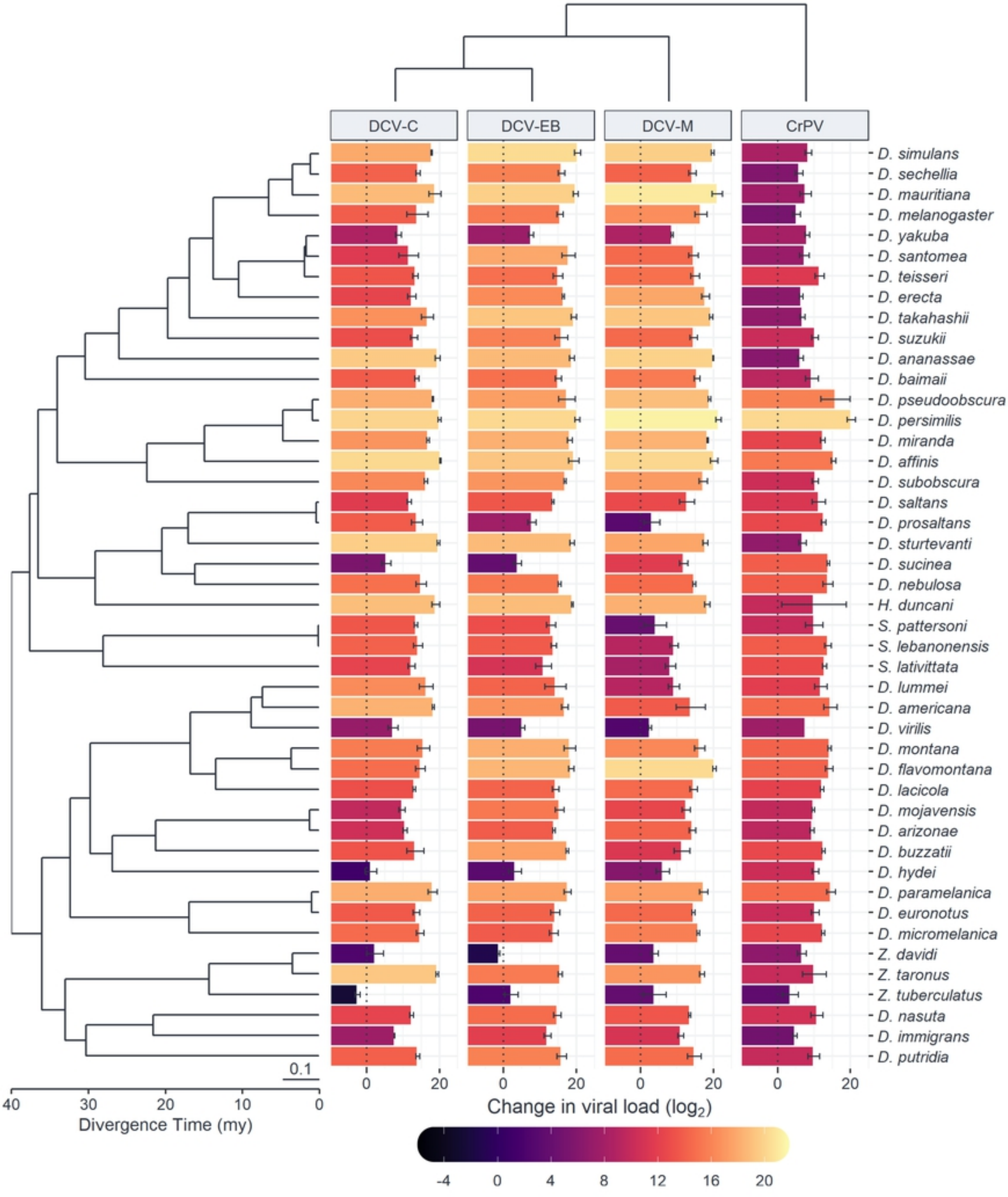

DCV-C
DCV-EB
DCV-M
CrPV
D. simulans
D. sechellia
D. mauritiana
D. melanogaster
D. yakuba
D. santomea
D. teisseri
D. erecta
D. takahashii
D. suzukii
D. ananassae
D. baimaii
D. pseudoobscura
D. persimilis
D. miranda
D. affinis
D. subobscura
D. saltans
D. prosaltans
D. sturtevanti
D. sucinea
D. nebulosa
H. duncani
S. pattersoni
S. lebanonensis
S. lativittata
D. lummei
D. americana
D. virilis
D. montana
D. flavomontana
D. lacicola
D. mojavensis
D. arizonae
D. buzzatii
D. hydei
D. paramelanica
D. euronotus
D. micromelanica
Z. davidi
Z. taronus
Z. tuberculatus
D. nasuta
D. immigrans
D. putridia
0.1
40
30
20
10
0
Divergence Time (my)
0
20
Change in viral load (log2)
-4
0
4
8
12
16
20

Figure 2

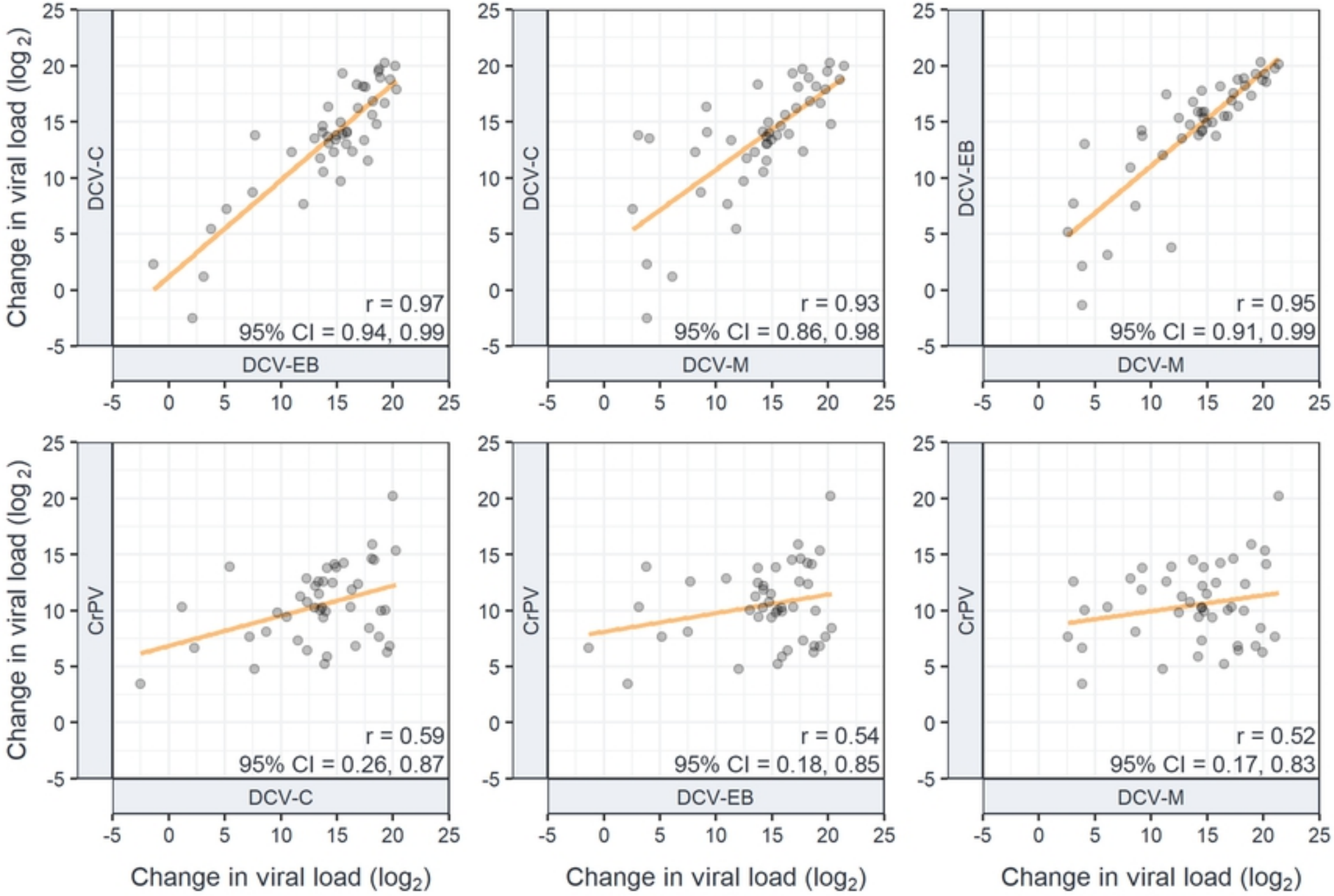

DCV-C
DCV-EB
r = 0.97
95% CI = 0.94, 0.99
DCV-C
DCV-M
r = 0.93
95% CI = 0.86, 0.98
DCV-EB
DCV-M
r = 0.95
95% CI = 0.91, 0.99
CrPV
DCV-C
r = 0.59
95% CI = 0.26, 0.87
CrPV
DCV-EB
r = 0.54
95% CI = 0.18, 0.85
CrPV
DCV-M
r = 0.52
95% CI = 0.17, 0.83
Change in viral load (log 2)
Change in viral load (log2)

Figure 3