Short report

Host shifts result in parallel genetic changes when viruses adapt to closely related species

Ben Longdon[*1,2], Jonathan P Day[2], Joel M Alves[2,3], Sophia CL Smith[2], Thomas M Houslay[1], John E McGonigle[2], Lucia Tagliaferri[2] and Francis M Jiggins[2]

[1] Biosciences, College of Life & Environmental Sciences, University of Exeter, Penryn Campus, TR10 9FE, UK
[2] Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK
[3] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal

*b.longdon2@exeter.ac.uk

Abstract

Host shifts, where a pathogen jumps into a new host species, are a major source of emerging infectious diseases. They frequently occur between related host species and often rely on the pathogen evolving adaptations that increase their fitness in the novel host species. To investigate genetic adaptation to novel hosts, we experimentally evolved replicate lineages of an RNA virus (Drosophila C Virus) in 19 different species of Drosophilidae and deep sequenced the viral genomes. We found a strong pattern of parallel evolution, where viral lineages from the same host were genetically more similar to each other than to lineages from other host species, indicating species-specific adaptation. When we compared viruses that had adapted to different host species, we found that parallel genetic changes were more likely to occur if the two host species were closely related. This suggests that when a virus adapts to one host it might also become better adapted to closely related host species. This may explain in part why host shifts tend to occur between related species, and may mean that when a new pathogen appears in a given species, closely related species may become vulnerable to the new disease.

Introduction

Host shifts – where a pathogen jumps into a new host species – are a major source of emerging infectious diseases. RNA viruses seem particularly prone to host shift [1-4], with HIV, Ebola virus and SARS coronavirus all having been acquired by humans from other host species [5-7]. Whilst some pathogens may be pre-adapted to a novel host, there is increasing evidence that adaptation to the new host is frequently essential for a host shift to succeed [8]. These adaptations may allow a pathogen to enter host cells, increase replication rates, avoid or suppress the host immune response, or optimise virulence or transmission [9, 10]. For example, in the 2013-2016 Ebola virus epidemic in West Africa, a mutation in the viral glycoprotein gene that arose early in the outbreak and rose to high frequency was found to increase infectivity in human cells, and decrease infectivity in bats, which are thought to be the source of Ebola virus [11, 12]. Likewise, a switch of a parvovirus from cats to dogs resulted in mutations in the virus capsid that allowed the virus to bind to cell receptors in dogs, but resulted in the virus losing its ability to infect cats [13].

In some instances adaptation to a novel host relies on specific mutations that arise repeatedly whenever a pathogen switches to a given host. For example, in the jump of HIV-1 from chimps to humans, codon 30 of the *gag* gene has undergone a change that increases virus replication in humans, and this has occurred independently in all three HIV-1 lineages [5, 14]. Similarly, five parallel mutations have been observed in the two independent epidemics of SARS coronavirus following its jump from palm civets into humans [15]. Similar patterns have been seen in experimental evolution studies, where parallel genetic changes occur repeatedly when replicate viral lineages adapt to a new host species in the lab. For example, when Vesicular Stomatitis Virus was passaged in human or dog cells, the virus evolved parallel mutations when evolved on the same cell type [16]. Likewise, a study passaging Tobacco Etch Potyvirus on four plant species found parallel mutations occurred only when the virus infected the same host species [17]. These parallel mutations provide compelling evidence that these genetic changes are adaptive, with the same mutations evolving independently in response to natural selection [18].

The host phylogeny is important for determining a pathogens ability to infect a novel host, with pathogens tending to replicate most efficiently when they infect a novel host that is closely related to their original host. This pattern is seen in nature, where host shifts tend to occur most frequently between closely related hosts, and in laboratory cross-infection studies, where viruses tend to replicate more rapidly when the new host is related to the pathogens natural host [2, 19-32].

These phylogenetic patterns suggest that mutations that adapt a virus to one host may also adapt it to closely related host species. Therefore, there may be a knock-on effect, where a host shift leaves closely related species vulnerable to the new disease. To investigate this, we asked whether viruses acquire the same genetic adaptations when adapting to closely related host species. We experimentally evolved replicate lineages of an RNA virus called Drosophila C Virus (DCV; Discistroviridae) in 19 species of Drosophilidae that vary in their relatedness and shared a common ancestor approximately 40 million years ago [33, 34]. We then sequenced the genomes of the evolved viral lineages and tested whether the same genetic changes arose when the virus was evolved in closely related host species.

Results

Parallel genetic changes occur in DCV lineages that have evolved in the same host species

To examine how viruses adapt to different host species we experimentally evolved DCV in 19 species of Drosophilidae by serially passaging the virus. In total we infected 22,095 adult flies and generated 173 independent replicate lineages (6-10 per host species). We deep sequenced the evolved virus genomes to generate over 740,000 300bp sequence reads from each viral lineage. Out of 8989 sites, 584 contained a SNP with a derived allele frequency >0.05 in at least one viral lineage, and 84 of these were tri-allelic. None of these variants were found at an appreciable frequency in five sequencing libraries produced from the ancestral virus, indicating that they had spread though populations during the experiment (Figure 1). In multiple cases these variants had nearly reached fixation (Figure 1).

A powerful way to detect adaptation is to test whether the same genetic changes occur in parallel when different populations encounter the same environment. Of the 584 SNPs, 102 had derived allele frequencies >0.05 in at least two viral lineages, and some had risen to high frequencies in multiple lineages (Figure 1). We estimated the genetic differentiation between viral lineages by calculating F_{ST}. We found that viral lineages that had evolved within the same host were genetically more similar to each other than to lineages from other host species (Figure 2; $P<0.001$). The observation of parallel evolution within the same host species indicates that the viruses are genetically adapting to the different host species.

To examine the genetic basis of parallel evolution, we individually tested whether each SNP in the DCV genome showed a signature of parallel evolution among viral lineages passaged in the same host species (i.e. we repeated the analysis in Figures 2 for each SNP). We identified 56 SNPs with a significant signal of parallel evolution within the same host species ($P<0.05$; significantly parallel SNPs are shown with a red asterisk in Figure 1; the false discovery rate is estimated to be 17% [35]).

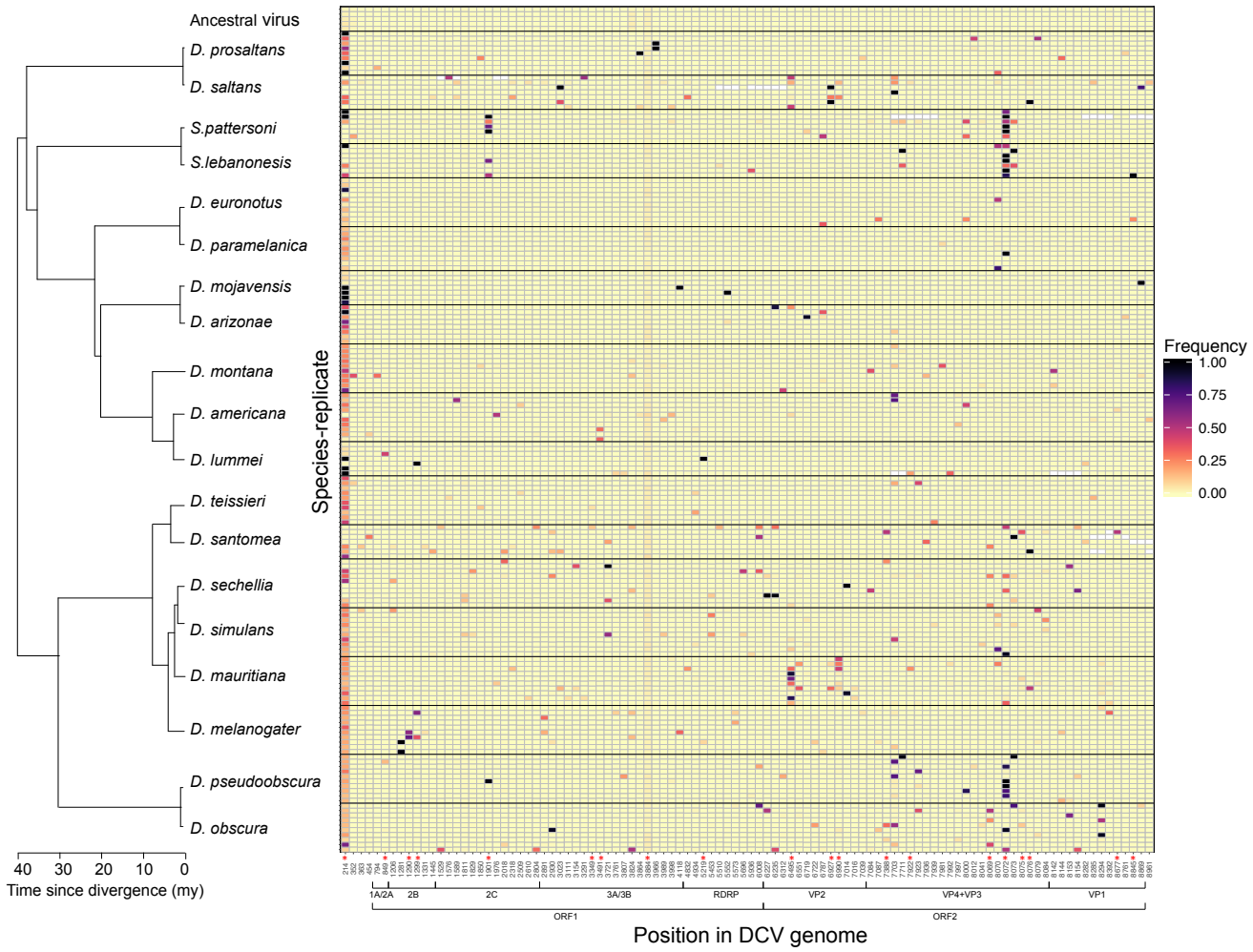

Figure 1. The frequency of SNPs in viral lineages that have evolved in different host species. Each row represents an independent viral lineage. Viruses that evolved in different host species are separated by black horizontal lines. Each column represents a polymorphic site in the DCV genome, and only sites where the derived allele frequency >0.05 in at least two lineages are shown. The intensity of shading represents the derived allele frequency. Sites where there are three alleles have the two derived allele frequencies pooled for illustrative purposes. SNPs whose frequencies are significantly correlated among lineages from the same host species are shown by red stars at the bottom the column (permutation test; $p<0.05$). Open reading frames (ORFs) and viral proteins based on predicted polyprotein cleavage sites ([36-39] UniProtKB http://www.uniprot.org/uniprot/O36966) are below the x axis. Sites with missing data are shown in white. The phylogeny was inferred under a relaxed molecular clock [31, 40] and the scale axis represents the approximate age since divergence in millions of years (my) based on estimates from: [33, 34].

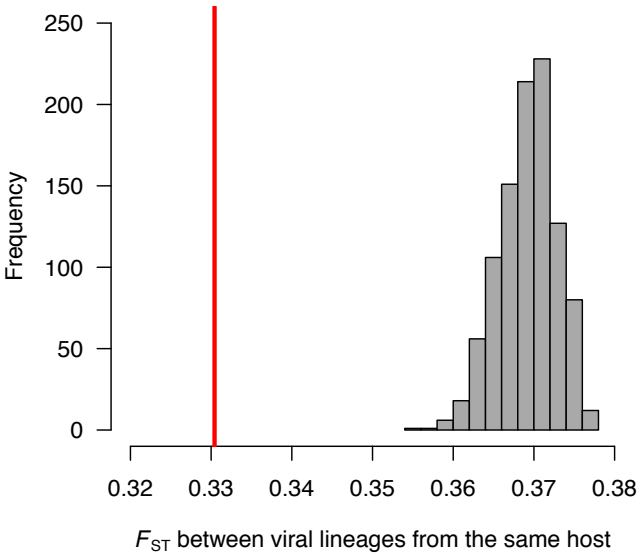

Figure 2. Viral lineages from the same host species were genetically more similar to each other than to lineages from different host species. The mean pairwise F_{ST} between all possible pairs of viral lineages from the same host species was calculated. The red line shows the observed value. The grey bars are the null distribution of this statistic obtained by permuting the viral lineages across host species 1000 times.

Viruses in closely related hosts are genetically more similar

We investigated if viruses passaged through closely related hosts showed evidence of parallel molecular adaptations. We calculated F_{ST} between all possible pairs of viral lineages that had evolved in different host species. We found that viral lineages from closely related hosts were more similar to each other than viral lineages from more distantly related hosts (Figure 3A). This is reflected in a significant positive relationship between virus F_{ST} and host genetic distance (Figure 3B, Permutation test: r=0.15, P=0.002). We lacked the statistical power to identify the specific SNPs that are causing the signature of parallel evolution in Figure 3 (false discovery rate >0.49 for all SNPs).

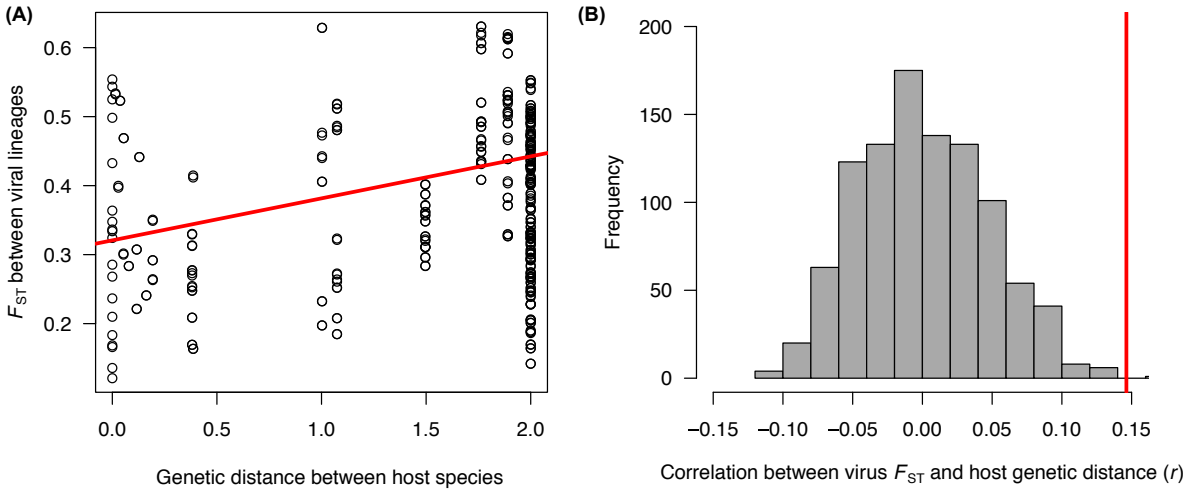

Figure 3. Viral lineages from more closely related host species are genetically more similar. (A) The correlation between the genetic differentiation of viral lineages and the genetic distance between the species they have evolved in. Linear regression line is shown in red. This plot includes comparisons between viral lineages that evolved in the same host species (genetic distance = 0). Genetic distances were scaled so that the distance from the root to the tip of the tree was one. (B) Pearson's correlation coefficient (*r*) of F_{ST} between pairs of viral lineage and the genetic distance between the host species they evolved in. The observed value is in red and the grey bars are the null distribution obtained by permutation. This analysis did not include comparisons between viral lineages from the same host species

Discussion

When a pathogen infects a novel host species, it finds itself in a new environment to which it must adapt [4, 9, 41, 42]. When DCV was passaged through different species of Drosophilidae, we found the same genetic changes arose repeatedly in replicate viral lineages in the same host species. Such repeatable parallel genetic changes to the same host environment are compelling evidence that these changes are adaptive [18]. We then examined whether these same genetic changes might adapt the virus to closely related host species, as these are likely to present a similar environment for the virus. We found that viruses evolved in closely related hosts were more similar to each other than viruses that evolved in more distantly related species. Therefore, mutations that are adaptive in one host species frequently arise when the virus infects closely related hosts. This finding of parallel genetic changes in closely related host species suggests that when a virus adapts to one host it might also become better adapted to closely related host species.

Such phylogenetic patterns of host adaptation may in part explain why pathogens tend to be more likely to jump between closely related host species. At a phenotypic level, similar patterns have been reported in nature [19, 27]. Likewise, in a large cross-infection experiment Drosophila sigma viruses (Rhabdoviridae) isolated from different species of *Drosophila*, the viruses tended to replicate most efficiently in species closely related to their natural hosts [32]. This suggests that these viruses had acquired adaptations to their host species that benefitted them when they infected closely related species. Our results demonstrate that this pattern is apparent at the level of specific nucleotides, and can arise very shortly after a host shift.

While the susceptibility of a novel host is strongly correlated to its relatedness to the pathogens' original host, it is also common to find exceptions to this pattern. This is seen both in nature when pathogens shift between very distant hosts [43, 44], and in laboratory cross-infection experiments [31, 32]. This pattern is also seen in our data where we also observe parallel genetic changes occurring between more distantly related hosts. For example, a mutation at position 8072 was not only near fixation in most of the lineages infecting two closely related species, but also occurred at a high frequency in replicate lineages in a phylogenetically distant host (Figure 1).

In conclusion, we have found that host relatedness can be important in determining how viruses evolve when they find themselves in a new host. This study suggests that while some genetic changes will be found only in specific hosts, we frequently see the same changes occurring in closely

199 related host species. This could mean that mutations that adapt a virus to one host might also adapt
200 it to closely related species.
201
202
203 **Methods**
204
205 *Virus production*
206
207 DCV is a positive sense RNA virus in the family Discistroviridae that was isolated from *D.*
208 *melanogaster*, which it naturally infects in the wild [45, 46]. To minimise the amount of genetic
209 variation in the DCV isolate we used to initiate the experimental evolution study, we aimed to isolate
210 single infectious clones of DCV using a serial dilution procedure. DCV was produced in Schneider's
211 Drosophila line 2 (DL2) cells [47] as described in [48]. Cells were cultured at 25°C in Schneider's
212 Drosophila Medium with 10% Fetal Bovine Serum, 100 U/ml penicillin and 100 µg/ml streptomycin
213 (all Invitrogen, UK). The DCV strain used was isolated from *D. melanogaster* collected in Charolles,
214 France [49]. DL2 cells were seeded into two 96-well tissue culture plates at approximately 10^4 cells in
215 100 µl of media per well. Cells were allowed to adhere to the plates by incubating at 25°C for five
216 hours or over-night. Serial 1:1 dilutions of DCV were made in complete Schneider's media, giving a
217 range of final dilutions from $1{:}10^8$ – $1{:}4x10^{14}$. 100 µl of these dilutions were then added to the cells
218 and incubated for 7 days, 8 replicates were made for each DCV dilution. Each well was then
219 examined for DCV infection of the DL2 cells, and a well was scored as positive for DCV infection if
220 clear cytopathic effects were present in the majority of the cells. The media was taken from the
221 wells with the greatest dilution factor that were scored as infected with DCV and stored at -80°C.
222 This processes was then repeated using the DCV samples from the first dilution series. One clone,
223 B6A, was selected for amplification and grown in cell culture as described above. Media containing
224 DCV was removed and centrifuged at 3000 x g for 5 minutes at 4°C to pellet any remaining cell
225 debris, before being aliquoted and stored at -80°C. The Tissue Culture Infective Dose 50 ($TCID_{50}$) of
226 the DCV was 6.32 x 10^9 infectious particles per ml using the Reed-Muench end-point method [50].
227
228 *Inoculating fly species*
229
230 We passaged the virus through 19 species of Drosophilidae, with 6-10 independent replicate
231 passages for each species. We selected species from across the phylogeny (that shared a common
232 ancestor approximately 40 million years ago [33, 34]), but included clades of closely related species
233 that recently shared common ancestors less than 5 million years ago (Figure 1). All fly stocks were
234 reared at 22°C. Stocks of each fly species were kept in 250ml bottles at staggered ages. Flies were
235 collected and sexed, and males were placed on cornmeal medium for 4 days before inoculation.
236 Details of the fly stocks used can be found in the supplementary materials.
237
238 4-11 day old males were infected with DCV using a 0.0125 mm diameter stainless steel needle
239 (26002–10, Fine Science Tools, CA, USA) dipped in DCV solution. For the first passage this was the
240 cloned DCV isolate in cell culture supernatant (described above), and then subsequently was the
241 virus extracted from the previous passage (described below). The needle was pricked into the
242 pleural suture on the thorax of flies, towards the midcoxa. Each replicate was infected using a new
243 needle and strict general cleaning procedures were used to minimise any risk of cross-contamination

244 between replicates. Species were collected and inoculated in a randomised order each passage. Flies
245 were then placed into vials of cornmeal medium and kept at 22°C and 70% relative humidity. Flies
246 were snap frozen in liquid nitrogen 3 days post-infection, homogenised in Ringer's solution (2.5µl
247 per fly) and then centrifuged at 12,000g for 10 mins at 4°C. The resulting supernatant was removed
248 and frozen at -80°C to be used for infecting flies in the subsequent passage. The remaining
249 homogenate was preserved in Trizol reagent (Invitrogen) and stored at -80°C for RNA extraction.
250 The 3 day viral incubation period was chosen based on time course and pilot data showing that viral
251 load reaches a maximum at approximately 3 days post-infection. This process was repeated for 10
252 passages for all species, except *D. montana* where only 8 passages were carried out due to the fly
253 stocks failing to reproduce. Each lineage was injected into a mean of 11 flies at each passage (range
254 4-18). Experimental evolution studies in different tissue types have seen clear signals of adaptation
255 in 100 virus generations [16]. Based on $\log_2$ change in RNA viral load we estimate that we have
256 passaged DCV for approximately 100-200 generations.
257
258 *Sequencing*
259
260 After passaging the virus, we sequenced evolved viral lineages from 19 host species, with a mean of
261 9 independent replicate lineages of the virus per species (range 6-10 replicates). cDNA was
262 synthesised using Invitrogen Superscript III reverse-transcriptase with random hexamer primers
263 (25°C 5mins, 50°C 50mins, 70°C 15mins). The genome of the evolved viruses, along with the initial
264 DCV ancestor (x5) were then amplified using Q5 high fidelity polymerase (NEB) in nine overlapping
265 PCR reactions (see supplementary Table S2 for PCR primers and cycle conditions). Primers covered
266 position 62-9050bp (8989bp) of the Genbank refseq (NC_001834.1) giving 97% coverage of the
267 genome. PCRs of individual genomes were pooled and purified with Ampure XP beads (Agencourt).
268 Individual Nextera XT libraries (Illumina) were prepared for each viral lineage. In total we sequenced
269 173 DCV pooled amplicon libraries on an Illumina MiSeq (Cambridge Genomic Service) v3 for 600
270 cycles to give 300bp paired-end reads.
271
272 *Bioinformatics and variant calling*

273 FastQC, version 0.11.2 [51] was used to assess read quality and primer contamination. Trimmomatic,
274 version 0.32 [52] was used to removed low quality bases and adaptor sequences, using the following
275 options: MINLEN=30 (Drop the read if it is below 30 base pairs), TRAILING=15 (cut bases of the end
276 of the read if below a threshold quality of 15), SLIDINGWINDOW=4:20 (perform a sliding window
277 trimming, cutting once the average quality within a 4bp window falls below a threshold of 20), and
278 ILLUMINACLIP=TruSeq3-PE.fa:2:20:10:1:true (remove adapter contamination; the values correspond
279 in order to: input fasta file with adapter sequences to be matched, seed mismatches, palindrome clip
280 threshold, simple clip threshold, minimum adapter length and logical value to keep both reads in
281 case of read-through being detected in paired reads by palindrome mode).

282 To generate a reference ancestral Drosophila C Virus sequence we amplified the ancestral starting
283 virus by PCR as above. PCR products were treated with exonuclease 1 and Antarctic phosphatase to
284 remove unused PCR primers and dNTPs and then sequenced directly using BigDye reagents (ABI) on
285 an ABI 3730 capillary sequencer in both directions (Source Bioscience, Cambridge, UK). Sequences
286 were edited in Sequencher (version 4.8; Gene Codes), and were manually checked for errors. Fastq

reads were independently aligned to this reference sequence (Genbank accession: MG570143) using BWA-MEM, version 0.7.10 {Li, 2009 #1605} with default options with exception of the parameter –M, which marks shorter split hits as secondary. 99.5% of reads had mapping phred quality scores of >60. The generated SAM files were converted to their binary format (BAM) and sorted by their leftmost coordinates with SAMtools, version 0.1.19 (website: http://samtools.sourceforge.net/) [53]. Read Group information (RG) was added to the BAM files using the module AddOrReplaceReadGroups from Picard Tools, version 1.126 (https://broadinstitute.github.io/picard).

The variant calling was then performed for each individual BAM using UnifiedGenotyper tool from GATK, version 3.3.0. As we were interested in calling low frequency variants in our viruses, we assumed a ploidy level of 100 (-sample_ploidy:100). The other parameters were set to their defaults except --stand_call_conf:30 (minimum phred-scaled confidence threshold at which variants should be called) and --downsample_to_coverage:1000 (down-sample each sample to 1000X coverage)

Host phylogeny

We used a trimmed version of an ultrametric (time-based) phylogeny produced previously [31]. This phylogeny was inferred using seven genes with a relaxed molecular clock model in BEAST (v1.8.0) [40, 54]. The tree was pruned to the 19 species used using the Ape package in R [55, 56].

Statistical Analysis

We examined the frequency of alternate alleles (single nucleotide polymorphisms: SNPs) in five ancestral virus replicates (aliquots of the same virus stock that was used to found the evolved lineages). SNPs in these ancestral viruses may represent pre-standing genetic variation, or may be sequencing errors. We found the mean SNP frequency was 0.000923 and the highest frequency of any SNP was 0.043 across the ancestral viruses. We therefore included a SNP in our analyses if its frequency was >0.05 in any of the evolved viral lineages. For all analyses we included all three alleles at triallelic sites.

Parallel evolution within species

As a measure of genetic differentiation we estimated F_{ST} between all the virus lineages based on the heterozygosity (H) of the SNPs we called [57]:

$$F_{ST} = \frac{H_b - H_w}{H_b} \qquad \text{(Equation 1)}$$

where H_b is the mean number of differences between pairs of sequence reads sampled from the two different lineages. H_w is mean number of differences between sequence reads sampled from within each lineage. H_b and H_w were calculated separately for each polymorphic site, and the mean across sites used in equation (1). H_w was calculated separately for the two lineages being compared, and the unweighted mean used in equation (1).

To examine whether there had been parallel evolution among viral lineages that had evolved within the same fly species, we calculated the mean F_{ST} between lineages that had evolved in the same fly species, and compared this to the mean F_{ST} between lineages that had evolved in different fly

species. We tested whether this difference was statistically significant using a permutation test. The fly species labels were randomly reassigned to the viral lineages, and we calculated the mean F_{ST} between lineages that had evolved in the same fly species. This was repeated 1000 times to generate a null distribution of the test statistic, and this was then compared to the observed value.

To identify individual SNPs with a signature of parallel evolution within species, we repeated this procedure separately for each SNP.

Parallel evolution between species

We next examined whether viral lineages that had evolved in different fly species tended to be more similar if the fly species were more closely related. Considering all pairs of viral lineages from different host species, we correlated pairwise F_{ST} with the genetic distance between the fly species. To test the significance of this correlation, we permuted the fly species over the *Drosophila* phylogeny and recalculated the Pearson correlation coefficient. This was repeated 1000 times to generate a null distribution of the test statistic, and this was then compared to the observed value. To identify individual SNPs whose frequencies were correlated with the genetic distance between hosts we repeated this procedure separately for each SNP.

We confirmed there was no relationship between rates of molecular evolution (SNP frequency) and either genetic distance from the host DCV was isolated from (*D. melanogaster*) or estimated viral population size (see supplementary Figures S1 and S2).

Sequence data (fastq files) are available in the NCBI SRA (Accession: SRP119720). BAM files, data and R scripts for analysis are available from the NERC data repository (funding requirement - awaiting doi, temporary link to data and scripts https://figshare.com/s/b119ba86def8bca58782).

Author contributions

Designing experiment: BL, FMJ. Lab work: JPD, SCLS, TMH, LT, BL. Bioinformatics analysis: JMA, JEM, FMJ, BL. Statistical analysis: BL, FMJ. Manuscript written by BL and FMJ with input from all authors.

Funding

BL and FMJ are supported by a NERC grant (NE/L004232/1 http://www.nerc.ac.uk/) and by an ERC grant (281668, DrosophilaInfection, http://erc.europa.eu/). BL is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 109356/Z/15/Z).

References

1. Cleaveland S, Laurenson MK, Taylor LH. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 2001;356(1411):991-9. PubMed PMID: WOS:000170315900003.

2. Davies TJ, Pedersen AB. Phylogeny and geography predict pathogen community similarity in wild primates and humans. Proceedings of the Royal Society B-Biological Sciences. 2008;275(1643):1695-701. doi: 10.1098/rspb.2008.0284. PubMed PMID: ISI:000256387500014.

3. Taylor LH, Latham SM, Woolhouse ME. Risk factors for human disease emergence. Philos Trans R Soc Lond B Biol Sci. 2001;356(1411):983-9. Epub 2001/08/23. doi: 10.1098/rstb.2001.0888. PubMed PMID: 11516376; PubMed Central PMCID: PMC1088493.

4. Woolhouse ME, Haydon DT, Antia R. Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol. 2005;20(5):238-44. Epub 2006/05/17. doi: S0169-5347(05)00038-8 [pii]
10.1016/j.tree.2005.02.009. PubMed PMID: 16701375.

5. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. Philosophical Transactions of the Royal Society B-Biological Sciences. 2010;365(1552):2487-94. doi: 10.1098/rstb.2010.0031. PubMed PMID: WOS:000280097000008.

6. Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, et al. Fruit bats as reservoirs of Ebola virus. Nature. 2005;438(7068):575-6. doi: 10.1038/438575a. PubMed PMID: WOS:000233593100030.

7. Li WD, Shi ZL, Yu M, Ren WZ, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310(5748):676-9. doi: 10.1126/science.1118391. PubMed PMID: WOS:000232997700042.

8. Woolhouse M, Adair K, Brierley L. RNA Viruses: A Case Study of the Biology of Emerging Infectious Diseases. Microbiol Spectr. 2013;1(1). doi: 10.1128/microbiolspec.OH-0001-2012. PubMed PMID: 26184815.

9. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The Evolution and Genetics of Virus Host Shifts. PLoS Pathog. 2014;10(11):e1004395. Epub 2014/11/07. doi: 10.1371/journal.ppat.1004395. PubMed PMID: 25375777; PubMed Central PMCID: PMC4223060.

10. Parrish CR, Kawaoka Y. The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Annual review of microbiology. 2005;59:553-86. Epub 2005/09/13. doi: 10.1146/annurev.micro.59.030804.121059. PubMed PMID: 16153179.

11. Diehl WE, Lin AE, Grubaugh ND, Carvalho LM, Kim K, Kyawe PP, et al. Ebola Virus Glycoprotein with Increased Infectivity Dominated the 2013-2016 Epidemic. Cell. 2016;167(4):1088-98 e6. doi: 10.1016/j.cell.2016.10.014. PubMed PMID: 27814506; PubMed Central PMCID: PMCPMC5115602.

12. Urbanowicz RA, McClure CP, Sakuntabhai A, Sall AA, Kobinger G, Muller MA, et al. Human Adaptation of Ebola Virus during the West African Outbreak. Cell. 2016;167(4):1079-87 e5. doi: 10.1016/j.cell.2016.10.013. PubMed PMID: 27814505; PubMed Central PMCID: PMCPMC5101188.

13. Truyen U, Evermann JF, Vieler E, Parrish CR. Evolution of canine parvovirus involved loss and gain of feline host range. Virology. 1996;215(2):186-9. Epub 1996/01/15. doi: 10.1006/viro.1996.0021. PubMed PMID: 8560765.

14. Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, et al. Adaptation of HIV-1 to its human host. Mol Biol Evol. 2007;24(8):1853-60. Epub 2007/06/05. doi: 10.1093/molbev/msm110. PubMed PMID: 17545188.

15. Liu W, Tang F, Fontanet A, Zhan L, Wang TB, Zhang PH, et al. Molecular epidemiology of SARS-associated coronavirus, Beijing. Emerg Infect Dis. 2005;11(9):1420-4. Epub 2005/10/19. doi: 10.3201/eid1109.040773. PubMed PMID: 16229772; PubMed Central PMCID: PMC3310602.

16. Remold SK, Rambaut A, Turner PE. Evolutionary genomics of host adaptation in vesicular stomatitis virus. Mol Biol Evol. 2008;25(6):1138-47. Epub 2008/03/21. doi: 10.1093/molbev/msn059. PubMed PMID: 18353798.

17. Bedhomme S, Lafforgue G, Elena SF. Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. Mol Biol Evol. 2012;29(5):1481-92. Epub 2012/02/10. doi: 10.1093/molbev/msr314. PubMed PMID: 22319146.

18. Bollback JP, Huelsenbeck JP. Parallel Genetic Evolution Within and Between Bacteriophage Species of Varying Degrees of Divergence. Genetics. 2009;181(1):225-34. doi: 10.1534/genetics.107.085225. PubMed PMID: WOS:000262595500021.
19. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host Phylogeny Constrains Cross-Species Emergence and Establishment of Rabies Virus in Bats. Science. 2010;329(5992):676-9. doi: 10.1126/science.1188836. PubMed PMID: WOS:000280602700037.
20. Faria NR, Suchard MA, Rambaut A, Streicker DG, Lemey P. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. Philos Trans R Soc Lond B Biol Sci. 2013;368(1614):20120196. Epub 2013/02/06. doi: 10.1098/rstb.2012.0196. PubMed PMID: 23382420; PubMed Central PMCID: PMC3678322.
21. Cooper N, Griffin R, Franz M, Omotayo M, Nunn CL, Fryxell J. Phylogenetic host specificity and understanding parasite sharing in primates. Ecol Lett. 2012;15(12):1370-7. Epub 2012/08/24. doi: 10.1111/j.1461-0248.2012.01858.x. PubMed PMID: 22913776.
22. Waxman D, Weinert LA, Welch JJ. Inferring host range dynamics from comparative data: the protozoan parasites of new world monkeys. Am Nat. 2014;184(1):65-74. Epub 2014/06/13. doi: 10.1086/676589. PubMed PMID: 24921601.
23. Huang S, Bininda-Emonds ORP, Stephens PR, Gittleman JL, Altizer S. Phylogenetically related and ecologically similar carnivores harbor similar parasite assemblages. Journal of Animal Ecology. 2013:n/a-n/a. doi: 10.1111/1365-2656.12160.
24. Hadfield JD, Krasnov BR, Poulin R, Nakagawa S. A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions. The American Naturalist. 2014;0(0):000. doi: 10.1086/674445.
25. Ramsden C, Holmes EC, Charleston MA. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol Biol Evol. 2009;26(1):143-53. Epub 2008/10/17. doi: msn234 [pii]
10.1093/molbev/msn234. PubMed PMID: 18922760.
26. de Vienne DM, Hood ME, Giraud T. Phylogenetic determinants of potential host shifts in fungal pathogens. Journal of Evolutionary Biology. 2009;22(12):2532-41. doi: 10.1111/j.1420-9101.2009.01878.x. PubMed PMID: WOS:000271785800019.
27. Gilbert GS, Webb CO. Phylogenetic signal in plant pathogen-host range. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(12):4979-83. doi: 10.1073/pnas.0607968104. PubMed PMID: WOS:000245256700040.
28. Tinsley MC, Majerus MEN. Small steps or giant leaps for male-killers? Phylogenetic constraints to male-killer host shifts. Bmc Evolutionary Biology. 2007;7. doi: 10.1186/1471-2148-7-238. PubMed PMID: WOS:000252786000001.
29. Russell JA, Goldman-Huertas B, Moreau CS, Baldo L, Stahlhut JK, Werren JH, et al. Specialization and geographic isolation among Wolbachia symbionts from ants and lycaenid butterflies. Evolution. 2009;63(3):624-40. Epub 2008/12/05. doi: 10.1111/j.1558-5646.2008.00579.x. PubMed PMID: 19054050.
30. Perlman SJ, Jaenike J. Infection success in novel hosts: An experimental and phylogenetic study of Drosophila-parasitic nematodes. Evolution. 2003;57(3):544-57. PubMed PMID: WOS:000182193800010.
31. Longdon B, Hadfield JD, Day JP, Smith SC, McGonigle JE, Cogni R, et al. The Causes and Consequences of Changes in Virulence following Pathogen Host Shifts. PLoS Pathog. 2015;11(3):e1004728. Epub 2015/03/17. doi: 10.1371/journal.ppat.1004728. PubMed PMID: 25774803.
32. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM. Host phylogeny determines viral persistence and replication in novel hosts. PLoS Pathogens. 2011;7((9)):e1002260. doi: 10.1371/journal.ppat.1002260.
33. Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. Estimating divergence dates and substitution rates in the Drosophila phylogeny. Molecular Biology and Evolution. 2012;29(11):3459-73.

473 34. Tamura K, Subramanian S, Kumar S. Temporal patterns of fruit fly (Drosophila) evolution
474 revealed by mutation clocks. Mol Biol Evol. 2004;21(1):36-44. Epub 2003/09/02. doi:
475 10.1093/molbev/msg236
476 msg236 [pii]. PubMed PMID: 12949132.
477 35. Storey JD. A direct approach to false discovery rates. J Roy Stat Soc B. 2002;64:479-98. doi:
478 Unsp 1369-7412/02/64479
479 Doi 10.1111/1467-9868.00346. PubMed PMID: WOS:000177425500009.
480 36. Jan E. Divergent IRES elements in invertebrates. Virus Res. 2006;119(1):16-28. doi:
481 10.1016/j.virusres.2005.10.011. PubMed PMID: 16307820.
482 37. Johnson KN, Christian PD. The novel genome organization of the insect picorna-like virus
483 Drosophila C virus suggests this virus belongs to a previously undescribed virus family. J Gen Virol.
484 1998;79 (Pt 1):191-203. Epub 1998/02/14. PubMed PMID: 9460942.
485 38. Nakashima N, Nakamura Y. Cleavage sites of the "P3 region" in the nonstructural polyprotein
486 precursor of a dicistrovirus. Arch Virol. 2008;153(10):1955-60. doi: 10.1007/s00705-008-0208-5.
487 PubMed PMID: 18810573.
488 39. Nakashima N, Uchiumi T. Functional analysis of structural motifs in dicistroviruses. Virus Res.
489 2009;139(2):137-47. doi: 10.1016/j.virusres.2008.06.006. PubMed PMID: 18621089.
490 40. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the
491 BEAST 1.7. Mol Biol Evol. 2012;29(8):1969-73. Epub 2012/03/01. doi: 10.1093/molbev/mss075.
492 PubMed PMID: 22367748; PubMed Central PMCID: PMC3408070.
493 41. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, et al. Cross-species virus
494 transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology
495 Reviews. 2008;72(3):457-70. doi: 10.1128/mmbr.00004-08. PubMed PMID: WOS:000258951200004.
496 42. Woolhouse ME, Gowtage-Sequeria S. Host range and emerging and reemerging pathogens.
497 Emerg Infect Dis. 2005;11(12):1842-7. Epub 2006/02/21. PubMed PMID: 16485468.
498 43. Webby RJ, Webster RG. Emergence of influenza A viruses. Philos Trans R Soc Lond B Biol Sci.
499 2001;356(1416):1817-28. Epub 2002/01/10. doi: 10.1098/rstb.2001.0997. PubMed PMID: 11779380;
500 PubMed Central PMCID: PMC1088557.
501 44. Weinert LA, Welch JJ, Suchard MA, Lemey P, Rambaut A, Fitzgerald JR. Molecular dating of
502 human-to-bovid host jumps by Staphylococcus aureus reveals an association with the spread of
503 domestication. Biol Lett. 2012;8(5):829-32. Epub 2012/05/26. doi: 10.1098/rsbl.2012.0290. PubMed
504 PMID: 22628096; PubMed Central PMCID: PMC3440972.
505 45. Christian PD. Studies of Drosophila C and A viruses in Australian populations of Drosophila
506 melanogaster: Australian National University; 1987.
507 46. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrai G, Quintana JF, et al. The
508 discovery, distribution and evolution of viruses associated with Drosophila melanogaster. PLOS
509 Biology. 2015;13(7): e1002210.
510 47. Teixeira L, Ferreira A, Ashburner M. The Bacterial Symbiont Wolbachia Induces Resistance to
511 RNA Viral Infections in Drosophila melanogaster. Plos Biology. 2008;6(12):2753-63. doi: DOI
512 10.1371/journal.pbio.1000002. PubMed PMID: ISI:000261913700016.
513 48. Longdon B, Cao C, Martinez J, Jiggins FM. Previous Exposure to an RNA Virus Does Not
514 Protect against Subsequent Infection in Drosophila melanogaster. Plos One. 2013;8(9):e73833. doi:
515 10.1371/journal.pone.0073833.
516 49. Jousset FX, Plus N, Croizier G, Thomas M. [Existence in Drosophila of 2 groups of picornavirus
517 with different biological and serological properties]. C R Acad Sci Hebd Seances Acad Sci D.
518 1972;275(25):3043-6. Epub 1972/12/18. PubMed PMID: 4631976.
519 50. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. The American
520 Journal of Hygiene. 1938;27:493-7.
521 51. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available at:
522 http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc.2010.

523 52. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
524 Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed PMID:
525 24695404; PubMed Central PMCID: PMCPMC4103590.
526 53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
527 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi:
528 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID:
529 PMCPMC2723002.
530 54. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC
531 Evolutionary Biology. 2007;7:214. doi: Artn 214
532 Doi 10.1186/1471-2148-7-214. PubMed PMID: ISI:000253468300001.
533 55. Team RDC. R: a language and environment for statistical computing. V 2.4. 2006.
534 56. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language.
535 Bioinformatics. 2004;20(2):289-90. PubMed PMID: 14734327.
536 57. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence
537 data. Genetics. 1992;132(2):583-9. Epub 1992/10/01. PubMed PMID: 1427045; PubMed Central
538 PMCID: PMC1205159.
539