PLOS | PATHOGENS

RESEARCH ARTICLE

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

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

1 Biosciences, College of Life & Environmental Sciences, University of Exeter, Penryn Campus, Penryn, United Kingdom, **2** Department of Genetics, University of Cambridge, Cambridge, United Kingdom, **3** CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal

* b.longdon2@exeter.ac.uk

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

Editor: Adam S. Lauring, University of Michigan, UNITED STATES

Received: January 15, 2018

Accepted: February 27, 2018

Published: April 12, 2018

Data Availability Statement: Sequence data (fastq files) are available in the NCBI SRA (Accession: SRP119720). BAM files, data and R scripts for analysis in the main text are available from the NERC data repository (https://doi.org/10.5285/4434a27d-5288-4f2e-88ac-4b1372e4d073).

Funding: BL and FMJ are supported by a Natural Environment Research Council grant (NE/L004232/1 http://www.nerc.ac.uk/) and by an European Research Council grant (281668, DrosophilaInfection, http://erc.europa.eu/). JMA

Abstract

Host shifts, where a pathogen invades and establishes in 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 changes in 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. When we compared viruses that had evolved in 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.

Author summary

Host shifts, where a pathogen jumps from one host species to another, are a major source of infectious disease. Hosts shifts are more likely to occur between related host species and often rely on the pathogen evolving adaptations that increase their fitness in the novel host. Here we have investigated how viruses evolve in different host species, by experimentally evolving replicate lineages of an RNA virus in 19 different host species that shared a common ancestor 40 million years ago. We then deep sequenced the genomes of these viruses to examine the genetic changes that have occurred in different host species that vary in their relatedness. We found that parallel mutations–that are indicative of selection–were significantly more likely to occur within viral lineages from the same host, and between viruses evolved in closely related species. This suggests that a mutation that may adapt a virus to a given host, may also adapt it to closely related host species.

was supported by a grant from the Portuguese Ministério da Ciência, Tecnologia e Ensino Superior (SFRH/BD/72381/2010). 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). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Host shifts–where a pathogen jumps into and establishes in 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 are increasing numbers of examples demonstrating that adaptation to the new host occurs following a host shift [8, 9]. 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 [10, 11]. 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 [12, 13]. 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 [14, 15].

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, 16]. Similarly, five parallel mutations have been observed in the two independent epidemics of SARS coronavirus following its jump from palm civets into humans [17]. 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 [18]. Likewise, a study passaging Tobacco Etch Potyvirus on four plant species found parallel mutations occurred only when the virus infected the same host species [19]. These parallel mutations provide compelling evidence that these genetic changes are adaptive, with the same mutations evolving independently in response to natural selection [20]. These studies have only used a limited number of hosts, and so do not provide information on how viral evolution occurs across a wide phylogenetic breadth of host species.

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 [2, 21–34]. Here, we asked whether viruses acquire the same genetic changes when evolving in the same and 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 [35, 36]. 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 evolve in different host species we serially passaged DCV in 19 species of Drosophilidae. 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 (Fig 1). In multiple cases these variants had nearly reached fixation (Fig 1).

We next examined whether the same genetic changes occur in parallel when different populations encounter the same host species. 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 (Fig 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 (Fig 2; $P<0.001$). Furthermore, we

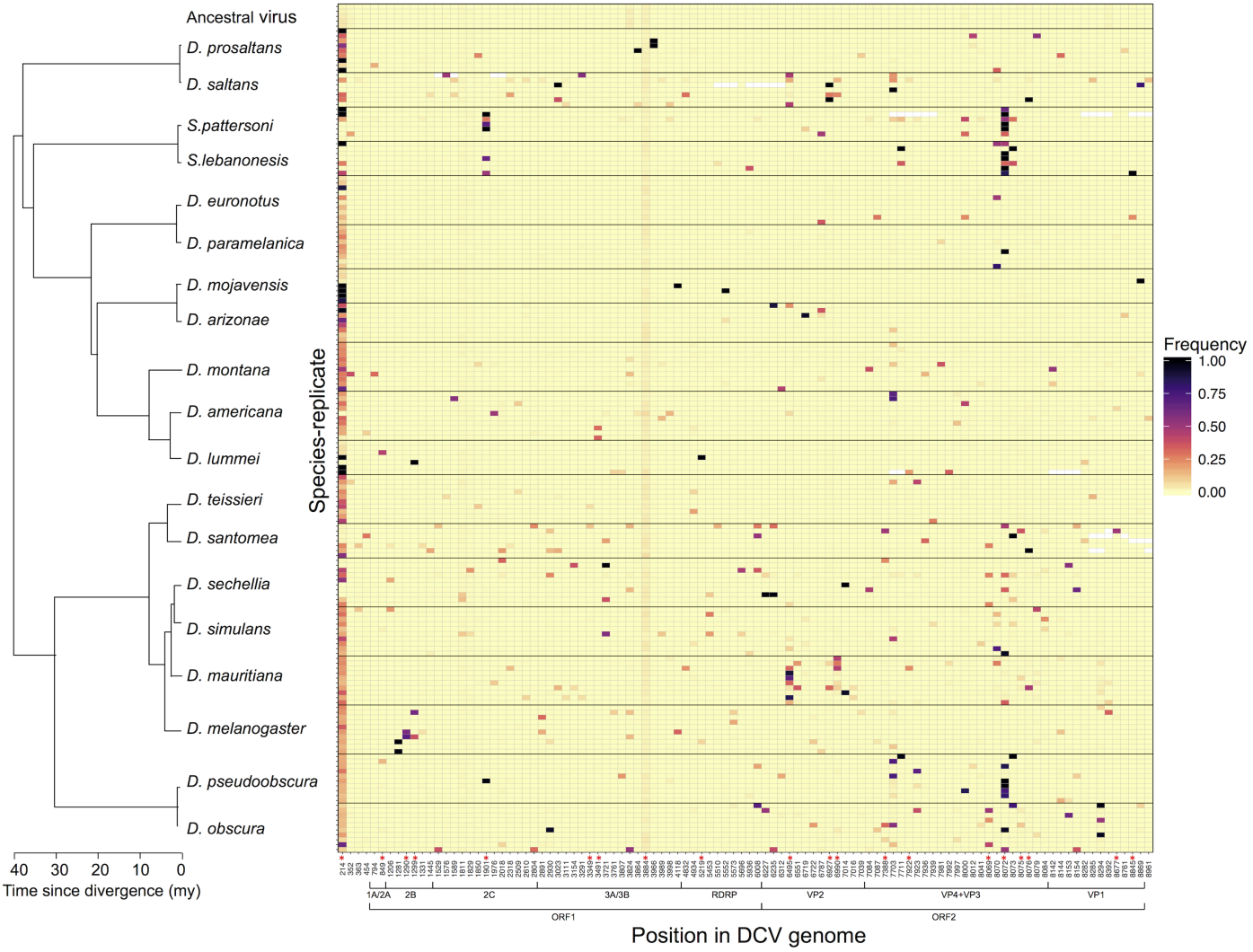

Fig 1. The frequency of SNPs in viral lineages that have evolved in different host species. Each row represents an independent viral lineage. Viruses that evolved in different host species are separated by black horizontal lines. Each column represents a polymorphic site in the DCV genome, and only sites 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. Sites with SNP frequencies that 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 [38–42] are below the x axis. Information on the distribution of mutations across the genome and whether they are synonymous or non-synonymous can be found in the supplementary results. Sites with missing data are shown in white. The phylogeny was inferred under a relaxed molecular clock [33, 43] and the scale axis represents the approximate age since divergence in millions of years (my) based on estimates from: [35, 36].

https://doi.org/10.1371/journal.ppat.1006951.g001

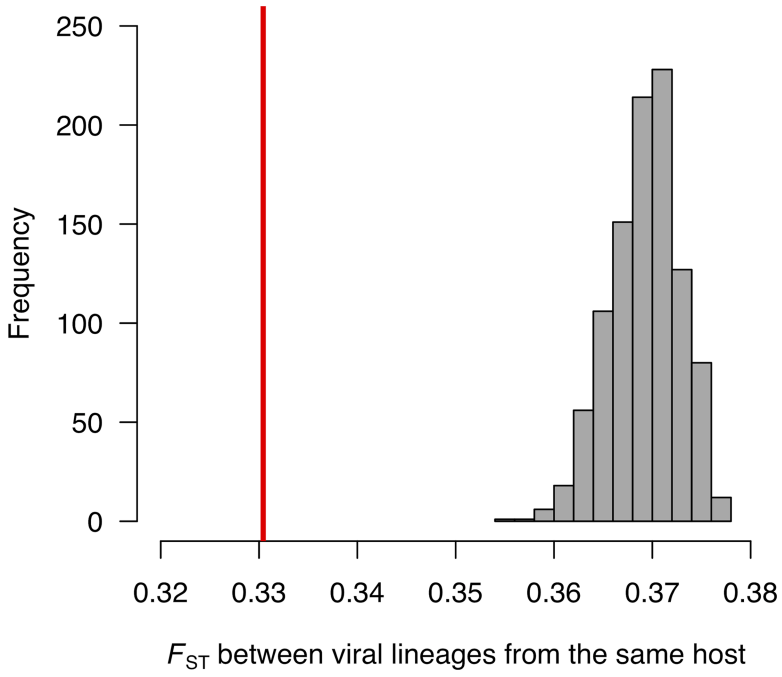

Fig 2. Viral lineages from the same host species were genetically more similar to each other than to lineages from different host species. The mean pairwise F_{ST} between all possible pairs of viral lineages from the same host species was calculated. The red line shows the observed value. The grey bars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

https://doi.org/10.1371/journal.ppat.1006951.g002

found no evidence of differences in substitution biases in the different host species (Fisher Exact Test: $p = 0.14$; see methods), suggesting that this pattern is not driven by changes in the types of mutations in 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 Fig 2 for each SNP). We identified 56 polymorphic sites with a significant signal of parallel evolution within the same host species ($P<0.05$; significantly parallel sites are shown with a red asterisk in Fig 1; the false discovery rate is estimated to be 17% [37]).

Viruses in closely related hosts are genetically more similar

We investigated if viruses passaged through closely related hosts showed evidence of parallel genetic changes. 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 (Fig 3A). This is reflected in a significant positive relationship between virus F_{ST} and host genetic distance (Fig 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 Fig 3 (false discovery rate >0.49 for all SNPs).

Two of the most striking examples of parallel evolution in related species are in *Scaptodrosophila pattersoni* and *S. lebanonensis*, which show two high frequency parallel mutations. These are a synonymous mutation in the 2C replicase protein at position 1901 and a triallelic non-synonymous mutation in a viral capsid protein at position 8072. However, the wider pattern of parallel evolution is not driven by these two examples, as the results remained significant after

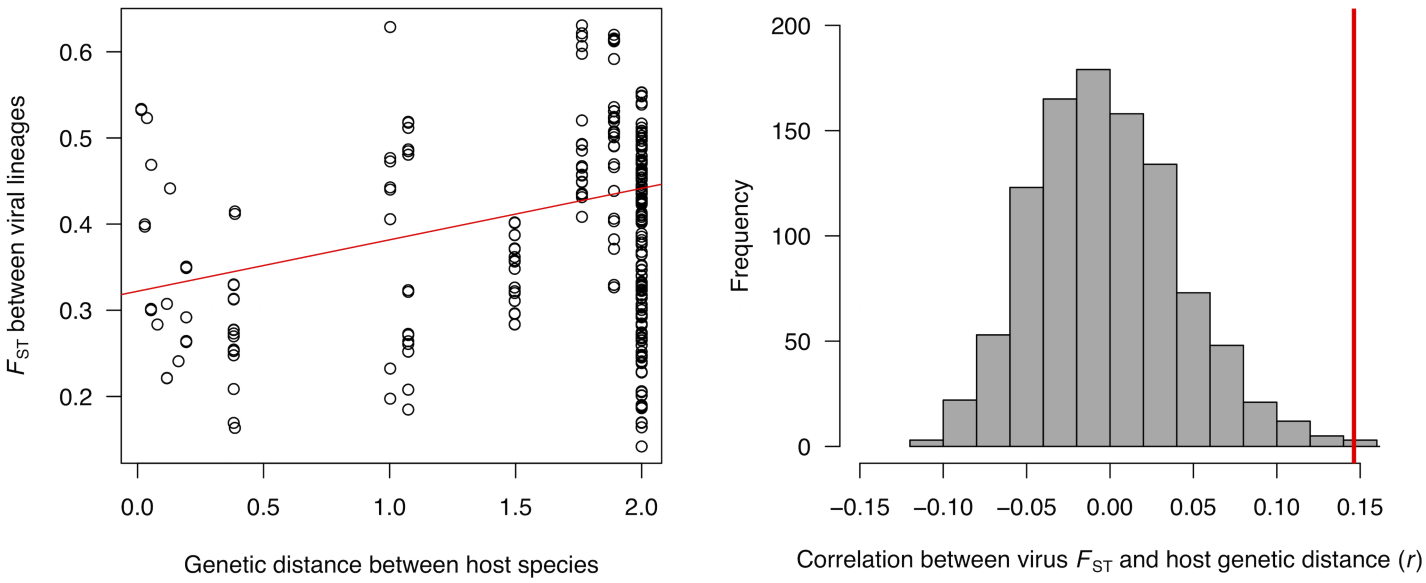

Fig 3. Viral lineages from more closely related host species are genetically more similar. (A) The correlation between the genetic differentiation of viral lineages and the genetic distance between the species they have evolved in. Linear regression line is shown in red. Genetic distances were scaled 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.

https://doi.org/10.1371/journal.ppat.1006951.g003

viruses that had evolved in these two species were removed from the dataset (within species parallelism: $P<0.001$; between species parallelism: $P = 0.013$).

Discussion

When a pathogen infects a novel host species, it finds itself in a new environment to which it must adapt [4, 8, 10, 44]. 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 [20]. We then examined whether these same genetic changes might occur in 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 evolve 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.

Phylogenetic patterns of host adaptation may in part explain why pathogens tend to be more likely to jump between closely related host species. 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, 21–34]. For example, in a large cross-infection experiment involving 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 [34]. 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 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 [45, 46], and in laboratory cross-infection experiments [33, 34]. 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 (Fig 1).

The function of these mutations is unknown, but in other systems adaptations after host shifts have been found to enhance the ability of the virus to bind to host receptors [11], increase replication rates [16] or avoid the host immune response [8, 10, 47]. Of the high frequency significant SNPs (shown in Fig 1 with red asterisk) nine occur in the non-structural proteins (ORF1: RNA dependant RNA polymerase, the putative protease and helicase proteins), and eleven occur in the capsid proteins (ORF2). Interestingly, none were in DCV-1A, which suppresses the host antiviral RNAi defences [48]. It will be of interest to examine the functions of the parallel mutations we detected, and characterise phenotypically how they affect viral infectivity and replication.

One mutation rose to a high frequency across all the host species (Fig 1, position 214 in the 3' un-translated region). This is unlikely to be an error in the genome sequencing, as it did not occur when we sequenced the ancestral virus. This may have been due to natural selection favouring this change in all species, perhaps because there was a strongly deleterious mutation at this site in the virus we cloned or due to the virus going from cell culture to being passaged *in vivo*.

Previous studies have elegantly demonstrated parallel evolution following host shifts (eg [18]). However, these are often in cell culture, and so do not reflect the heterogeneity of tissue and cell types in whole animals that occur in studies *in vivo* (although see [49] that suggests otherwise). The complex nature of different tissue types *in vivo* coupled with a limited number of generations may explain why some parallel SNPs have remained at a low frequency in this study. Following a host shift, viruses must sometimes acquire specific mutations that allow them to be transmitted in their new host [9, 10]. As we artificially inoculated the virus, this aspect of adaptation to a new host is missing from our study.

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 related host species. 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.

Methods

Virus production

DCV is a positive sense RNA virus in the family Discistroviridae that was isolated from *D. melanogaster*, which it naturally infects in the wild [50, 51]. To minimise the amount of genetic variation in the DCV isolate we used to initiate the experimental evolution study, we aimed to isolate single infectious clones of DCV using a serial dilution procedure. DCV was produced in Schneider's Drosophila line 2 (DL2) cells [52] as described in [53]. Cells were cultured at 25˚C in Schneider's Drosophila Medium with 10% Fetal Bovine Serum, 100 U/ml penicillin and 100 μg/ml streptomycin (all Invitrogen, UK). The DCV strain used was isolated from

D. melanogaster collected in Charolles, France [54]. DL2 cells were seeded into two 96-well tissue culture plates at approximately 10^4 cells in 100 μl of media per well. Cells were allowed to adhere to the plates by incubating at 25˚C for five hours or over-night. Serial 1:1 dilutions of DCV were made in complete Schneider's media, giving a range of final dilutions from $1:10^8$–$1:4x10^{14}$. 100 μl of these dilutions were then added to the cells and incubated for 7 days, 8 replicates were made for each DCV dilution. Each well was then examined for DCV infection of the DL2 cells, and a well was scored as positive for DCV infection if clear cytopathic effects were present in the majority of the cells. The media was taken from the wells with the greatest dilution factor that were scored as infected with DCV and stored at -80˚C. This processes was then repeated using the DCV samples from the first dilution series. One clone, B6A, was selected for amplification and grown in cell culture as described above. Media containing DCV was removed and centrifuged at 3000 x g for 5 minutes at 4˚C to pellet any remaining cell debris, before being aliquoted and stored at -80˚C. The Tissue Culture Infective Dose 50 ($TCID_{50}$) of the DCV was 6.32 x 10^9 infectious particles per ml using the Reed-Muench end-point method [55].

Inoculating fly species

We passaged the virus through 19 species of Drosophilidae, with 6–10 independent replicate passages for each species. We selected species from across the phylogeny (that shared a common ancestor approximately 40 million years ago [35, 36]), but included clades of closely related species that recently shared common ancestors less than 5 million years ago (Fig 1). All fly stocks were reared at 22˚C. Stocks of each fly species were kept in 250ml bottles at staggered ages. Flies were collected and sexed, and males were placed on cornmeal medium for 4 days before inoculation. Details of the fly stocks used can be found in the supplementary materials.

4–11 day old males were infected with DCV using a 0.0125 mm diameter stainless steel needle (26002–10, Fine Science Tools, CA, USA) dipped in DCV solution. For the first passage this was the cloned DCV isolate in cell culture supernatant (described above), and then subsequently was the virus extracted from the previous passage (described below). The needle was pricked into the pleural suture on the thorax of flies, towards the midcoxa. Each replicate was infected using a new needle and strict general cleaning procedures were used to minimise any risk of cross-contamination between replicates. Species were collected and inoculated in a randomised order each passage. Flies were then placed into vials of cornmeal medium and kept at 22˚C and 70% relative humidity. Flies were snap frozen in liquid nitrogen 3 days post-infection, homogenised in Ringer's solution (2.5μl per fly) and then centrifuged at 12,000g for 10 mins at 4˚C. The resulting supernatant was removed and frozen at -80˚C to be used for infecting flies in the subsequent passage. The remaining homogenate was preserved in Trizol reagent (Invitrogen) and stored at -80˚C for RNA extraction. The 3 day viral incubation period was chosen based on time course and pilot data showing that viral load reaches a maximum at approximately 3 days post-infection. This process was repeated for 10 passages for all species, except *D. montana* where only 8 passages were carried out due to the fly stocks failing to reproduce. Each lineage was injected into a mean of 11 flies at each passage (range 4–18). Experimental evolution studies in different tissue types have seen clear signals of adaptation in 100 virus generations [18]. Based on $\log_2$ change in RNA viral load we estimate that we have passaged DCV for approximately 100–200 generations.

Sequencing

After passaging the virus, we sequenced evolved viral lineages from 19 host species, with a mean of 9 independent replicate lineages of the virus per species (range 6–10 replicates). cDNA was

synthesised using Invitrogen Superscript III reverse-transcriptase with random hexamer primers (25˚C 5mins, 50˚C 50mins, 70˚C 15mins). The genome of the evolved viruses, along with the initial DCV ancestor (x5) were then amplified using Q5 high fidelity polymerase (NEB) in nine overlapping PCR reactions (see supplementary Table S2 for PCR primers and cycle conditions). Primers covered position 62-9050bp (8989bp) of the Genbank refseq (NC_001834.1) giving 97% coverage of the genome. PCRs of individual genomes were pooled and purified with Ampure XP beads (Agencourt). Individual Nextera XT libraries (Illumina) were prepared for each viral lineage. In total we sequenced 173 DCV pooled amplicon libraries on an Illumina MiSeq (Cambridge Genomic Service) v3 for 600 cycles to give 300bp paired-end reads.

Bioinformatics and variant calling

FastQC, version 0.11.2 [56] was used to assess read quality and primer contamination. Trimmomatic, version 0.32 [57] was used to removed low quality bases and adaptor sequences, using the following options: MINLEN = 30 (Drop the read if it is below 30 base pairs), TRAILING = 15 (cut bases of the end of the read if below a threshold quality of 15), SLIDINGWINDOW = 4:20 (perform a sliding window trimming, cutting once the average quality within a 4bp window falls below a threshold of 20), and ILLUMINACLIP = TruSeq3-PE.fa:2:20:10:1:true (remove adapter contamination; the values correspond in order to: input fasta file with adapter sequences to be matched, seed mismatches, palindrome clip threshold, simple clip threshold, minimum adapter length and logical value to keep both reads in case of read-through being detected in paired reads by palindrome mode).

To generate a reference ancestral Drosophila C Virus sequence we amplified the ancestral starting virus by PCR as above. PCR products were treated with exonuclease 1 and Antarctic phosphatase to remove unused PCR primers and dNTPs and then sequenced directly using BigDye reagents (ABI) on an ABI 3730 capillary sequencer in both directions (Source Bioscience, Cambridge, UK). Sequences 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/) [58]. 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 a phylogeny produced previously [33]. This time-based tree (where the distance from the root to the tip is equal for all taxa) was inferred using seven genes with a relaxed molecular clock model in BEAST (v1.8.0) [43, 59]. The tree was pruned to the 19 species used using the Ape package in R [60, 61].

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 [62]:

$$F_{ST} = \frac{H_b - H_w}{H_b} \quad \text{(Eq1)}$$

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 Eq (1). H_w was calculated separately for the two lineages being compared, and the unweighted mean used in Eq (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 S1 and S2 Figs) using generalised linear mixed models that include the phylogeny as a random effect in the MCMCglmm package in R [63] as described previously [34]. We also examined the distribution of SNPs and whether they were synonymous or non-synonymous (see supplementary results).

To test whether there were systematic differences in the types of mutations occurring in the different host species, we classified all the SNPs into the six possible types (A/G, A/T, A/C, G/T, G/C and C/T). We then counted the number of times each type of SNP arose in each host species at a frequency above 5% and in at least one biological replicate (SNPs in multiple biological replicates were only counted once). This resulted in a contingency table with 6 columns and 19 rows. We tested for differences between the species in the relative frequency of the 6 SNP types by simulation [64].

Accession numbers

Sequence data (fastq files) are available in the NCBI SRA (Accession: SRP119720). BAM files, data and R scripts for analysis in the main text are available from the NERC data repository (https://doi.org/10.5285/4434a27d-5288-4f2e-88ac-4b1372e4d073).

Supporting information

S1 Supporting information. Supplementary methods, results, tables and figures.
(PDF)

Acknowledgments

Thanks to the Drosophila species stock centre for providing fly stocks and four anonymous reviewers for constructive comments.

Author Contributions

Conceptualization: Ben Longdon, Francis M. Jiggins.

Data curation: Ben Longdon, Jonathan P. Day, Joel M. Alves, Francis M. Jiggins.

Formal analysis: Ben Longdon, Joel M. Alves, Francis M. Jiggins.

Funding acquisition: Ben Longdon, Francis M. Jiggins.

Investigation: Ben Longdon, Jonathan P. Day, Joel M. Alves, Sophia C. L. Smith, Thomas M. Houslay, John E. McGonigle, Lucia Tagliaferri.

Methodology: Ben Longdon, Jonathan P. Day, Joel M. Alves, John E. McGonigle, Francis M. Jiggins.

Project administration: Ben Longdon, Francis M. Jiggins.

Resources: Ben Longdon, Joel M. Alves, Francis M. Jiggins.

Software: Joel M. Alves, John E. McGonigle, Francis M. Jiggins.

Supervision: Ben Longdon, Jonathan P. Day, Francis M. Jiggins.

Validation: Ben Longdon, Francis M. Jiggins.

Visualization: Ben Longdon, Francis M. Jiggins.

Writing – original draft: Ben Longdon, Francis M. Jiggins.

Writing – review & editing: Jonathan P. Day, Joel M. Alves, Sophia C. L. Smith, Thomas M. Houslay, John E. McGonigle, Lucia Tagliaferri.

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. https://doi.org/10.1098/rspb.2008.0284 PubMed PMID: ISI:000256387500014. PMID: 18445561
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. https://doi.org/10.1098/rstb.2001.0888 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] https://doi.org/10.1016/j.tree.2005.02.009 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. https://doi.org/10.1098/rstb.2010.0031 PubMed PMID: WOS:000280097000008. PMID: 20643738
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. https://doi.org/10.1038/438575a PubMed PMID: WOS:000233593100030. PMID: 16319873
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. https://doi.org/10.1126/science.1118391 PubMed PMID: WOS:000232997700042. PMID: 16195424
8. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, et al. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews. 2008; 72(3):457–70. https://doi.org/10.1128/MMBR.00004-08 PubMed PMID: WOS:000258951200004. PMID: 18772285
9. Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, James SL, et al. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science. 2012; 336 (6088):1541–7. Epub 2012/06/23. https://doi.org/10.1126/science.1222526 PMID: 22723414; PubMed Central PMCID: PMC3426314.
10. 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. https://doi.org/10.1371/journal.ppat.1004395 PMID: 25375777; PubMed Central PMCID: PMC4223060.
11. 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. https://doi.org/10.1146/annurev.micro.59.030804.121059 PMID: 16153179.
12. 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. https://doi.org/10.1016/j.cell.2016.10.014 PMID: 27814506; PubMed Central PMCID: PMCPMC5115602.
13. 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. https://doi.org/10.1016/j.cell.2016.10.013 PMID: 27814505; PubMed Central PMCID: PMCPMC5101188.
14. Shackelton LA, Parrish CR, Truyen U, Holmes EC. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc Natl Acad Sci U S A. 2005; 102(2):379–84. Epub 2005/01/01. https://doi.org/10.1073/pnas.0406765102 PMID: 15626758; PubMed Central PMCID: PMC544290.
15. 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. https://doi.org/10.1006/viro.1996.0021 PMID: 8560765.
16. 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. https://doi.org/10.1093/molbev/msm110 PMID: 17545188.
17. 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. https://doi.org/10.3201/eid1109.040773 PMID: 16229772; PubMed Central PMCID: PMC3310602.
18. 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. https://doi.org/10.1093/molbev/msn059 PMID: 18353798.

19. 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. https://doi.org/10.1093/molbev/msr314 PMID: 22319146.

20. Bollback JP, Huelsenbeck JP. Parallel Genetic Evolution Within and Between Bacteriophage Species of Varying Degrees of Divergence. Genetics. 2009; 181(1):225–34. https://doi.org/10.1534/genetics.107.085225 PubMed PMID: WOS:000262595500021. PMID: 19001294

21. 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. https://doi.org/10.1126/science.1188836 PubMed PMID: WOS:000280602700037. PMID: 20689015

22. 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. https://doi.org/10.1098/rstb.2012.0196 PMID: 23382420; PubMed Central PMCID: PMC3678322.

23. 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. https://doi.org/10.1111/j.1461-0248.2012.01858.x PMID: 22913776.

24. 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. https://doi.org/10.1086/676589 PMID: 24921601.

25. 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. https://doi.org/10.1111/1365-2656.12160 PMID: 24289314

26. 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. https://doi.org/10.1086/674445 PMID: 24464193

27. 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] https://doi.org/10.1093/molbev/msn234 PMID: 18922760.

28. 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. https://doi.org/10.1111/j.1420-9101.2009.01878.x PubMed PMID: WOS:000271785800019. PMID: 19878406

29. 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. https://doi.org/10.1073/pnas.0607968104 PubMed PMID: WOS:000245256700040. PMID: 17360396

30. Tinsley MC, Majerus MEN. Small steps or giant leaps for male-killers? Phylogenetic constraints to male-killer host shifts. Bmc Evolutionary Biology. 2007; 7. https://doi.org/10.1186/1471-2148-7-238 PubMed PMID: WOS:000252786000001. PMID: 18047670

31. 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. https://doi.org/10.1111/j.1558-5646.2008.00579.x PMID: 19054050.

32. 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. PMID: 12703944

33. 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. https://doi.org/10.1371/journal.ppat.1004728 PMID: 25774803.

34. 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. https://doi.org/10.1371/journal.ppat.1002260 PMID: 21966271

35. 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. https://doi.org/10.1093/molbev/mss150 PMID: 22683811

36. Tamura K, Subramanian S, Kumar S. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol. 2004; 21(1):36–44. Epub 2003/09/02. https://doi.org/10.1093/molbev/msg236 [pii]. PMID: 12949132.

37. Storey JD. A direct approach to false discovery rates. J Roy Stat Soc B. 2002; 64:479–98. doi: Unsp 1369-7412/02/64479 https://doi.org/10.1111/1467-9868.00346 PubMed PMID: WOS:000177425500009.

38. Jan E. Divergent IRES elements in invertebrates. Virus Res. 2006; 119(1):16–28. https://doi.org/10.1016/j.virusres.2005.10.011 PMID: 16307820.

39. Johnson KN, Christian PD. The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family. J Gen Virol. 1998; 79 (Pt 1):191–203. Epub 1998/02/14. https://doi.org/10.1099/0022-1317-79-1-191 PMID: 9460942.

40. Nakashima N, Nakamura Y. Cleavage sites of the "P3 region" in the nonstructural polyprotein precursor of a dicistrovirus. Arch Virol. 2008; 153(10):1955–60. https://doi.org/10.1007/s00705-008-0208-5 PMID: 18810573.

41. Nakashima N, Uchiumi T. Functional analysis of structural motifs in dicistroviruses. Virus Res. 2009; 139(2):137–47. https://doi.org/10.1016/j.virusres.2008.06.006 PMID: 18621089.

42. UniProtKB [cited 2017]. Available from: http://www.uniprot.org/uniprot/O36966.

43. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29(8):1969–73. Epub 2012/03/01. https://doi.org/10.1093/molbev/mss075 PMID: 22367748; PubMed Central PMCID: PMC3408070.

44. Woolhouse ME, Gowtage-Sequeria S. Host range and emerging and reemerging pathogens. Emerg Infect Dis. 2005; 11(12):1842–7. Epub 2006/02/21. https://doi.org/10.3201/eid1112.050997 PMID: 16485468.

45. Webby RJ, Webster RG. Emergence of influenza A viruses. Philos Trans R Soc Lond B Biol Sci. 2001; 356(1416):1817–28. Epub 2002/01/10. https://doi.org/10.1098/rstb.2001.0997 PMID: 11779380; PubMed Central PMCID: PMC1088557.

46. Weinert LA, Welch JJ, Suchard MA, Lemey P, Rambaut A, Fitzgerald JR. Molecular dating of human-to-bovid host jumps by Staphylococcus aureus reveals an association with the spread of domestication. Biol Lett. 2012; 8(5):829–32. Epub 2012/05/26. https://doi.org/10.1098/rsbl.2012.0290 PMID: 22628096; PubMed Central PMCID: PMC3440972.

47. Sauter D, Schindler M, Specht A, Landford WN, Munch J, Kim KA, et al. Tetherin-Driven Adaptation of Vpu and Nef Function and the Evolution of Pandemic and Nonpandemic HIV-1 Strains. Cell Host & Microbe. 2009; 6(5):409–21. https://doi.org/10.1016/J.Chom.2009.10.004 PubMed PMID: ISI:000272539700006. PMID: 19917496

48. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes & development. 2006; 20(21):2985–95. PubMed PMID: ISI:000241767900009.

49. Allison AB, Kohler DJ, Ortega A, Hoover EA, Grove DM, Holmes EC, et al. Host-Specific Parvovirus Evolution in Nature Is Recapitulated by In Vitro Adaptation to Different Carnivore Species. Plos Pathogens. 2014; 10(11). doi: ARTN e1004475 https://doi.org/10.1371/journal.ppat.1004475 PubMed PMID: WOS:000345515800014. PMID: 25375184

50. Christian PD. Studies of Drosophila C and A viruses in Australian populations of Drosophila melanogaster: Australian National University; 1987.

51. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrai G, Quintana JF, et al. The discovery, distribution and evolution of viruses associated with Drosophila melanogaster. PLOS Biology. 2015; 13(7): e1002210. https://doi.org/10.1371/journal.pbio.1002210 PMID: 26172158

52. Teixeira L, Ferreira A, Ashburner M. The Bacterial Symbiont Wolbachia Induces Resistance to RNA Viral Infections in Drosophila melanogaster. Plos Biology. 2008; 6(12):2753–63. https://doi.org/10.1371/journal.pbio.1000002 PubMed PMID: ISI:000261913700016. PMID: 19222304

53. Longdon B, Cao C, Martinez J, Jiggins FM. Previous Exposure to an RNA Virus Does Not Protect against Subsequent Infection in Drosophila melanogaster. Plos One. 2013; 8(9):e73833. https://doi.org/10.1371/journal.pone.0073833 PMID: 24040086

54. Jousset FX, Plus N, Croizier G, Thomas M. [Existence in Drosophila of 2 groups of picornavirus with different biological and serological properties]. C R Acad Sci Hebd Seances Acad Sci D. 1972; 275 (25):3043–6. Epub 1972/12/18. PMID: 4631976.

55. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. The American Journal of Hygiene. 1938; 27:493–7.

56. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc.2010.

57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404; PubMed Central PMCID: PMCPMC4103590.

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943; PubMed Central PMCID: PMCPMC2723002.

59. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 2007; 7:214. doi: Artn 214 https://doi.org/10.1186/1471-2148-7-214 PubMed PMID: ISI:000253468300001. PMID: 17996036

60. Team RDC. R: a language and environment for statistical computing. V 2.4. 2006.

61. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20(2):289–90. PMID: 14734327.

62. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992; 132(2):583–9. Epub 1992/10/01. PMID: 1427045; PubMed Central PMCID: PMC1205159.

63. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. Journal of Statistical Software. 2010; 33(2):1–22. PubMed PMID: WOS:000275203300001.

64. Patefield WM. Algorithm AS 159: An Efficient Method of Generating Random R × C Tables with Given Row and Column Totals. Journal of the Royal Statistical Society Series C (Applied Statistics). 1981; 30(1):91–7. https://doi.org/10.2307/2346669