

Gauging genetic diversity of generalists: A test of genetic and ecological generalism with RNA virus experimental evolution

Lele Zhao and Siobain Duffy*

Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, 14 College Farm Road, New Brunswick, NJ 08901, USA

*Corresponding author: E-mail: duffy@sebs.rutgers.edu

Abstract

Generalist viruses, those with a comparatively larger host range, are considered more likely to emerge on new hosts. The potential to emerge in new hosts has been linked to viral genetic diversity, a measure of evolvability. However, there is no consensus on whether infecting a larger number of hosts leads to higher genetic diversity, or whether diversity is better maintained in a homogeneous environment, similar to the lifestyle of a specialist virus. Using experimental evolution with the RNA bacteriophage phi6, we directly tested whether genetic generalism (carrying an expanded host range mutation) or environmental generalism (growing on heterogeneous hosts) leads to viral populations with more genetic variation. Sixteen evolved viral lineages were deep sequenced to provide genetic evidence for population diversity. When evolved on a single host, specialist and generalist genotypes both maintained the same level of diversity (measured by the number of single nucleotide polymorphisms (SNPs) above 1%, $P = 0.81$). However, the generalist genotype evolved on a single host had higher SNP levels than generalist lineages under two heterogeneous host passaging schemes ($P = 0.001$, $P < 0.001$). RNA viruses' response to selection in alternating hosts reduces standing genetic diversity compared to those evolving in a single host to which the virus is already well-adapted.

Key words: phage evolution; genetic diversity; generalist; dsRNA virus.

1. Introduction

Measures of genetic diversity, including nucleotide diversity, phylogenetic distance, and quasispecies size, are frequently used as proxies for viral evolvability, making them useful metrics for understanding viral emergence (Dutta et al. 2008; Bordería, Stapleford, and Vignuzzi 2011; Day 2015). The evolvability associated with high levels of viral genetic diversity has been observed in both laboratory and clinical settings. For instance, more diverse populations of Influenza, Hepatitis C, and HIV-1 are more likely to evade antiviral therapy (Mas et al. 2010). Quasispecies complexity—a measure of genetic

diversity—has been suggested to be predictive of evolvability for individual pathogens (Mas et al. 2004). Genetic diversity is also positively correlated with viral fitness (Arenas, Lorenzo-Redondo, and Lopez-Galindez 2016) and virulence in emerging viral pathogens (Lauring and Andino 2010; Pita and Roossinck 2013). Understanding viral population genetic diversity is therefore a crucial component in the study of fast-evolving viruses. However, most research on viral genetic diversity focuses on clinical samples and disease outcomes (Fusaro et al. 2016; Hasing et al. 2016), and there have been fewer controlled studies on genetic diversity in viral populations (Korboukh et al. 2014; McCrone and Lauring 2016; Debbink et al. 2017).

Specialism and generalism are relative terms in evolution—a generalist is able to exist and thrive in more environments than a specialist. In virology, specialism usually restricts a virus to a single resource niche, while host generalism typically means infecting more hosts than a specialist (Duffy, Turner, and Burch 2006; García-Arenal and Fraile 2013; McLeish, Fraile, and García-Arenal 2018). The wider host range is usually caused by a genetic difference, often a single amino acid change, which leads to a more generalist genotype (Duffy, Turner, and Burch 2006; Hillung et al. 2014). Specialist and generalist genotypes can each outcompete the other under different ecological conditions: the former will thrive under constant environmental selection pressure, while the latter performs better given fluctuating environmental changes (Kassen 2002). ‘Evolutionary rescue’ of specialists can also occur when varied host heterogeneity favors the adaptation of specialists in mixed specialist and generalist populations. How exactly generalists cope with ecological changes is an open area of research, and one way could be by maintaining higher levels of genetic diversity. Indeed, mathematical models imply that generalists should have higher levels of genetic diversity (Nowak et al. 1991; Morand, Manning, and Woolhouse 1996; Haydon and Woolhouse 1998), but these predictions have not been tested in viruses. Successful specialists are hypothesized to maintain less genetic diversity (Liberian and Feldman 1986), which has been experimentally verified (Buckling, Wills, and Colegrave 2003), but the converse, that generalists are more diverse, has also not yet been directly tested in viruses (but has been observed in *Caenorhabditis* nematodes species, Li et al. 2014).

While experimental studies that could resolve the relationship between generalism and genetic diversity have mixed results (Kassen 2002), it remains conventional wisdom that generalist pathogens are more prepared to survive and reproduce in new environments (Dennehy et al. 2013). In fact, emerging and re-emerging pathogens are most likely host generalists (Woolhouse and Gowtage-Sequeria 2005). Directly and indirectly selected generalist virus populations infect novel hosts with higher initial fitness than specialist populations (Turner et al. 2010), and bacterial generalists have an increased ability to invade novel environments (Ketola et al. 2013). Although, these studies did not directly assess the level of genetic diversity in these evolvable generalists, these generalists all displayed features associated with higher genetic diversity such as emergence potential and higher fitness. In contrast, intensive sequencing of dengue virus as it infected multiple tissue types showed reduced genetic diversity compared to those in a single tissue (Lequime et al. 2016). Further complicating the relationship between generalism and genetic diversity is the theoretical prediction that using one resource (ecological specialism) or multiple resources (ecological generalism) should select for different arrays of mutations, but not necessarily different levels of overall population diversity (Aguirre, Lázaro, and Manrubia 2009).

As the literature is mixed with predictions of the relationship between generalism and population genetic diversity, we designed an experiment to directly address this relationship. Because ecological generalist populations experience changing environmental selective pressures provided by diverse hosts, genetic diversity may be maintained at higher levels than ecological specialists. Although there have been a few studies of genetic diversity of generalists (e.g. Saito and Tojo 2016), none have been in a controlled experiment where comparison to an isogenic specialist is possible. Using the RNA *Pseudomonas* bacteriophage phi6 as a model system, which has a high mutation

rate (Lythgoe and Chao 2003) and can have large population sizes (Elena 2016), we examined whether generalist genotypes evolving in a single environment (ecological specialism) generated similar levels of diversity as specialist genotypes, and whether alternating host environments (ecological generalism) led to higher levels of genetic diversity. This will determine whether standing genetic diversity is a pleiotropic effect of the host range mutation conferring a generalist genotype (equal diversity for generalists evolved in a single and multiple hosts) or if ecological history is the more significant determinant of population genetic diversity (unequal diversity). The ~13 kb genomes of phi6 allowed for significant read depth even when multiplexing on MiSeq flow cells, and we were able to directly compare the number of single nucleotide polymorphisms (SNPs), as an indicator of genetic diversity, from experimentally evolved populations. Our results showed that genetic generalism does not produce higher levels of population genetic diversity. Combining results from deep sequencing and phenotypic assays, we observed selection pressure from generalist ecologies purging instead of promoting population genetic diversity.

2. Material and methods

2.1 Strains and culture conditions

Phi6 is a lipid enveloped double-strand RNA *Pseudomonas* bacteriophage with three segments: Small, Medium, and Large (Semancik, Vidaver, and Van Etten 1973; Vidaver, Koski, and Van Etten 1973). The starting genotypes of phi6 are twice-plaque purified wild-type phi6 (specialist, S, strain #21781-B1, American Type Culture Collection, Bethesda, MD), and its isogenic host range mutant (generalist, G) that has one nonsynonymous mutation in the p3 attachment protein gene, E8G (Duffy, Turner, and Burch 2006). The specialist can infect *Pseudomonas syringae* pathovar *phaseolicola* (P) strain HB10Y (ATCC no. 21781), while the generalist can infect two additional hosts (*P. syringae* pathovar *tomato* (T), *Pseudomonas pseudoalcaligenes* East River isolate A (E)) (Duffy, Turner, and Burch 2006). An additional bacterial strain neither the specialist nor generalist could readily infect, *P. syringae* pathovar *atofaciens* (A), was used for host range mutation frequency assays. All of these additional *P. syringae* strains were streaked from glycerol stocks originally obtained from G. Martin (Cornell University, Ithaca, NY), and the *P. pseudoalcaligenes* strain was obtained from L. Mindich (Public Health Research Institute, Rutgers University, Newark, NJ).

All cultures were grown in LC media (Luria-Bertani lysogeny broth at pH 7.5), and phages were propagated by mixing with either 200 µl P, 50 µl T, or 5 µl E bacteria overnight culture in 3 ml 0.7 per cent agar top layer poured on 1.5 per cent agar LC plates as previously described (Duffy, Turner, and Burch 2006). Volumes of each bacteria culture were predetermined to ensure similar initial optical density, OD600: 0.03 AU (absorbance unit), and to accommodate the different growth rates of the bacterial strains (Duffy, Turner, and Burch 2006). All incubation was at 25 °C. Phage lysates were stored in 4 °C for short periods of time (<5 days), and were archived in 40 per cent glycerol at –20 °C as previously described (Duffy, Turner, and Burch 2006).

2.2 Experimental evolution

The phi6 specialist (S) and phi6 generalist (G) were experimentally passaged for 30 days. Specialist and generalist single plaques were first raised on their most recent host of propagation

(S on P, G on T) to high titer lysates. The genotypes were raised on different host to maintain a stable genotype in the starting lysate given their ecological history and source of isolation, as G was an isogenic mutant from S (see Section 4 for further explanation; [Duffy, Turner, and Burch 2006](#)). Then the two strains were serially passaged only on P for thirty passages (S/P, G/P to designate passaging on P), which allows assessment of the effects of genetic generalism in a common ecological selection scheme. Additionally, the generalist population was passaged on two alternating host schemes: on hosts P and T (G/PT) and hosts P and E (G/PE), the results of which can be compared to G/P to see the effects of different ecological histories on population genetic diversity. During each passage, we harvested ~350 plaques (range 272–431), diluted and plated from previous day's incubation. Incubation time for each passage was 22–24 h, which allows for approximately five generations' growth ([Burch and Chao 1999](#)), therefore the sixteen lineages (four lineages for each passage scheme) were evolved for a total of ~150 generations. The bottleneck size was chosen to minimize the effect of drift and reduce the potential for overlapping plaques, which creates competition for resources and allow gene exchange in this segmented RNA virus ([Dennehy et al. 2013](#)).

2.3 Phenotypic assays

Every third passages, the four G/P lineages were tested for the potential loss of T host range. Diluted phage lysates were plated with both P (200 μ l) and T (10 μ l) on the plate, and clear and turbid plaques were enumerated. Generalists would infect both hosts and produce a clear plaque, turbid plaques would result from a loss of T host range. Every 10th passage, all sixteen lineages (at least 10^6 pfu of phage) were plated on 200 μ l host A to obtain the population's mutation frequency in the novel host A. The lysates from passage 10, 20, and 30 were also tested for heat shock tolerance by titrating on host P before and after 5 min incubation in a 50 °C heat block ([Dessau et al. 2012](#)).

2.4 Fitness assays

Genotype S and G lysates from Day 0, G/PT and G/PE lysates from Day 29, and all lysates from Day 30 were tested for their relative fitness during 24 h incubation on P (descendants of the generalist were also tested after 24 h incubation on hosts T and E). Approximately 1,000 plaque-forming units (pfu) of phage were mixed with equal pfu of a common competitor (host T common competitor (TCC) and host E common competitor (ECC) for host T and E testing, respectively) in paired growth assays ([Chao 1990](#); [Chao, Tran, and Tran 1997](#)), as previously described ([Duffy, Burch, and Turner 2007](#)). Six technical replicates were tested for each population, except for two populations with five replicates specified in Section 3. Ratios of phage population to common competitor in the mixtures were enumerated from the initial mix and after 24 h of growth on the relevant host. Phage were distinguished and enumerated by plating on a mixed bacterial lawn containing P (200 μ l) and A (10 μ l); as the common competitors are able to infect A and P, therefore produced clear plaques, while the populations of interest cannot readily infect A, but only P, therefore produced turbid plaques. The relative fitness of the phage population is simply a ratio of values from initial mixture and after 24 h ([Duffy, Turner, and Burch 2006](#)).

Statistical analyses of fitness data, including ANOVA and Tukey's honestly significant difference (HSD) tests, were performed in R ([R Development Core Team 2010](#)).

2.5 Library preparation

Viral RNA was extracted from S and G lysates from Day 0, G/P, G/PT, and G/PE lysates from Day 29 (when each was reared on P), and all lysates from Day 30 using QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA). Viral RNA were purified by 1 per cent low melt agarose gel electrophoresis (100 V, 40 min, IBI Scientific, IA) and Gelase digestion following manufacturer instructions (GELase™, Lucigen). Individual RNA samples at a final concentration of ~15 ng/ μ l were prepared into Illumina RNA libraries using Illumina TruSeq RNA Library Prep Kit (Illumina, CA) ([Zhao et al. 2019](#)). Viral RNA libraries were divided into four batches, each containing one Day 30 lineage from each passaging scheme, and Day 29 from the corresponding G/P, G/PT, and G/PE lineages. Single-end 150-cycle sequencing was done on Illumina MiSeq housed in Foran Hall, Rutgers University (SEBS Genome Cooperative).

2.6 NGS data analysis

Raw reads were trimmed and filtered with cutadapt 1.12 (Q score cutoff: 30, minimum length cutoff: 75 bp, adapters and terminal Ns of reads removed) ([Martin 2011](#)). The reference genomes for the S and G genotypes were obtained from Day 0 Illumina sequencing results, and these full-length sequences were identical to the sequences obtained through Sanger sequencing of the same stocks ([Zhao et al. 2019](#)). The cleaned up reads were mapped to phi6 specialist and generalist genotype reference genomes using bwa mem with default settings ([Li and Durbin 2010](#); [Li 2013](#)). Although 5.74–59.09 per cent (median: 38.26%) of the NGS reads mapped to *Pseudomonas* host genomes, S/P, G/P, and G/PT population samples produced above 1,000 \times coverage for 91–97 per cent of the genome. G/PE population samples had above 1,000 \times coverage for 80–97 per cent of genome because five samples had lower than 1,000 \times large segment coverage. Additional file format conversion was performed using SAMtools ([Li et al. 2009](#)). Genome coverage data produced by Integrative Genomics Viewer IGVTools (count options: window size 1 and –bases) ([Robinson et al. 2011](#)). Whole genomes variant calling was performed using VarScan (default parameters with adjusted minimum coverage 500, minimum variant frequency 0.001, P-value threshold 0.05) ([Koboldt et al. 2012](#)). SNP data from VarScan were further analyzed (Shapiro–Wilk normality test, Brown–Forsythe test, ANOVA, and Tukey's HSD tests) and visualized in R ([R Development Core Team 2010](#)) and Excel (Microsoft, Redmond, WA). Reads were deposited in the sequence read archive accession number SRR8440639-68.

3. Results

3.1 Phenotypic assays

To test the effect of the generalist genotype on accumulated genetic diversity, it is important to assure that the G/P populations do not revert their host range mutation while being passed on single host. We monitored the loss of host range in G/P lineages by enumerating the proportion of phages unable to infect T anymore. The reversion frequency was stable throughout the 30 days with a mean of 1.1 per cent and median of 0.7 per cent of plaques appearing turbid. This showed that all four lineages of G/P populations were stable generalists throughout the experiment and revertants never became a large proportion of the population.

As a proxy for genetic diversity, we assessed the frequency of mutations to two novel environments: host range mutation

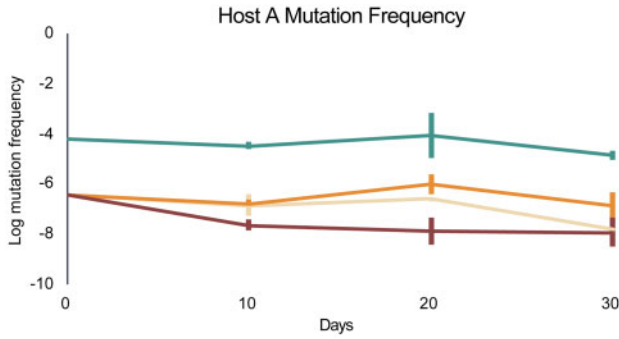


Figure 1. Average frequency of host A mutations among lineages. S/P in blue, G/P in beige, G/PT in orange, G/PE in maroon. Data collected every 10-days during the 30-day passaging, points are average values of four lineages, error bars are standard deviations.

to novel host A and withstanding heat shock. After sampling the lineages at Days 10, 20, and 30, we found that mutation frequency to infect host A is dependent on initial genotype and relatively invariant over the 30 days: S/P populations, and the S ancestor, had significantly ($P = 0.002$) higher mutation frequencies (mean: 1.21×10^{-4} , range: 5.99×10^{-6} to 6.65×10^{-4}) than any population with a G genotype (mean: 2.68×10^{-7} , range: $0-3.71 \times 10^{-7}$) (Fig. 1). In contrast, heat shock tolerance survival rates fluctuated for most populations over the 30 days and were not dependent on initial genotype (Fig. 2). Two G/PE populations became ten times less tolerant of heat shock during passaging, while all other G populations did not lose heat shock tolerance by Day 30 compared to the ancestor. We note that there are potential batch effects in our heat shock measurements, as they were conducted on Day 10, 20, and 30 of the passaging experiment. This may explain some of the fluctuation in values over the experiment.

3.2 Genetic diversity comparison

We deep sequenced Day 30 phage populations to quantify the number of SNPs. We also sequenced the Day 29 lysate of G/P lineages as an internal quality control and Day 29 lysate of G/PT and G/PE to account for any differences in SNP frequency that could occur due to the last host experienced. Because of host alternating passaging, G/PT and G/PE populations were grown on P on Day 29, and grown on T or E on Day 30, respectively. Raw numbers of SNPs called by VarScan were scaled by sequencing batch, which included all samples that shared the same MiSeq lane. We compared the scaled number of SNPs with a frequency above 0.1 per cent for the sequenced twenty-eight populations (seven groups with four replicates each: S/P, G/P-29, G/P-30, G/PT-29, G/PT-30, G/PE-29, G/PE-30). The data were tested for normality (Shapiro–Wilk test, $P = 0.29-0.96$) and homogeneity (Brown–Forsythe test, $P = 0.86$) before ANOVA were carried out. Most populations are indistinguishable from each other, however, a significantly higher number of SNPs with a frequency above 0.1 per cent were called from G/P-29 than that of G/PT-30 (adjusted $P = 0.04$), and G/PE-29 (adjusted $P = 0.001$). This is statistical evidence that ecological specialism leads to higher diversity than two alternating host passaging schemes that mimicked ecological generalism. In addition, SNPs with a frequency above 0.1 per cent called from G/PE-29 were significantly lower than that of G/PE-30 (adjusted $P = 0.03$), indicating a loss of diversity when the Day 29 lysate was plated on host P. Further analyses showed that these patterns were being driven

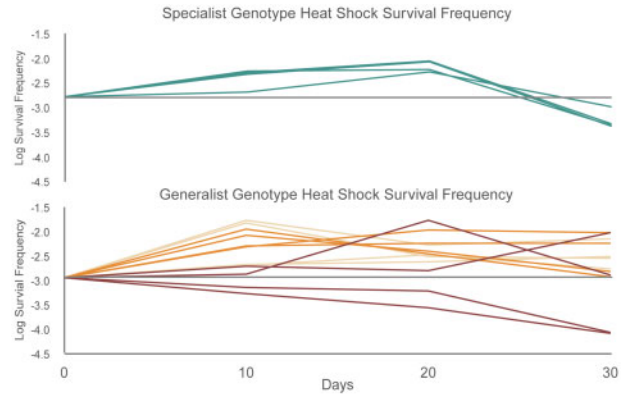


Figure 2. Heat shock tolerance mutation frequency of all lineages. At the top, heat shock tolerance of four S/P lineages (blue); below is the heat shock tolerance of four G/P lineages (beige), four G/PT lineages (orange), four G/PE lineages (maroon). Data collected every 10 days during the 30-day passaging. The horizontal gray line is the heat shock survival of the S ancestor (top) and G ancestor (bottom).

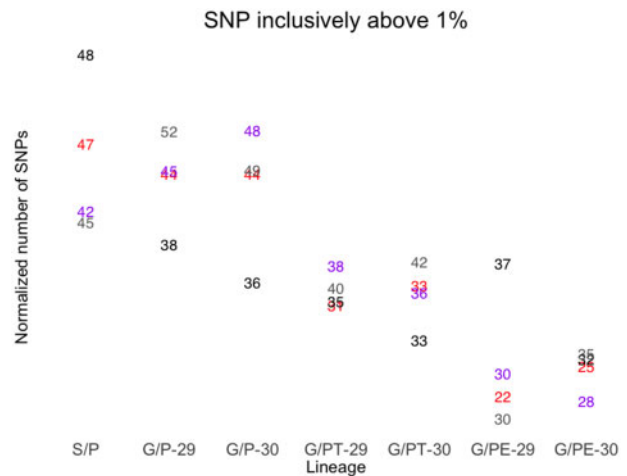


Figure 3. Number of SNPs above 1 per cent across all lineages of all passaging schemes. Each data point is labeled with the raw SNP count, and the position of the data point is normalized among samples sharing the same MiSeq sequencing run. This means that some raw SNP count numbers will appear lower on the y-axis than a smaller raw SNP count number from a different MiSeq run (and vice versa). The four colors indicate four different sequencing runs—the first replicate of each passaging scheme was run in one MiSeq flow cell (colored red), as were the second replicates (colored purple), the third replicates (colored black), and the fourth replicates (colored gray).

by lower frequency SNPs (between 0.1 and 1%, which comprise most of each population’s SNPs, $79.7 \pm 7.5\%$). Examining only the SNPs with a frequency between 0.1 and 1 per cent, identical pattern was observed as with all SNPs >0.1 per cent: most Days 29 and 30 populations were still statistically indistinguishable from each other (all data scaled by sequencing batch were normally distributed (Shapiro–Wilk test, $P = 0.10-0.98$) with equal variance (Brown–Forsythe test, $P = 0.83$)), excepting that G/P-29 populations still had more SNPs than G/PT-30 (adjusted $P = 0.08$) and G/PE-29 (adjusted $P = 0.005$), and the G/PE-29 populations had fewer SNPs than the G/PE-30 population (adjusted $P = 0.03$). Higher frequency SNPs (above 1% frequency) were more numerous in the S/P populations than the G/PT and G/PE populations (adjusted $P = 0.0002-0.02$), Fig. 3 (all data scaled by sequencing batch were normally distributed (Shapiro–Wilk test,

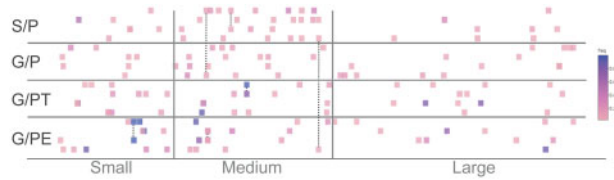


Figure 4. Top ten SNPs from each Day 30 population, four replicate populations for each treatment. Each square is a SNP placed according to its genomic position on the segmented $\phi 6$ genome. Color indicates frequency of the SNP. Vertical dashed lines show parallel SNPs by connecting the squares in the identical location from two populations. From left to right, with uppercase letters for amino acids within the indicated protein, and lowercase letters for non-coding region nucleotides, the lines are at P5 G101C on the Small segment; a830g, t869c, P6 S166G, P3 Q130R, P13 T65A on the Medium segment. All other specific SNPs shown are listed in [Supplementary Files S1–S28](#).

$P = 0.19–0.64$) with equal variance (Brown–Forsythe test, $P = 0.53$)). G/P populations continued to have significantly more SNPs than G/PE populations (adjusted $P = 0.0007–0.002$) and marginally more SNPs than G/PT populations (adjusted $P = 0.06–0.13$). There were no significant differences in SNP counts above 1 per cent between S/P and G/P populations, G/P-29 and G/P-30 populations, G/PT-29 and G/PT-30 populations, G/PE-29 and G/PE-30 populations, and G/PT and G/PE populations (both Days 29 and 30, [Fig. 3](#)).

While counts at different SNP frequency cutoffs are essential for comparisons of overall diversity between treatments, we also examined the top ten most frequent individual SNPs themselves ([Fig. 4](#)). These most frequent SNPs were fairly stable, even in alternate host passaging; there was an average overlap of 8.75 SNPs among the top ten from Days 29 and top ten SNPs from Day 30 for all tested lineages. Differences were typically due to an SNP ranking just under the top ten on one of the two days. However, there were relatively few shared SNPs among replicates or between treatments. Only six examples of parallel evolution were observed among these most frequent SNPs. Two of the parallel SNPs were in an intergenic region on the Medium segment: one lineage each of S/P and G/P shared the a830g, 11.28 and 9.99 per cent frequency; two G/PE lineages shared t869c, 50.05, and 5.83 per cent frequency ([Fig. 4](#)). One site showed the same fixed substitution in the lytic protein on the Small segment: two G/PE lineages, G101C in P5, 99.73 and 99.61 per cent frequency. The remaining parallel sites were in genes on the Medium segment: S166G in the membrane fusion protein P6 (two S/P lineages, 12.87 and 7.83%), Q130R in the host attachment protein P3 (two G/PT lineages, 98.15 and 79.01%), and T65A in the nonessential protein P13 (one lineage each from S/P and G/PE, 4.64 and 5.81%). At both parallel sites and at unique sites in the genome, [Fig. 4](#) shows that only ecological generalism (G/PT and G/PE) brought SNPs to very high frequency (highest frequency SNPs: 79.01–99.81%). This indicates stronger positive selection occurring in the alternating host passages than in the lineages evolved only on host P. The highest frequency SNP among the S/P lineages was 54.12 per cent, falling short of a completed selective sweep over 30 days of evolution.

3.3 Evolved population fitness

We measured the sixteen populations' fitness on the *Pseudomonas* hosts each could infect. All Day 30 populations, Day 29 of G/PT and G/PE, and Day 0 S and Day 0 G were mixed and grown with a common competitor to show their relative fitness on their shared host P ([Fig. 5](#)). Because these were genetically diverse populations instead of comparing the fitness of



Figure 5. Relative fitness of populations on P. Both lysates grown from the Day 0 ancestors are squares (specialist S is black, generalist G is gray). The Day 29 populations are diamonds and Day 30 populations are circles. S/P populations are blue, G/P populations are beige, G/PT populations are orange, G/PE populations are maroon. Values are means of six replicates (with the exception of G/PT lineage one at Day 30 and G/PE lineage four at Day 30 having five replicates), error bars are standard deviations.

clonal genotypes, the standard deviations among the technical replicates were large. There were no differences in fitness on host P among four lineages of Day 30 lysates from S/P ($P = 0.38$), G/P ($P = 0.69$), G/PT ($P = 0.30$), and G/PE ($P = 0.27$) and Day 29 lysates from G/PE ($P = 0.37$). However, the four lineages from Day 29 G/PT passing scheme were significantly different in their fitness on P ($P = 0.00258$). There was no difference in fitness of Day 0 S and G populations (two sample t-test, $P = 0.76$). We grouped all relative fitness measurements within each of the following categories and compared their difference in mean: Day 0 S, Day 0 G, Day 29 G/PT, Day 29 G/PE, Day 30 S/P, Day 30 G/P, Day 30 G/PT, Day 30 G/PE. There were significant differences among the groups ($P = 0.004$), specifically Day 30 G/PT populations had lower relative fitness compared to Day 30 S/P (adjusted $P = 0.03$), Day 30 G/P (adjusted $P = 0.04$), and Day 29 G/PE (adjusted $P = 0.004$), while other categories were statistically indistinguishable.

We tested the relative fitness of the generalist populations on hosts T and E separately ([Fig. 6](#)). The initial fitness of the G ancestor was higher on host T, which is another strain of the same species as the typical lab host P, than on the different species E. There were significant differences in fitness on host T among four lineages of Day 30 G/P ($P = 0.005$), Day 30 G/PT ($P < 0.001$), Day 30 G/PE ($P < 0.001$), Day 29 G/PT ($P = 0.002$), and Day 29 G/PE ($P = 0.004$). There was no difference in T fitness between Days 29 and 30 populations for G/PT (adjusted $P = 0.97$) or G/PE (adjusted $P = 0.36$). Understandably, the G/PT populations had a higher T fitness than the populations that had not been exposed to host T for part of the 30 days: the Day 0 generalist, G/P and G/PE populations (adjusted $P < 0.001$). When measuring fitness on host E, there were significant differences in fitness on host E among four lineages of Day 30 G/PT ($P < 0.001$), Day 30 G/PE ($P = 0.02$), but insignificant different among the lineages for Day 30 G/P ($P = 0.21$), Day 29 G/PT ($P = 0.62$), and Day 29 G/PE ($P = 0.10$). Yet again there was no difference in E fitness between Days 29 and 30 populations of G/PT (adjusted $P = 1$) and G/PE (adjusted $P = 0.59$). G/PE populations had significantly higher E fitness values than all other populations (adjusted $P < 0.001$), which had similar fitness values to each other (adjusted $P = 0.99–1$). Relative fitness gains were larger on host E than on host T because of the poorer starting fitness of the generalist on E (0.12, [Fig. 6](#)) compared to T (0.60, [Fig. 6](#)), which is consistent with the greater number of substitutions in the G/PE populations compared to the G/PT populations: selection was stronger on host E than host T.

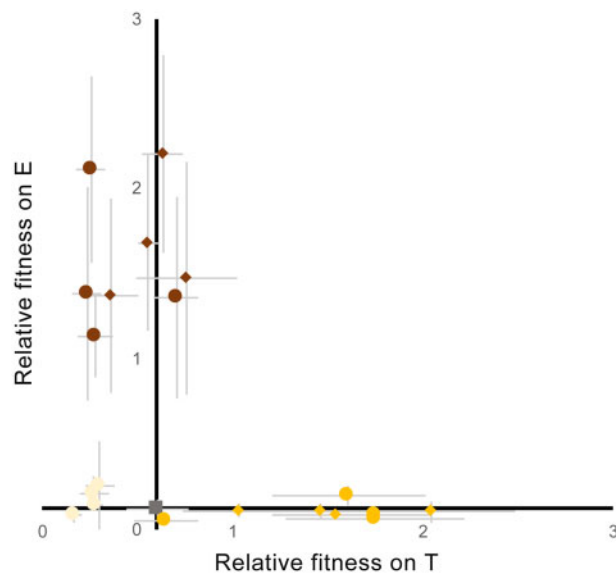


Figure 6. Relative fitness of populations on T and E. The lysates grown from the Day 0 generalist G ancestor is shown as the gray square, which serves as the origin for the axes (the scale of each axis is relative fitness compared to the appropriate common competitor). The Day 29 populations are diamonds and Day 30 populations are circles. G/P populations are beige, G/PT populations are orange, G/PE populations are maroon. Values are means of six replicates, error bars are standard deviations.

4. Discussion

Elevated mutation rates and large population sizes allow RNA viruses to maintain higher levels of population genetic diversity compared to prokaryotes and eukaryotes, ensuring an ample supply of variation for potential acquisition of and adaptation to new niches. For this reason, virologists often use viral genetic diversity as a measure of evolvability, virulence, and potential for emergence (Lauring and Andino 2010; Bordería, Stapleford, and Vignuzzi 2011; Pita and Roossinck 2013; Day 2015). With experimental evolution, we directly compared the effect of generalist genotype and generalist ecology on RNA virus population genetic diversity, and found that neither kind of generalism led to higher population genetic diversity. Although previous studies showed genetic diversity to positively correlate with host range size among divergent viral species sharing a common ancestor (Schneider and Roossinck 2000), we observed an expanded host range mutant of wild-type phi6 (the generalist) to have similar levels of genetic diversity compared to wild-type phi6 in the same passaging scheme. It was also unexpected that a generalist ecology did not lead to more genetic diversity, though others have noted that some ecological histories constrain organisms from finding the optimal genotype in variable environment (Jasmin and Kassen 2007). Experimental design involving comparing generalist and specialist ecology often focused on evolved population fitness, sometimes showing increased fitness in ecological generalist environment compared to ecological specialist environment with only speculative molecular basis for the phenotype (Alto et al. 2013) or showing fitness varying in heterogeneous environments without significant support in population genetic polymorphism (Morley et al. 2016). Both specialist and generalist viral populations are thought to be under continuous selective pressure for higher fitness, and to combat host defenses (Kassen 2002). However, generalist viruses are known to face more selective pressures compared to specialists as they infect different hosts with

different fitness optima (Elena, Agudelo-Romero, and Lalić 2009). As long as selection is present, it will purge genetic diversity within the population—barring relatively rare situations such as negative frequency dependent selection (Elena, Codoñer, and Sanjuán 2003). As phi6 is not capable of homologous recombination, diversity is reduced with each hard selective sweep (Amos and Harwood 1998; Wootton et al. 2002). Our most frequent SNPs (above 1%) in the alternating passage populations show signs of hard selective sweeps: two SNPs above 90 per cent in 2/4 G/PT lineages; six SNPs above 90 per cent in the G/PE lineages, and a commensurate loss of diversity elsewhere in the genome (Dennehy et al. 2013).

Our phenotypic screenings for genetic diversity were misleading compared to the sequencing data. We recently observed that host range mutations in P3 constrain further emergence onto host A (Zhao et al. 2019), but it was not known whether this epistatic constraint would relax with additional changes to the phi6 genome over evolutionary time (including other changes in P3, e.g. Q130R). Our results show that the mutational neighborhood available to the generalists is as narrow for the 30 day evolved populations as it is for their ancestor, which compromised the utility of the assay as a proxy for genetic variation. This is a reminder to assess phenotypes for background epistatic interactions when using as genotypic proxies. No similar epistatic constraint was observed in heat shock survival, but the G/PE populations showed the greatest variance when they were not the most diverse populations genetically.

In viral experimental evolution studies, it is common that viruses are exclusively passaged on novel hosts (Cuevas, Moya, and Sanjuán 2009; Ciota et al. 2014), which speeds up adaptive molecular evolution (Pepin, Domsic, and McKenna 2008; Wasik et al. 2016). When hosts are alternated, the typical lab host (or natural host) is seldom involved (Coffey and Vignuzzi 2011), and more commonly viruses alternate between two novel hosts (e.g. Turner and Elena 2000; Turner et al. 2010). These simplified experimental designs are informative, but they may be less applicable to real world scenarios of viral host use. Many studies of viral spillover and emergence suggest that passaging schemes involving both a novel host and the original, or reservoir, host are more realistic than having the pathogen exclusively replicate in the novel host (Allison et al. 2013; Troupin et al. 2016). This informed our experimental design, but this also likely explains why we saw fewer fixed substitutions than analogous studies that passaged viruses on exclusively novel hosts (Duffy, Burch, and Turner 2007; Turner et al. 2010).

Though host range mutations can have long-lasting epistatic effects (for instance, limiting emergence on host A), our results do not link genetic generalism overall to evolvability. Instead, ecological history appears to be the larger determinant of standing genetic diversity in fast-evolving RNA virus populations, with positive selection reducing population diversity. We observed several hard selective sweeps in the alternating passage populations, but there is a growing consensus that soft sweeps are common in viral populations, aiding standing genetic variation despite selective pressure (Wilson, Pennings, and Petrov 2017). The contrast between our work and the deduced soft sweeps in viral populations is expected: the effects of selection in one environment (such as serial host passaging) cannot be generalized to other environments and selective pressures. It has been shown that population genetic diversity becomes structured differently in different environments (Draghi and Wagner 2008), and it was experimentally shown in phi6 that different host ecologies selected and maintained different pools of genetic variation (Bono et al. 2015). Our experimental design allows for multiple

generations in a single host in daily growth, instead of switching hosts as frequently as every generation. This design mimics viral alternation of individual eukaryotic hosts or a phage switching between patches of bacteria, but a more frequent host alternation, such as is possible in a mixed host environment, may produce different results in terms of response to selection and standing genetic diversity (Bono et al. 2015).

Phi6 growth on its original host P had much reduced selection pressure compared to the alternating passages, and previous phage experimental evolution studies have linked lower evolvability to growth in higher selection environments (Pepin and Wichman 2008). Some groups eliminate this effect through pre-adaptation to novel hosts (Pepin, Domsic, and McKenna 2008; Pepin and Wichman 2008; Domingo-Calap, Cuevas, and Sanjuán 2009), but such pre-adaptation necessarily alters the starting genotype of the generalist. We wanted to test both the effect of genetic and ecological generalism, which was facilitated by using the un-pre-adapted host range mutant that was isogenic with the wild-type phi6 specialist. Since we saw no effect of genetic generalism, future work aiming to only examine ecological generalism might be well-served by pre-adapting viruses to all hosts that will be used. On the other hand, our experimental design mimics how viruses expand their host range in nature. It is unlikely that a generalist genotype, newly arisen through mutation, is near the peak of multiple host fitness landscapes, implying newly generalist genotypes experiencing novel generalist ecologies are likely always to experience stronger selection and lowered genetic diversity relative to specialist ecologies. The strong selection experienced by viruses in novel hosts may be in contrast to generalism in cellular organisms like bacteria and phytophagous insects. Bacteria often have operons to toggle larger metabolic networks, which allows the same genotype to have different proteins expressed when consuming a single or different combination of carbon sources (Dandekar et al. 2015). Similarly, generalist insects are not arising de novo in populations, but instead describe a long-established host range for species or sub-species (Ali and Agrawal 2012). Tests of the link between genetic diversity and generalism in viruses may not be applicable to cellular generalists, and it would be worthwhile to separately test the relationship between generalism on standing genetic diversity in cellular organisms.

Acknowledgements

We thank the three anonymous reviewers for their careful reading of the manuscript and useful comments. L.Z. also thanks Natasia Jacko, LaShanda Williams, and Chen Zhao for support and helpful advice at various stages of the project.

Funding

This work was supported by the National Science Foundation [grant number DEB 1453241] and the Graduate School of New Brunswick, Rutgers, The State University of New Jersey (award to L.Z.).

Data availability

All sequence data is publicly available in the sequence read archive SRR8440639-68.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

References

- Aguirre, J., Lázaro, E., and Manrubia, S. C. (2009) 'A Trade-off between Neutrality and Adaptability Limits the Optimization of Viral Quasispecies', *Journal of Theoretical Biology*, 261: 148–55.
- Ali, J. G., and Agrawal, A. A. (2012) 'Specialist Versus Generalist Insect Herbivores and Plant Defense', *Trends in Plant Science*, 17: 293–302.
- Allison, A. B. et al. (2013) 'Frequent Cross-Species Transmission of Parvoviruses among Diverse Carnivore Hosts', *Journal of Virology*, 87: 2342–7.
- Alto, B. W. et al. (2013) 'Stochastic Temperatures Impede RNA Virus Adaptation', *Evolution*, 67: 969–79.
- Amos, W., and Harwood, J. (1998) 'Factors Affecting Levels of Genetic Diversity in Natural Populations', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353: 177–86.
- Arenas, M., Lorenzo-Redondo, R., and Lopez-Galindez, C. (2016) 'Influence of Mutation and Recombination on HIV-1 In Vitro Fitness Recovery', *Molecular Phylogenetics and Evolution*, 94: 264–70.
- Bono, L. M. et al. (2015) 'Evolutionary Rescue and the Coexistence of Generalist and Specialist Competitors: An Experimental Test', *Proceedings of the Royal Society B: Biological Sciences*, 282: 20151932.
- Bordería, A. V., Stapleford, K. A., and Vignuzzi, M. (2011) 'RNA Virus Population Diversity: Implications for Inter-Species Transmission', *Current Opinion in Virology*, 1: 643–8.
- Buckling, A., Wills, M. A., and Colegrave, N. (2003) 'Adaptation Limits Diversification of Experimental Bacterial Populations', *Science*, 302: 2107–9.
- Burch, C. L., and Chao, L. (1999) 'Evolution by Small Steps and Rugged Landscapes in the RNA Virus Phi6', *Genetics*, 151: 921–7.
- Chao, L. (1990) 'Fitness of RNA Virus Decreased by Muller's Ratchet', *Nature*, 348: 454–5.
- , Tran, T. T., and Tran, T. T. (1997) 'The Advantage of Sex in the RNA Virus phi6', *Genetics*, 147: 953–9.
- Ciota, A. T. et al. (2014) 'Consequences of In Vitro Host Shift for St. Louis Encephalitis Virus', *Journal of General Virology*, 95: 1281–8.
- Coffey, L. L., and Vignuzzi, M. (2011) 'Host Alternation of Chikungunya Virus Increases Fitness While Restricting Population Diversity and Adaptability to Novel Selective Pressures', *Journal of Virology*, 85: 1025–35.
- Cuevas, J. M., Moya, A., and Sanjuán, R. (2009) 'A Genetic Background with Low Mutational Robustness Is Associated with Increased Adaptability to a Novel Host in an RNA Virus', *Journal of Evolutionary Biology*, 22: 2041–8.
- Dandekar, T. et al. (2015) 'Salmonella—How a Metabolic Generalist Adopts an Intracellular Lifestyle during Infection', *Frontiers in Cellular and Infection Microbiology*, 4: 191.
- Day, T. (2015) 'Information Entropy as a Measure of Genetic Diversity and Evolvability in Colonization', *Molecular Ecology*, 24: 2073–83.
- Debbink, K. et al. (2017) 'Vaccination Has Minimal Impact on the Intra-host Diversity of H3N2 Influenza Viruses', *PLoS Pathogens*, 13: e1006194.
- Dennehy, J. J. et al. (2013) 'Frequent Coinfection Reduces RNA Virus Population Genetic Diversity', *Journal of Heredity*, 104: 704–12.

- Dessau, M. et al. (2012) 'Selective Pressure Causes an RNA Virus to Trade Reproductive Fitness for Increased Structural and Thermal Stability of a Viral Enzyme', *PLoS Genetics*, 8: e1003102.
- Domingo-Calap, P., Cuevas, J. M., and Sanjuán, R. (2009) 'The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages', *PLoS Genetics*, 5: e1000742.
- Draghi, J., and Wagner, G. P. (2008) 'Evolution of Evolvability in a Developmental Model', *Evolution; International Journal of Organic Evolution*, 62: 301–15.
- Duffy, S., Burch, C. L., and Turner, P. E. (2007) 'Evolution of Host Specificity Drives Reproductive Isolation among RNA Viruses', *Evolution*, 61: 2614–22.
- , Turner, P. E., and Burch, C. L. (2006) 'Pleiotropic Costs of Niche Expansion in the RNA Bacteriophage $\phi 6$ ', *Genetics*, 172: 751–7.
- Dutta, R. N. et al. (2008) 'Rapid Adaptive Amplification of Preexisting Variation in an RNA Virus', *Journal of Virology*, 82: 4354–62.
- Elena, S. F. (2016) 'Evolutionary Transitions during RNA Virus Experimental Evolution', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371: 20150441.
- , Agudelo-Romero, P., and Lalić, J. (2009) 'The Evolution of Viruses in Multi-Host Fitness Landscapes', *The Open Virology Journal*, 3: 1–6.
- , Codoñer, F. M., and Sanjuán, R. (2003) 'Intraclonal Variation in RNA Viruses: Generation, Maintenance and Consequences', *Biological Journal of the Linnean Society*, 79: 17–26.
- Fusaro, A. et al. (2016) 'Unexpected Interfarm Transmission Dynamics during a Highly Pathogenic Avian Influenza Epidemic', 90: 6401–11.
- García-Arenal, F., and Fraile, A. (2013) 'Trade-Offs in Host Range Evolution of Plant Viruses', *Plant Pathology*, 62: 2–9.
- Hasing, M. E. et al. (2016) 'A Next Generation Sequencing-Based Method to Study the Intra-Host Genetic Diversity of Norovirus in Patients with Acute and Chronic Infection', *BMC Genomics*, 17: 1–11.
- Haydon, D. T., and Woolhouse, M. E. J. (1998) 'Immune Avoidance Strategies in RNA Viruses: Fitness Continuums Arising from Trade-Offs Between Immunogenicity and Antigenic Variability', *Journal of Theoretical Biology*, 193: 601–12.
- Hillung, J. et al. (2014) 'Experimental Evolution of an Emerging Plant Virus in Host Genotypes That Differ in Their Susceptibility to Infection', *Evolution*, 68: 2467–80.
- Jasmin, J.-N., and Kassen, R. (2007) 'Evolution of a Single Niche Specialist in Variable Environments', *Proceedings. Biological Sciences*, 274: 2761–7.
- Kassen, R. (2002) 'The Experimental Evolution of Specialists, Generalists, and the Maintenance of Diversity', *Journal of Evolutionary Biology*, 15: 173–90.
- Ketola, S. et al. (2013) 'No Evidence of Long-Term Benefits of Arthroscopic Acromioplasty in the Treatment of Shoulder Impingement Syndrome: Five-Year Results of a Randomised Controlled Trial', *Bone and Joint Research*, 2: 132–9.
- Koboldt, D. C. et al. (2012) 'VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing', *Genome Research*, 22: 568–76.
- Korboukh, V. K. et al. (2014) 'RNA Virus Population Diversity: An Optimum for Maximal Fitness and Virulence', *Journal of Biological Chemistry*, 289: 29531–44.
- Lauring, A. S., and Andino, R. (2010) 'Quasispecies Theory and the Behavior of RNA Viruses', *PLoS Pathogens*, 6: e1001005.
- Lequime, S. et al. (2016) 'Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-host Genetic Diversity in Mosquitoes', *PLoS Genetics*, 12: e1006111.
- Li, H. (2013) 'Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM', arXiv: 1303.3997.
- , and Durbin, R. (2010) 'Fast and Accurate Long-Read Alignment with Burrows–Wheeler Transform', *Bioinformatics*, 26: 589–95.
- et al. (2009) 'The Sequence Alignment/Map Format and SAMtools', *Bioinformatics*, 25: 2078–9.
- Li, S. et al. (2014) 'Specialist Versus Generalist Life Histories and Nucleotide Diversity in Caenorhabditis Nematodes', *Proceedings. Biological Sciences*, 281: 20132858.
- Lieberman, U., and Feldman, M. W. (1986) 'Modifiers of Mutation Rate: A General Reduction Principle', *Theoretical Population Biology*, 30: 125–42.
- Lythgoe, K., and Chao, L. (2003) 'Mechanisms of Coexistence of a Bacteria and a Bacteriophage in a Spatially Homogeneous Environment', *Ecology Letters*, 6: 326–34.
- McCrone, J. T., and Lauring, A. S. (2016) 'Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling', *Journal of Virology*, 90: 6884–95.
- McLeish, M. J., Fraile, A., and García-Arenal, F. (2018) 'Ecological Complexity in Plant Virus Host Range Evolution', Chap. 9, in C.M., Malmstrom (ed.) *Advances in Virus Research*, Vol. 101, pp. 293–339. Academic Press.
- Martin, M. (2011) 'Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads', *EMBnet. Journal*, 17: 10–12.
- Mas, A. et al. (2010) 'Unfinished Stories on Viral Quasispecies and Darwinian Views of Evolution', *Journal of Molecular Biology*, 397: 865–77.
- et al. (2004) 'Hepatitis C Virus Population Analysis of a Single-Source Nosocomial Outbreak Reveals an Inverse Correlation Between Viral Load and Quasispecies Complexity', *The Journal of General Virology*, 85: 3619–26.
- Morand, S., Manning, S. D., and Woolhouse, M. E. J. (1996) 'Parasite-Host Coevolution and Geographic Patterns of Parasite Infectivity and Host Susceptibility', *Proceedings: Biological Sciences*, 263: 119–28.
- Morley, V. J. et al. (2016) 'Evolution in Spatially Mixed Host Environments Increases Divergence for Evolved Fitness and Intrapopulation Genetic Diversity in RNA Viruses', *Virus Evolution*, 2: vev022.
- Nowak, M. A. et al. (1991) 'Antigenic Diversity Thresholds and the Development of AIDS', *Science*, 254: 963–9.
- Pepin, K. M., Domsic, J., and McKenna, R. (2008) 'Genomic Evolution in a Virus under Specific Selection for Host Recognition', *Infection, Genetics and Evolution*, 8: 825–34.
- , and Wichman, H. A. (2008) 'Experimental Evolution and Genome Sequencing Reveal Variation in Levels of Clonal Interference in Large Populations of Bacteriophage $\phi X174$ ', *BMC Evolutionary Biology*, 8: 85.
- Pita, J. S., and Roossinck, M. J. (2013) 'Mapping Viral Functional Domains for Genetic Diversity in Plants', *Journal of Virology*, 87: 790–7.
- R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Robinson, J. T. et al. (2011) 'Integrative Genomics Viewer', *Nature Biotechnology*, 29: 24–6.
- Saito, R., and Tojo, K. (2016) 'Comparing Spatial Patterns of Population Density, Biomass, and Genetic Diversity Patterns of the Habitat Generalist Mayfly *Isonychia japonica* Ulmer (Ephemeroptera: Isonychiidae) in the Chikuma–Shinano River Basin', *Freshwater Science*, 35: 724–37.

- Schneider, W. L., and Roossinck, M. J. (2000) 'Evolutionarily Related Sindbis-Like Plant Viruses Maintain Different Levels of Population Diversity in a Common Host', *Journal of Virology*, 74: 3130–4.
- Semancik, J. S., Vidaver, A. K., and Van Etten, J. L. (1973) 'Characterization of Segmented Double-Helical RNA from Bacteriophage Phi6', *Journal of Molecular Biology*, 78: 617–25.
- Troupin, C. et al. (2016) 'Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts', *PLoS Pathogens*, 12: e1006041.
- Turner, P. E., and Elena, S. F. (2000) 'Cost of Host Radiation in an RNA Virus', *Genetics*, 156: 1465–70.
- et al. (2010) 'Role of Evolved Host Breadth in the Initial Emergence of an RNA Virus', *Evolution; International Journal of Organic Evolution*, 64: 3273–86.
- Vidaver, A. K., Koski, R. K., and Van Etten, J. L. (1973) 'Bacteriophage Phi6: A Lipid-Containing Virus of *Pseudomonas Phaseolicola*', *Journal of Virology*, 11: 799–805.
- Wasik, B. R. et al. (2016) 'Generalized Selection to Overcome Innate Immunity Selects for Host Breadth in an RNA Virus', *Evolution*, 70: 270–81.
- Wilson, B. A., Pennings, P. S., and Petrov, D. A. (2017) 'Soft Selective Sweeps in Evolutionary Rescue', *Genetics*, 205: 1573–86.
- Woolhouse, M. E. J., and Gowtage-Sequeria, S. (2005) 'Host Range and Emerging and Reemerging Pathogens', *Emerging Infectious Diseases*, 11: 1842–7.
- Wootton, J. C. et al. (2002) 'Genetic Diversity and Chloroquine Selective Sweeps in *Plasmodium falciparum*', *Nature*, 418: 320–3.
- Zhao, L. et al. (2019) 'Existing Host Range Mutations Constrain Further Emergence of RNA Viruses', *Journal of Virology*, 93: e01385-18.