Viral coinfection is shaped by bacterial ecology and virus-virus interactions across diverse microbial taxa and environments

Samuel L. Díaz Muñoz[#]

Center for Genomics and Systems Biology
Department of Biology
New York University, New York, NY, USA

#Address correspondence to: Samuel L. Díaz Muñoz, sam.diazmunoz@nyu.edu

Abstract

Viral coinfection is a common across taxa and environments. Coinfection can enable genetic exchange, alter the dynamics of infections, and change the course of viral evolution. Despite the importance of coinfection to viral ecology and evolution, the factors that influence the frequency and extent of viral coinfection remain largely unexplored. Here I employ an extensive data set of virus-host interactions representing 6,564 microbial hosts and 13,103 viruses, to test the importance of bacterial traits and virus-virus interactions in shaping coinfection dynamics across a wide variety of taxa and environments. Using data from phage-host infection matrices, I found that bacterial ecology was the most important factor explaining variation (>28%) in the potential for coinfection. Realized (actual) coinfection was affected by bacterial defense mechanisms at the single-cell level. In a natural environment, the presence of CRISPR spacers in marine bacteria limited coinfections with active viruses by ~50%, despite the absence of spacer matches in any active infection. Analysis of viral infections mined from published bacterial and archaeal sequence data (n= 5,492 hosts), showed prophages limited coinfection of host cultures by other prophages, but not extrachromosomal viruses. At the single-cell level, prophages virtually eliminated coinfection. Virus-virus interactions also enhanced coinfection with culture coinfection by ssDNA and dsDNA viruses twice as likely to occur than ssDNA-only coinfections. Collectively, these results suggest bacterial ecology and virus-virus interactions are strong drivers of coinfection across different taxa and environments. These findings highlight that virus-virus interactions constitute an important selective pressure on viruses that is often underappreciated.

Introduction

Viruses outnumber hosts by a significant margin (Bergh *et al.*, 1989; Suttle, 2007; Weinbauer, 2004; Rohwer & Barott, 2012). In this situation, infection of more than one virus in a host (coinfection) might be expected to be a rather frequent occurrence potentially leading to virus-virus interactions (Bergh *et al.*, 1989; Díaz-Muñoz & Koskella, 2014; Suttle, 2007; Weinbauer, 2004; Rohwer & Barott, 2012). Across many different viral groups, virus-virus interactions within a host can alter genetic exchange (Worobey & Holmes, 1999), modify relative fitness (Refardt, 2011; Dropulić *et al.*, 1996), and change the course of viral evolution (Turner & Chao, 1998). Sustained within-host competition can lead to the evolution of viral strategies to compete against coinfecting viruses, such as frequency-dependent reproductive strategies (Turner & Chao, 1999; 2003) and adaptive lysis timing (Leggett *et al.*, 2013). Experimental studies provide evidence of the powerful consequences of virus-virus interactions for the fate of the host (Vignuzzi *et al.*, 2006; Li *et al.*, 2010; Abrahams *et al.*, 2009) and subsequent viral evolution (Ghedin *et al.*, 2005). Yet, there is little information regarding the ecological dimensions of coinfection and virus-virus interactions. This dearth of information may explain why virus-virus interactions are underappreciated as a selective force on viruses (DaPalma *et al.*, 2010). Given that most laboratory studies of viruses focus on a single virus at a time (DaPalma *et al.*, 2010), understanding the drivers and dynamics of coinfection and virus-virus interactions is a pressing frontier for viral ecology.

Recent studies of bacteriophages have started shedding light on the ecology of viral coinfection. In particular, mounting evidence indicates that many bacterial hosts can be infected by more than one phage (Koskella & Meaden, 2013; Flores *et al.*, 2013; 2011), suggesting there is potential for viral coinfection. Studies mining sequence data to uncover virus-host relationships have uncovered widespread coinfection in publicly available bacterial and archaeal genome sequence data (Roux *et al.*, 2015) and provided, for the first time, single-cell level information on viruses associated to specific hosts isolated from the environment in a culture-independent manner (Roux *et al.*, 2014; Labonté *et al.*, 2015). Collectively, these studies suggest that there is a large potential for

coinfection and that this potential is realized at both the host culture and single cell level. A summary of these studies suggests roughly half of hosts can be or are infected multiply, by an average of >2 viruses (Table 1). For the first time, there is extensive evidence across various methodologies, taxa, environments, and levels of coinfection that coinfection is widespread and virus-virus interactions may be a frequent occurrence.

Table 1. *Viral coinfection is prevalent across various methodologies, taxa, environments, and levels of coinfection.*

	Potential coinfection	Culture-level coinfection	Single-cell coinfection
Number of viruses in coinfections	4.89 (± 4.61)	3.377 ± 1.804	2.37 ± 0.83
Prop of bacteria with multiple infections	0.654	0.538	0.450
Reference	(Flores *et al.*, 2011)	(Roux *et al.*, 2015)	(Roux *et al.*, 2014)

Yet, if coinfection is a frequent occurrence in bacterial and archaeal hosts, what are the factors influencing coinfection patterns? What explains variation in this widespread phenomenon? To determine the frequency and extent of coinfection, there are two necessary conditions. First, as a necessary but not sufficient criterion, hosts must be able to be infected by the viruses independently, i.e. there must be a potential for coinfection. Studies of phage host range have provided a window into the potential for coinfection. For instance, in a single bacterial species there can be wide variation in phage host range (Holmfeldt *et al.*, 2007), and thus, the potential for coinfection. A larger scale, quantitative study of phage-bacteria infection networks suggests a continuum of potential coinfection, with some hosts susceptible to few viruses and others to many (Flores *et al.*, 2011). However, information regarding the ecological and biological correlates of coinfection at the strain or species levels remains elusive. At broad scales, geographic separation may play a role in potential coinfection (Flores *et al.*, 2013). Thus, bacterial ecology and identity are the primary candidates for drivers of potential for coinfection, at least in lytic viruses examined by phage-bacteria host range studies.

Second, the potential for coinfection is not always realized, so a second necessary condition is that both the bacteria and infecting viruses allow simultaneous or sequential infection. An extensive collection of studies provides insight into the bacterial and viral mechanisms that may affect coinfection. Bacteria, understandably reluctant to welcome

viruses, possess a collection of mechanisms of defense against viral infection, including restriction enzymes (Murray, 2002; Linn & Arber, 1968) and CRISPR-Cas systems (Horvath & Barrangou, 2010). The latter have been shown to be an adaptive immune system for bacteria, protecting from future infection by the same phage (Barrangou *et al.*, 2007) and preserving the memory of viral infections past (Held & Whitaker, 2009). Metagenomic studies of CRISPR in natural environments suggest rapid coevolution of CRISPR arrays (Tyson & Banfield, 2008), but little is known regarding *in-situ* protective effects of CRISPR on cells, which should now possible with single-cell genomics.

Viruses also have mechanisms to mediate infection by other viruses, some of which were identified in some of the earliest lab studies of bacteriophages (Ellis & Delbruck, 1939; Delbruck, 1946). An example of a well-described phenomenon of virus-virus interactions is superinfection immunity conferred by lysogens (Bertani, 1953), which can inhibit coinfection of cultures and single cells (Bertani, 1954). While this mechanism has been described in several species, its frequency at broader taxonomic scales and its occurrence in natural settings is not well known. Most attention in virus-virus interactions has focused on mechanisms limiting coinfection, with the assumption that coinfection invariably reduces host fitness (Berngruber *et al.*, 2010). However, some patterns of non-random coinfection suggest elevated coinfection (Dang *et al.*, 2004; Cicin-Sain *et al.*, 2005; Turner *et al.*, 1999) and there are viral mechanisms that promote co-infection (Joseph *et al.*, 2009). Systematic coinfection has been proposed (Roux *et al.*, 2012) to explain findings of chimeric viruses of mixed nucleic acids in metagenome reads (Diemer & Stedman, 2012; Roux *et al.*, 2013). This suspicion was confirmed in a study of marine bacteria that found highly non-random patterns of coinfection between ssDNA and dsDNA viruses in a lineage of marine bacteria (Roux *et al.*, 2014), but the frequency of this phenomenon across bacterial taxa remains to be uncovered. Thus, detailed molecular studies of coinfection dynamics and virome sequence data are generating questions ripe for testing across diverse taxa and environments.

Here I employ an extensive data set of virus-host interactions to test the importance of bacterial traits and virus-virus interactions in explaining coinfection dynamics and

patterns across a wide variety of taxa and environments. Specifically, I aim to answer the following questions:

1) How do bacterial traits and sampling conditions explain variation in estimates of potential coinfection (how many phages can infect hosts)?
2) Do prophages limit the scope of coinfection of host cultures?
3) Do ssDNA and dsDNA viruses show evidence of preferential coinfection?
4) Do prophages limit coinfection of single cells?
5) Does the CRISPR bacterial defense mechanism limit coinfection of single cells?

Results suggest that bacterial ecology and identity explain most of the variability in potential coinfection. Bacterial defense and virus-virus interactions were important mediators of coinfection dynamics. At the culture level integrated viruses limited coinfection by other prophages, but not extrachromosomal viruses, whereas CRISPR spacers and integrated viruses in single cells severely limited any further infection. However, systematic coinfection of host cultures by ssDNA and dsDNA viruses suggested virus-virus interactions can also promote coinfection.

Materials and Methods

Data Sets

I assembled data collectively representing 13,103 viral infections in 6,564 bacterial and archaeal hosts from diverse environments. These data are composed of three data sets that provide an increasingly fine-grained examination of coinfection from potential to realized coinfection at the culture (pure cultures or single colonies, not necessarily single cells) and single-cell levels.

The first data set is composed of bacteriophage host-range infection matrices documenting the results of experimental attempts at lytic infection in cultured phage and hosts (2011) and provides information on potential coinfection by compiling results from 38 published studies. The host-range infection data are matrices of infection success or failure via the “spot test”, briefly, a drop of phage lysate is “spotted” on a bacterial lawn and lysing of bacteria is noted as presence or absence. This data set represents studies

with varying sample compositions, in terms of bacteria and phage species, bacterial trophy, source of samples, bacterial association, and isolation habitat.

The second data set is derived from viral sequence information mined from published microbial genome sequence data on NCBI databases. Thus, this second data set provided information on actual (as opposed to potential) coinfection of cultures, including integrated and extrachromosomal viruses representing 12,498 viral infections in 5,492 bacterial and archaeal hosts. The data set includes data on viruses that are incorporated into the host genome (prophages) as well as extrachromosomal viruses detected in the genome assemblies (representing chronic, carrier state, and 'extrachromosomal prophage' infections). Genomes of microbial strains were primarily generated from colonies or pure cultures (except for 27 hosts known to be represented by single cells). Thus, although these data could represent coinfection potential coinfections at the single cell level, they are more conservatively regarded as culture coinfections.

The third data set included single-cell amplified genomes, providing information on coinfection and virus-virus interactions within single cells. This data set is composed of data from a study identifying viruses of 127 single cells of SUP05 marine bacteria in an oxygen minimum zone in the ocean (Roux *et al.*, 2014). These single-cell data represent a combined 143 viral infections including past and current (active) infections. The data set also identifies past infections (CRISPRs and prophages) and current infections (that is current at the time of isolation, e.g. ongoing lytic infections) in bacterial cells.

A list and description of data sources are included in Supplementary Table 1, and the raw data used in this paper are deposited in the FigShare data repository (FigShare doi:10.6084/m9.figshare.2072929).

Factors explaining potential coinfection

To test the potential influence of such factors on the estimate of phage infecting each host, I conducted a factorial analysis of variance (ANOVA). The independent (factorial) variables tested were the study type/source (natural, coevolution, artificial), bacterial taxon, habitat from which bacteria and phages were isolated, bacterial trophy

185 (photosynthetic and heterotrophic), and bacterial association (e.g. pathogen, free-living).
186 Geographic origin and phage taxa were present in the metadata, but were largely
187 incomplete; therefore they were not included in the analyses. Because the infection
188 matrices were derived from different studies testing varying numbers of phages, I
189 conducted ANOVA on the proportion of phage tested that infected a given host. The
190 same data were also analyzed using an arc-sine transform ANOVA and a binomial
191 logistic regression.

192 ***Effect of integrated prophages on coinfection of cultures***

193 To determine whether prophages affected the frequency and extent of coinfection in host
194 cultures, I examined host cultures infected exclusively by prophages or
195 extrachromosomal viruses (representing chronic, carrier state, and 'extrachromosomal
196 prophage' infections) and all prophage-infected cultures. I tested whether prophage-only
197 coinfections were infected by a different average number of viruses compared to
198 extrachromosomal-only using a Wilcoxon Rank Sum test. I tested whether prophage
199 infected cultures were more likely than not to be coinfections, and whether these
200 coinfections were more likely to occur with additional prophages or extrachromosomal
201 viruses.

202 ***Culture coinfection by ssDNA and dsDNA***

203 To examine whether ssDNA and dsDNA viruses exhibited non-random patterns of
204 culture coinfection, I compared the frequency of dsDNA-ssDNA mixed coinfections
205 against ssDNA-only coinfections among all host cultures coinfected with at least one
206 ssDNA virus, using a binomial test.

207 ***Effect of prophages and CRISPR on coinfection***

208 I investigated the effect of prophages and CRISPR spacers on coinfection in single cells,
209 by analyzing a data set of single amplified genomes (Roux *et al.*, 2014). This genomic
210 data set identified viral infections, including past, temperate and lytic infections in 67
211 SUP-05 bacteria (sulfur-oxidizing Gammaproteobacteria) isolated from different depths
212 of the oxygen minimum zone in the Saanich Inlet on Vancouver Island, British
213 Columbia.

214

To examine the effect of past infections on the frequency of cells undergoing current (active) infections, I examined cells that harbored putative defective prophages and CRISPR spacers in the genome and calculated the proportion of those that also had current (active) infections. To determine the extent of coinfection in cells with putative defective prophages and CRISPR spacers, I calculated the average number of active viruses infecting these cells. I examined differences in the frequency of current infection between bacteria with or without past infections using a proportion test and differences in the average amount of current viral infections using a Wilcoxon rank sum test.

Statistical Analyses

I conducted all statistical analyses in the R statistical programming environment (Team, 2011) and generated graphs using the ggpolt2 package. Means are presented as means ± standard deviation, unless otherwise stated. Data and code for analyses and figures are available in the Figshare data repository (doi: 10.6084/m9.figshare.2072929).

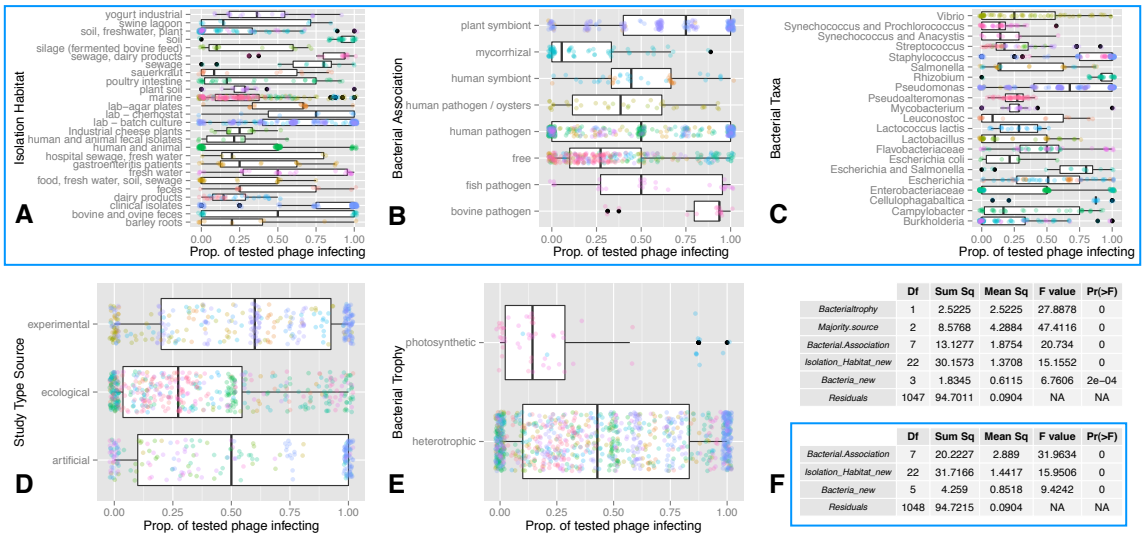
Results

Bacterial ecology and identity explain variation in potential for coinfection

To test whether sampling or bacterial and phage characteristics affected estimates of potential coinfection, I conducted an analysis of variance on 38 studies from the host-range data set for which these metadata were available. The full model explained 37.25% of the variance in potential for coinfection, with bacterial/phage ecology explaining >28% of the variance. The isolation habitat of phage and bacteria was the factor that explained the most variation (19.98%) between potential coinfection estimates. Microbes isolated from clinical isolates had the highest median potential coinfection, followed by sewage/dairy products, soil, sewage, and laboratory chemostats. All these habitats had more than 75% of tested phage infecting each host on average (Figure 1A). In absolute terms, the average host in each of these habitats could be infected by 3-15 different phages. Bacterial association explained 8.70% of the variance in potential coinfection. Bacteria that were pathogenic to cows had the highest potential coinfection followed by plant symbionts; both had more than 75% of tested phage infecting each bacteria (Figure 1B). In absolute terms, the average host that was a pathogenic to cows could be infected 15 phages on average, whereas plant symbionts could be infected by 3. The type of study

that served as the source of isolated strains explained 5.68% of variation, with coevolutionary studies and artificial pairings of laboratory strains (i.e. strains selected for their availability in laboratory stock collections) having higher potential coinfection than ecological studies (Figure 1D). Finally, the only other statistically significant factors explaining variation in potential coinfection were bacterial trophy (1.67%) and bacterial taxa (1.22%) and (Figure 1C, 1E). A reduced ANOVA model, generated by stepwise model selection with AIC, explained 30% of the variance with only three factors (Figure 1, outlined in blue): isolation habitat (21.02%), bacterial association (13.40%), and bacterial taxa (2.82%).

Model criticism suggested that ANOVA assumptions were reasonably met (Supplementary Figure 1) despite using proportion data (Warton & Hui, 2011). ANOVA on the arc-sine transform of the proportions and a binomial logistic regression provided qualitatively similar results (data not shown, see associated code in FigShare repository).



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bacterialtrophy	1	2.5225	2.5225	27.8878	0
Majority.source	2	8.5768	4.2884	47.4116	0
Bacterial.Association	7	13.1277	1.8754	20.734	0
Isolation_Habitat_new	22	30.1573	1.3708	15.1552	0
Bacteria_new	3	1.8345	0.6115	6.7606	2e–04
Residuals	1047	94.7011	0.0904	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bacterial.Association	7	20.2227	2.889	31.9634	0
Isolation_Habitat_new	22	31.7166	1.4417	15.9506	0
Bacteria_new	5	4.259	0.8518	9.4242	0
Residuals	1048	94.7215	0.0904	NA	NA

Figure 1. Bacterial-phage ecology and identity explain most of the variation in potential coinfection. Potential coinfection is the number of phages that can infect a bacterial host, here measured as the proportion of tested phages infecting each host (represented by points). Point colors correspond to hosts in the same study. Note data points are offset by a random and small amount (jittered) to enhance visibility and reduce overplotting. All factors explaining a statistically significant proportion of the variation in the full model are depicted (A-E). Those factors selected after stepwise model selection using AIC are indicated with a blue outline around the plot. ANOVA tables for the full and reduced (blue outline) models are presented in F.

Integrated viruses limit coinfection of host cultures by other prophages, but not extrachromosmal viruses

Virus-virus interactions can occur within cultures of cells, so I next determined whether integration of the virus into the host genome affected culture coinfection. To test whether prophage-infected host cultures reduced the probability of other viral infections, I examined all host cultures infected by prophages. A majority of prophage-infected host cultures were coinfections (56.48% of n = 3,134), a modest but statistically distinguishable increase over a 0.5 null (Binomial Exact: $p = 4.344e^{-13}$, Figure 2A inset). Of these coinfected host cultures (n=1770), cultures with more than one prophage (32.54%) were two times less frequent than those with prophages and extrachromosomal viruses (Binomial Exact: $p < 2.2e^{-16}$, Figure 2A). Therefore, integrated prophages appear to reduce the chance of the culture being infected with additional prophages, but not additional extrachromosomal viruses. Accordingly, host cultures co-infected exclusively by extrachromosomal viruses (n=675) were infected by 3.25 ± 1.34 viruses, compared to 2.54 ± 1.02 prophages (n=575); these quantities showed a statistically significant difference (Wilcoxon Rank Sum: W = 125398.5, $p < 2.2e^{-16}$, Figure 2B).

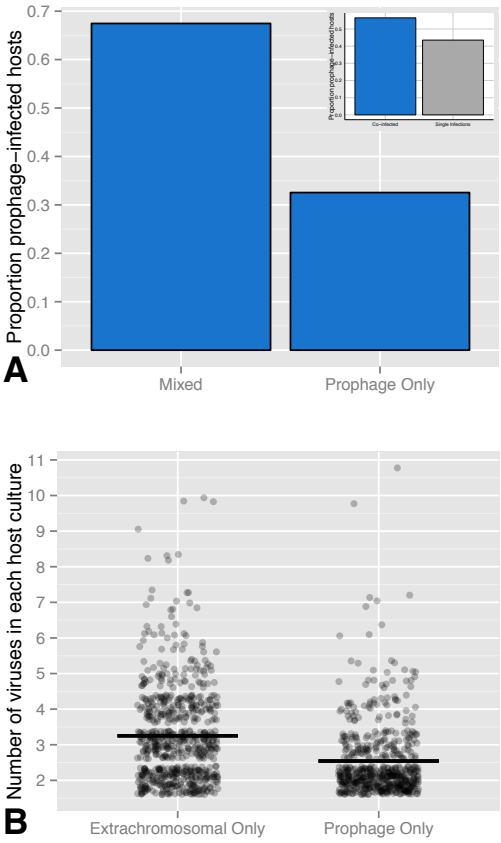


Figure 2. Host cultures infected with prophages limit coinfection by other prophages, but not extrachromosomal viruses. A slight, but statistically significant, majority of prophage-infected host cultures were coinfected (A-inset). Of these, host cultures containing multiple prophages were less frequent than those containing prophages and extrachromosomal (e.g. chronic, carrier state) infections (A). On average (black horizontal bars), extrachromosomal-only coinfections involved more viruses than prophage-only coinfections (B).

Non-random coinfection of host cultures by ssDNA and dsDNA viruses suggests mechanisms enhancing coinfection

To determine whether non-random coinfection patterns could increase the likelihood of coinfection, I tested patterns of ssDNA and dsDNA infection in the hosts using data from viruses found in NCBI bacterial and archaeal genome sequence data. Coinfected host cultures containing ssDNA viruses (n = 331), were more likely to have dsDNA or

unclassified viruses (70.69%), than multiple ssDNA infections (exact binomial: p= $3.314e^{-14}$). These coinfections were >2 times more likely to involve at least one dsDNA viruses than none (exact binomial: p = $2.559e^{-11}$, Figure 3).

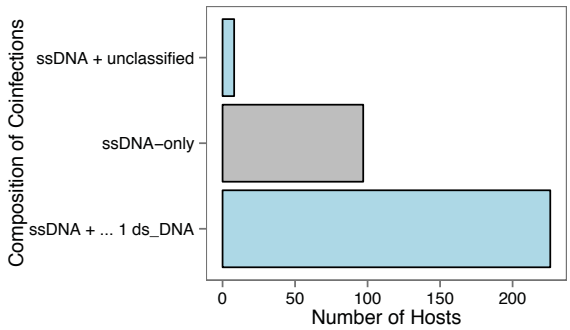


Figure 3. Culture coinfections between ssDNA-dsDNA viruses are more common than expected by chance. Host cultures with dsDNA-ssDNA coinfections, were more frequent than ssDNA-only coinfections (shaded in gray).

Putative defective prophages severely limit coinfection at single cell level

To test whether past infections integrated into the host genome could affect coinfection of single cells in a natural environment, I examined a single cell amplified genomics data set of SUP05 marine bacteria. Cells with putative defective prophages were less likely to have current infections: 9.09% of cells with defective prophages had current infections, compared to 74.55% of cells that had no prophages (X-squared = 14.2607, df = 1, p-value = 0.00015). Bacteria with defective prophages were currently infected by an average of 0.09 ± 0.30 phages, whereas bacteria without prophages were infected by 1.22 ± 1.05 phages (Figure 4A). No host with defective prophages had current coinfections (i.e., > 1 active virus infection).

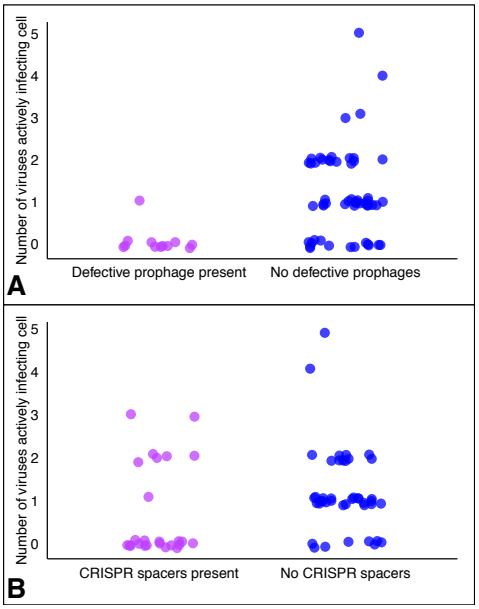


Figure 4. Past viral infections limit coinfections of single cells of SUP-05 marine bacteria. Each point represents a bacterial cell; the number of current (active) infections is depicted on the y-axis. A) Prophage-infected cells (left) are less likely to have current infections (note number of points on either side) and are infected by fewer viruses that cells not infected with prophages (right, note higher average number of active viruses). B) Cells with CRISPR spacers are less likely to have current infections and are infected by fewer viruses (left), that cells without CRISPR spacers (right).

CRISPR spacers limit coinfection at single cell level without spacer matches

In SUP-05 marine bacteria, cells with CRISPR spacers were also less likely to have current viral infections as those without spacers (X-squared = 14.0308, df = 1, p-value = 0.00018). The effects of CRISPR were more

moderate than defective prophages, with 32.00% of bacteria with CRISPRs having current viral infections, compared to 80.95% percent of bacteria without CRISPR spacers. Bacteria with CRISPR spacers had 0.68 ± 1.07 current phage infections compared to 1.21 ± 1.00 for those without spacers (Figure 4B). In contrast to putative defective prophages, hosts with CRISPR spacers could have current infections and coinfections with up to 3 phages.

Discussion

Summary of findings

The results of this study provide both a broad scale and a fine-grained examination of the bacterial and viral factors affecting coinfection dynamics. Across a broad range of taxa and environments, I found evidence for the importance of bacteria-phage ecology and bacterial identity in shaping potential coinfection. Across an even broader range of taxa, results suggest that prophages limit coinfection of cultures by other prophages, but not by extrachromosomal viruses, providing the most comprehensive test of the phenomenon of superinfection immunity conferred by lysogens. Conversely, I found evidence of increased culture coinfection by ssDNA and dsDNA phages, suggesting mechanisms that may enhance coinfection. At a fine-scale, single cell data enabled testing of the effects of prophages and CRISPR spacers on coinfection in a natural environment. In light of the increasing awareness of the widespread occurrence of viral coinfection, this study provides a foundation for future work on the frequency, mechanisms, and dynamics of viral coinfection and its ecological and evolutionary consequences.

Bacterial correlates of coinfection

Analysis of the factors influencing the potential for coinfection pointed to the importance of microbial ecology, with isolation habitat of the bacteria/phage pairings explaining the most variation. This finding suggests that the diverse and complex patterns of bacterial susceptibility and phage observed at the scale of bacterial species (Holmfeldt *et al.*, 2007), may be best explained by local ecological factors. Bacterial taxon was the only other factor identified in the most reduced model, but explained little of the variation in potential coinfection. This could be because the bacteria represented in the host-range data set were varied predominantly at the strain or species level (Flores *et al.*, 2011),

whereas infection patterns are less variable at higher taxonomic scales levels (Flores *et al.*, 2013; Roux *et al.*, 2015). Notably, the type of study strains were selected (i.e. coevolution experiments, lab strain stocks, or the environment) explained a modest proportion of variation in the potential for coinfection, suggesting that viral coinfection is just as important in laboratory settings as in the environment.

Although all tested factors showed significant statistical associations with the potential for coinfection, there was still substantial unexplained variation (~65%) in the potential for coinfection. Thus, these factors should be regarded as starting points for future experimental examinations. Some of the tested factors were only represented by one study, limiting the generality of inferences. Moreover, other factors not examined, such as geography and bacterial phylogeny, could plausibly affect the potential for viral coinfection. First, the geographic origin of strains can affect infection specificity such that bacteria isolated from one location are likely to be infected by more phage isolated from the same location. This result was observed in a study analyzing a spatially explicit host-phage infection matrix of marine microbes (Flores *et al.*, 2013). This pattern could be due to the influence of local adaptation of phages to their hosts (Koskella *et al.*, 2011) and represents an interesting avenue for further research. Second, phylogenetic patterns are another possible factor influencing potential coinfection. A visual inspection of published host infection matrices with phylogenies (Koskella & Meaden, 2013; Liu *et al.*, 2015) suggests that particular bacterial lineages can exhibit dramatic differences in the number of phages able to infect. Further studies with detailed phylogenetic, spatial, and infection information will be necessary to test the influence of bacterial evolution and ecology on the potential for coinfection. More importantly, the importance of these factors in explaining variation in actual coinfections should be determined, as the potential for coinfection is not necessarily realized.

Bacterial defense was another important factor influencing coinfection patterns. At a single cell level, the presence of CRISPR spacers reduced the extent of active viral infections, even though these spacers matched none of the infecting viruses identified (Roux *et al.*, 2014). These results provide the first evidence from a natural environment

389 that CRISPR's protective effects extend beyond viruses with exact matches to the
390 particular spacers within the cell (Fineran *et al.*, 2014; Semenova *et al.*, 2011). Although
391 very specific sequence matching is thought to be required for CRISPR-Cas-based
392 immunity (Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Mojica *et al.*, 2005), the system
393 can tolerate mismatches in protospacers (within and outside of the seed region: Semenova
394 *et al.*, 2011; Fineran *et al.*, 2014) enabling protection (interference) against related phages
395 by a mechanism of enhanced spacer acquisition termed priming (Fineran *et al.*, 2014).
396 The seemingly broader protective effect of CRISPR-Cas beyond specific sequences may
397 help explain continuing effectiveness of CRISPR-Cas (Fineran *et al.*, 2014) in the face of
398 rapid viral coevolution for escape (Heidelberg *et al.*, 2009; Andersson & Banfield, 2008;
399 Tyson & Banfield, 2008).

400 ### *The role of virus-virus interactions in coinfection*

401 The results suggest that virus-virus interactions play a role in limiting *and* enhancing
402 coinfection of cultures and single cells. At the level of host cultures prophages limited
403 coinfection of host cultures by other prophages, but not extrachromosomal viruses. As
404 these were culture coinfections and not necessarily single-cell coinfections, these results
405 are consistent with a single-cell study of *Salmonella* cultures showing that lysogens can
406 activate cell subpopulations that are transiently immune from viral infection (Cenens *et*
407 *al.*, 2015). Prophages, more specifically, putative defective prophages, had a more
408 dramatic impact at the single-cell level in SUP05 marine bacteria, severely limiting active
409 viral infection and virtually excluding coinfection, extending findings of laboratory
410 studies to the natural environment. The results on culture-level and single-cell coinfection
411 come from very different data sets, which should be examined carefully before drawing
412 general patterns. First, the culture-level data set is composed of an analysis of all publicly
413 available bacterial and archaeal genome sequences in NCBI databases. These sequences
414 show a bias towards particular taxonomic groups (e.g. model study species) and those
415 that are easy to grow in pure culture. However, in some ways this limitation makes the
416 data set more remarkable, as these host cultures were presumably assumed to be virus-
417 free (i.e. pure cultures). The single cell data set is limited to just one host type isolated in
418 a particular environment, as opposed to the 5,444 hosts in the NCBI data set. This
419 limitation prohibits taxonomic generalizations about the effects on prophages on single

cells, but generalizes laboratory findings to a natural environment. Additionally, the prophages in the single cell study were termed 'putative defective prophages' (Roux *et al.*, 2014), which could mean that bacterial domestication of phage functions (Bobay *et al.*, 2014; Asadulghani *et al.*, 2009), rather than phage-phage interactions in a strict sense, would explain protection from infection in these single cells. In view of these current limitations, a wider taxonomic and ecological range of culture and single-cell sequence data should elucidate the role of lysogenic viruses in affecting coinfection dynamics. Interactions in coinfection between temperate bacteriophages can affect viral fitness (Refardt, 2011), suggesting latent infections are a profitable avenue for future research on virus-virus interactions.

Other virus-virus interactions examined in this study, appeared to increase the chance of coinfection. While, prophages strongly limited coinfection in single cells, Roux et al.'s (2014) original analysis of this same data set found strong evidence of enhanced coinfection (i.e. higher than expected by random chance) between dsDNA and ssDNA Microviridae phages in SUP05_03 lineage. I extend the taxonomic applicability of this result by providing evidence that ssDNA-dsDNA culture coinfections occur more frequently than would be expected by chance. Thus, enhanced coinfection, perhaps due to the long replicative cycle of some ssDNA viruses (e.g. Innoviridae: Rakonjac *et al.*, 2011), is a major factor explaining findings of phages with chimeric genomes composed of different types of nucleic acids (Roux *et al.*, 2013; Diemer & Stedman, 2012). Collectively, these results highlight the importance of virus-virus interactions as part of the suite of evolved viral strategies to mediate frequent interactions with other viruses, from limiting to promoting coinfection depending on the evolutionary and ecological context (Turner & Duffy, 2009).

Implications and applications

Collectively, these results suggest bacterial ecology and virus-virus interactions are important drivers of the widespread phenomenon of viral coinfection. An important implication is that virus-virus interactions will constitute an important selective pressure on viral evolution. The importance of virus-virus interactions may have been underappreciated because of an overestimation of the importance of superinfection

exclusion (Dulbecco, 1952). Paradoxically, superinfection avoidance may actually highlight the selective force of virus-virus interactions. In an evolutionary sense, this viral trait exists precisely because the potential for coinfection is high. If this is correct, then the variability in the potential for coinfection, as found in this study, suggests that the manifestation of superinfection exclusion will vary across viral groups according to their ecological context. Accordingly, some viral mechanisms will promote coinfection, as found in this study with ssDNA/dsDNA coinfections and other studies (Dang *et al.*, 2004; Cicin-Sain *et al.*, 2005; Turner *et al.*, 1999). I found substantial variation in potential coinfection, suggesting that the selective pressure for coinfection is going to vary across local ecologies. This is in agreement with observations of variation in viral genetic exchange (which requires coinfection) rates across different geographic localities in a variety of viruses (Díaz-Muñoz *et al.*, 2013; Trifonov *et al.*, 2009; Held & Whitaker, 2009).

These results have clear implications, not only for the study of viral ecology in general, but for practical biomedical and agricultural applications of phages and bacteria. Phage therapy is often predicated on the specificity of phages on a particular bacterium, but intentional coinfection could be an important part of the arsenal as part of combined or cocktail phage therapy. This study also suggests that viral coinfection in the microbiome should be examined, as part of the influence of the virome as an integral part of the microbiome (Pride *et al.*, 2012; Minot *et al.*, 2011; Reyes *et al.*, 2010). Finally, if these results apply in eukaryotic viruses, as evidence suggests (DaPalma *et al.*, 2010), variation in viral coinfection rates should be considered in the context of treating and preventing infections, as coinfection likely represents the default condition of human hosts (Wylie *et al.*, 2014). Coinfection and virus-virus interactions have been implicated in changing disease outcomes for hosts (Vignuzzi *et al.*, 2006), altering epidemiology of viral diseases (Nelson *et al.*, 2008), and impacting antimicrobial therapies (Birger *et al.*, 2015). In sum, the results of this study suggest that the ecological context, mechanisms, and evolutionary consequences of virus-virus interactions should be considered as an important subfield in the study of viruses.

Acknowledgements

Simon Roux kindly provided extensive assistance with previously published data sets. I am indebted to Joshua Weitz and Britt Koskella for providing helpful critiques and advice on an earlier version of this manuscript. A Faculty Fellowship to SLDM from New York University supported this work.

References

Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, *et al.* (2009). Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants. *J Virol* **83**:3556–3567.

Andersson AF, Banfield JF. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**:1047–1050.

Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, *et al.* (2009). The defective prophage pool of Escherichia coli O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *Plos Pathog* **5**:e1000408.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, *et al.* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**:1709–1712.

Bergh Ø, Børsheim KY, Bratbak G, Heldal M. (1989). High abundance of viruses found in aquatic environments. *Nature* **340**:467–468.

Berngruber TW, Weissing FJ, Gandon S. (2010). Inhibition of superinfection and the evolution of viral latency. *J Virol* **84**:10200–10208.

Bertani G. (1953). Lysogenic versus lytic cycle of phage multiplication. *Cold Spring Harbor Symposia on Quantitative Biology* **18**:65–70.

Bertani G. (1954). Studies on lysogenesis. III. Superinfection of lysogenic Shigella dysenteriae with temperate mutants of the carried phage. *J Bacteriol* **67**:696–707.

Birger RB, Kouyos RD, Cohen T, Griffiths EC, Huijben S, Mina M, *et al.* (2015). The potential impact of coinfection on antimicrobial chemotherapy and drug resistance. *Trends Microbiol* **23**:537–544.

Bobay L-M, Touchon M, Rocha EPC. (2014). Pervasive domestication of defective prophages by bacteria. *P Natl Acad Sci-Biol* **111**:12127–12132.

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, *et al.* (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**:960–

964.

Cenens W, Makumi A, Govers SK, Lavigne R, Aertsen A. (2015). Viral Transmission Dynamics at Single-Cell Resolution Reveal Transiently Immune Subpopulations Caused by a Carrier State Association. *PLoS Genet* **11**:e1005770.

Cicin-Sain L, Podlech J, Messerle M, Reddehase MJ, Koszinowski UH. (2005). Frequent Coinfection of Cells Explains Functional In Vivo Complementation between Cytomegalovirus Variants in the Multiply Infected Host. *J Virol* **79**:9492–9502.

Dang Q, Chen JB, Unutmaz D, Coffin JM, Pathak VK, Powell D, *et al.* (2004). Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. *Proc Natl Acad Sci U S A* **101**:632–637.

DaPalma T, Doonan BP, Trager NM, Kasman LM. (2010). A systematic approach to virus-virus interactions. *Virus Research* **149**:1–9.

Delbruck M. (1946). Bacterial viruses or bacteriophages. *Biological Reviews* **21**:30–40.

Diemer GS, Stedman KM. (2012). A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biology Direct* **7**:13.

Díaz-Muñoz SL, Koskella B. (2014). Bacteria–Phage Interactions in Natural Environments. *Advances in Applied Microbiology* **89**:135–183.

Díaz-Muñoz SL, Tenaillon O, Goldhill D, Brao K, Turner PE, Chao L. (2013). Electrophoretic mobility confirms reassortment bias among geographic isolates of segmented RNA phages. *BMC Evol Biol* **13**:206.

Dropulić B, Hěrmánková M, Pitha PM. (1996). A conditionally replicating HIV-1 vector interferes with wild-type HIV-1 replication and spread. *Proc Natl Acad Sci U S A* **93**:11103–11108.

Dulbecco R. (1952). Mutual exclusion between related phages. *J Bacteriol* **63**:209–217.

Ellis EL, Delbruck M. (1939). The growth of bacteriophage. *J Gen Physiol* **22**:365–384.

Fineran PC, Gerritzen MJH, Suárez-Diez M, Künne T, Boekhorst J, van Hijum SAFT, *et al.* (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *P Natl Acad Sci-Biol* **111**:E1629–38.

Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. (2011). Statistical structure of host–phage interactions. *Proc Natl Acad Sci U S A* **108**:E288–E297.

Flores CO, Valverde S, Weitz JS. (2013). Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal* **7**:520–532.

Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, *et al.* (2005). Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**:1162–1166.

Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. (2009). Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *Plos One*.

Held NL, Whitaker RJ. (2009). Viral biogeography revealed by signatures in Sulfolobus islandicusgenomes. *Environmental Microbiology* **11**:457–466.

Holmfeldt K, Middelboe M, Nybroe O, Riemann L. (2007). Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. *Applied and Environmental Microbiology* **73**:6730–6739.

Horvath P, Barrangou R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**:167–170.

Joseph SB, Hanley KA, Chao L, Burch CL. (2009). Coinfection rates in phi 6 bacteriophage are enhanced by virus-induced changes in host cells. *Evol Appl* **2**:24–31.

Koskella B, Meaden S. (2013). Understanding bacteriophage specificity in natural microbial communities. *Viruses* **5**:806–823.

Koskella B, Thompson JN, Preston GM, Buckling A. (2011). Local Biotic Environment Shapes the Spatial Scale of Bacteriophage Adaptation to Bacteria. *The American Naturalist* **177**:440–451.

Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, *et al.* (2015). Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *The ISME Journal* **9**:2386–2399.

Leggett HC, Benmayor R, Hodgson DJ, Buckling A. (2013). Experimental evolution of adaptive phenotypic plasticity in a parasite. *Current Biology* **23**:139–142.

Li C, Hatta M, Nidom CA, Muramoto Y, Watanabe S, Neumann G, *et al.* (2010). Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence. *P Natl Acad Sci-Biol* **107**:4687–4692.

Linn S, Arber W. (1968). Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. *Proc Natl Acad Sci U S A* **59**:1300–1306.

Liu J, Yan R, Zhong Q, Ngo S, Bangayan NJ, Nguyen L, *et al.* (2015). The diversity and host interactions of Propionibacterium acnes bacteriophages on human skin. *The ISME Journal*.

Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, *et al.* (2011). The human gut

virome: Inter-individual variation and dynamic response to diet. *Genome Research* **21**:1616–1625.

Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**:174–182.

Murray NE. (2002). 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. *Microbiology (Reading, Engl)* **148**:3–20.

Nelson MI, Viboud C, Simonsen L, Bennett RT, Griesemer SB, St George K, *et al.* (2008). Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza A Virus Since 1918 Kawaoka, Y (ed). *Plos Pathog* **4**:e1000012.

Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA, *et al.* (2012). Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *The ISME Journal* **6**:915–926.

Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M. (2011). Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr Issues Mol Biol* **13**:51–76.

Refardt D. (2011). Within-host competition determines reproductive success of temperate bacteriophages. *The ISME Journal* **5**:1451–1460.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, *et al.* (2010). nature09199. *Nature* **466**:334–338.

Rohwer F, Barott K. (2012). Viral information. *Biol Philos* **28**:283–297.

Roux S, Enault F, Bronner G, Vaulot D, Forterre P, Krupovic M. (2013). Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nature Communications* **4**:2700.

Roux S, Hallam SJ, Woyke T, Sullivan MB. (2015). Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife* **4**.

Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, *et al.* (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta- genomics. *eLife* **3**.

Roux S, Krupovic M, Poulet A, Debroas D, Enault F. (2012). Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. *Plos One* **7**:e40418.

Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, *et al.* (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *P Natl Acad Sci-Biol* **108**:10098–10103.

Suttle CA. (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology* **5**:801–812.

Team RDC. (2011). R: A Language and Environment for Statistical Computing.

Trifonov V, Khiabanian H, Rabadan R. (2009). Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. *N Engl J Med* **361**:115–119.

Turner P, Burch C, Hanley K, Chao L. (1999). Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage phi 6. *J Virol* **73**:2420–2424.

Turner P, Chao L. (2003). Escape from Prisoner's Dilemma in RNA phage phi 6. *Am Nat* **161**:497–505.

Turner P, Chao L. (1999). Prisoner's dilemma in an RNA virus. *Nature* **398**:441–443.

Turner P, Chao L. (1998). Sex and the evolution of intrahost competition in RNA virus phi 6. *Genetics* **150**:523–532.

Turner PE, Duffy S. (2009). Evolutionary ecology of multiple phage adsorption and infection. In:*Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses*, Abedon, ST (ed), Cambridge University Press: Cambridge.

Tyson GW, Banfield JF. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology* **10**:200–207.

Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**:344–348.

Warton DI, Hui FKC. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology* **92**:3–10.

Weinbauer MG. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**:127–181.

Worobey M, Holmes EC. (1999). Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* **80 (Pt 10)**:2535–2543.

Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM. (2014). Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biol* **12**:71.