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

Running head: Drivers of viral coinfection

Samuel L. Díaz Muñoz#
Center for Genomics and Systems Biology + Department of Biology
12 Waverly Place
New York University, New York, NY, USA 10003

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

Abstract

Infection of more than one virus in a host, coinfection, is common across taxa and environments. Viral coinfection can enable genetic exchange, alter the dynamics of infections, and change the course of viral evolution. Yet, the factors influencing the frequency and extent of viral coinfection remain largely unexplored. Here, employing three microbial data sets of virus-host interactions covering cross-infectivity, culture coinfection, and single-cell coinfection (total: 6,564 microbial hosts, 13,103 viruses), I found evidence that ecology and virus-virus interactions are recurrent factors shaping coinfection patterns. Host ecology was a consistent and strong predictor of coinfection across all three datasets: potential, culture, and single-cell coinfection. Host phylogeny or taxonomy was a less consistent predictor, being weak or absent in potential and single-cell coinfection models, yet it was the strongest predictor in the culture coinfection model. Virus-virus interactions strongly affected coinfection. In the largest test of superinfection exclusion to date, prophage infection reduced culture coinfection by other prophages, with a weaker effect on extrachromosomal virus coinfection. At the single-cell level, prophages eliminated coinfection. Virus-virus interactions also *increased* culture coinfection with ssDNA-dsDNA coinfections >2x more likely than ssDNA-only coinfections. Bacterial defense limited single-cell coinfection in marine bacteria CRISPR spacers reduced coinfections by ~50%,

despite the absence of spacer matches in any active infection. Collectively, these results suggest the environment bacteria inhabit and the interactions among surrounding viruses are two factors consistently shaping viral coinfection patterns. These findings highlight the role of virus-virus interactions in coinfection with implications for phage therapy, microbiome dynamics, and viral infection treatments.

Introduction

Viruses outnumber hosts by a significant margin (Bergh *et al.*, 1989; Suttle, 2007; Weinbauer, 2004; Rohwer and Barott, 2012; Wigington *et al.*, 2016). In this situation, infection of more than one strain or type of virus in a host (coinfection) might be expected to be a rather frequent occurrence potentially leading to virus-virus interactions (Díaz-Muñoz and Koskella, 2014; Bergh *et al.*, 1989; Rohwer and Barott, 2012; Suttle, 2007; Weinbauer, 2004). Across many different viral groups, virus-virus interactions within a host can alter genetic exchange (Worobey and Holmes, 1999), modify viral evolution (Refardt, 2011; Roux *et al.*, 2015; Dropulić *et al.*, 1996; Ghedin *et al.*, 2005; Turner and Chao, 1998), and change the fate of the host (Vignuzzi *et al.*, 2006; Li *et al.*, 2010; Abrahams *et al.*, 2009). Yet, there is little information regarding the ecological dimensions of coinfection and virus-virus interactions. Given that most laboratory studies of viruses focus on a single virus at a time (DaPalma *et al.*, 2010), understanding the drivers of coinfection and virus-virus interactions is a pressing frontier for viral ecology.

Coinfection can be assessed at different scales and with different methods. At the broadest scale, the ability of two or more viruses to independently infect the same host –cross-infectivity– is a necessary but not sufficient criterion to determine coinfection. Thus, cross-infectivity or the potential for coinfection, represents a baseline shaping coinfection patterns. At increasingly finer

scales, coinfection can refer to >1 virus infecting the same multicellular host (e.g., a multicellular eukaryote or a colony of bacterial cells, here called culture coinfection), or to the infection of a single cell. Each of these measures of coinfection may be estimated using different methods that jointly provide a comprehensive picture of the ecology of viral coinfection.

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 type of phage (Koskella and Meaden, 2013; Flores *et al.*, 2013; 2011). 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 and Suttle, 2013). 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 have the potential to be infected (i.e. are cross-infective) or are actually infected by an average of >2 viruses (Table 1). Thus, there is extensive evidence across various methodologies, taxa, environments, 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.*

	Cross-infectivity (*potential* coinfection)	Culture-level coinfection	Single-cell coinfection
Number of viruses	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 explaining variation in this widespread phenomenon? The literature suggests four factors that are

likely to play a role: host ecology, host taxonomy or phylogeny, host defense mechanisms, and virus-virus interactions. The relevance and importance of these are likely to vary for cross-infectivity, culture coinfection, and single-cell coinfection.

Cross-infectivity has been examined in studies of phage host-range, which have provided some insight into the factors affecting potential coinfection. 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, quantitative study of phage-bacteria infection networks in multiple taxa also found wide variation in cross-infectivity in narrow taxonomic ranges (strain or species level), with some hosts susceptible to few viruses and others to many (Flores *et al.*, 2011). However, at broader host taxonomic scales, cross-infectivity followed a modular pattern, suggesting that higher taxonomic ranks could influence cross-infectivity (Flores *et al.*, 2013). Thus, host taxonomy may show a scale dependent effect in shaping coinfection patterns. Flores et al. (2013) also highlighted the potential role of ecology in shaping coinfection patterns as geographic separation played a role in cross-infectivity. Thus, bacterial ecology and phylogeny (particularly at taxonomic ranks above species) are the primary candidates for drivers of coinfection.

Culture and single cell coinfection, require cross-infectivity, but also require both the bacteria and infecting viruses to allow simultaneous or sequential infection. Thus, in addition to bacterial phylogeny and ecology, we can expect additional factors shaping coinfection, namely bacterial defense and virus-virus interactions. An extensive collection of studies provides evidence that bacterial and viral mechanisms may affect coinfection. Bacteria, understandably reluctant to

welcome viruses, possess a collection of mechanisms of defense against viral infection (Labrie *et al.*, 2010), including restriction enzymes (Murray, 2002; Linn and Arber, 1968) and CRISPR-Cas systems (Horvath and 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 and Whitaker, 2009). Metagenomic studies of CRISPR in natural environments suggest rapid coevolution of CRISPR arrays (Tyson and Banfield, 2008), but little is known regarding the *in-situ* protective effects of CRISPR on cells, which should now be possible with single-cell genomics.

Viruses also have mechanisms to mediate infection by other viruses, some of which were identified in early lab studies of bacteriophages (Ellis and Delbruck, 1939; Delbruck, 1946). An example of a well-described phenomenon of virus-virus interactions is superinfection immunity conferred by lysogenic viruses (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 specific viral mechanisms that promote co-infection have been identified (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 and 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 large virus-host interaction data sets to date to provide a comprehensive picture of the factors shaping coinfection (from potential coinfection, to culture and single-cell coinfection) by answering the following questions:

1) How do ecology, bacterial phylogeny/taxonomy, bacterial defense mechanisms, and virus-virus interactions explain variation in estimates of viral coinfection?
2) What is the relative importance of each of these factors in potential, culture, and single-cell coinfection?
3) Do prophages limit viral coinfection?
4) Do ssDNA and dsDNA viruses show evidence of preferential coinfection?
5) Does the CRISPR bacterial defense mechanism limit coinfection of single cells?

The results of this study suggest that microbial host ecology and virus-virus interactions are consistently important mediators of the frequency and extent of coinfection. Host taxonomy and CRISPR defense mechanisms also shaped culture and single-cell coinfection patterns, respectively. Virus-virus interactions served to limit and 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 (Table S1). These data are composed of three data sets that provide an increasingly fine-grained examination of coinfection from cross-infectivity (potential coinfection), to coinfection at the culture (pure cultures or single colonies, not necessarily single cells) and single-cell levels. The data set examining cross-infectivity assessed infection experimentally with laboratory cultures, while other two data sets (culture and single-cell coinfection) used sequence data to infer infection.

The first data set on cross-infectivity (potential coinfection) is composed of bacteriophage host-range infection matrices documenting the results of experimental attempts at lytic infection in cultured phage and hosts (Flores *et al.*, 2011). It compiles results from 38 published studies, encompassing 499 phages and 1,005 bacterial hosts. Additionally, I entered new metadata (ecosystem and ecosystem category) to match the culture coinfection data set (see below), to enable comparisons between both data sets. 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 on culture coinfection is derived from viral sequence information mined from published microbial genome sequence data on NCBI databases (Roux *et al.*, 2015). Thus, this second data set provided information on actual (as opposed to potential) coinfection of

cultures, representing 12,498 viral infections in 5,492 bacterial and archaeal hosts. The 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 at the single-cell level, they are more conservatively regarded as culture coinfections. In addition, I imported metadata from the US Department of Energy, Joint Genome Institute, Genomes Online Database (GOLD: Mukherjee *et al.*, 2016) to assess the ecological and host factors (ecosystem, ecosystem category, and energy source) that could influence culture coinfection. I curated these records and added missing metadata (n = 4964 records) by consulting strain databases (NCBI BioSample, DSMZ, BEI Resources, PATRIC and ATCC) and the primary literature.

The third data set included single-cell amplified genomes, providing information on coinfection and virus-virus interactions within single cells (Roux *et al.*, 2014). This genomic data set is covers viral infections in 127 single cells of SUP05 marine bacteria (sulfur-oxidizing Gammaproteobacteria) isolated from different depths of the oxygen minimum zone in the Saanich Inlet near Vancouver Island, British Columbia (Roux *et al.*, 2014). These single-cell data represent a combined 143 viral infections including past infections (CRISPRs and prophages) and active infections (active at the time of isolation, e.g. ongoing lytic infections).

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 cross-infectivity (potential coinfection) and coinfection

To test the factors that potentially influence coinfection –ecology, host taxonomy/phylogeny, host defense mechanisms, and virus-virus interactions– I conducted regression analyses on each of the three data sets, representing an increasingly fine scale analysis of coinfection from cross-infectivity (potential coinfection), to culture coinfection, and finally to single-cell coinfection. The data sets represent three distinct phenomena related to coinfection and thus the variables and data in each one are necessarily different. Therefore ecological factors and bacterial taxonomy were tested in all data sets, but virus-virus interactions, for example, were not evaluated in the cross-infectivity data set because they do not apply (i.e., measures potential and not actual coinfection). Table 2 details the data used to test each of the factors that potentially influence coinfection.

Table 2. *Factors explaining coinfection tested with each of the three data sets and (specific variable tested).*

	Cross-infectivity (*potential* coinfection)	Culture-level coinfection	Single-cell coinfection
Ecology	Yes (habitat, association)	Yes (habitat)	Yes (depth)
Taxonomic Group	Yes (rank)	Yes (rank)	Yes (rRNA cluster)
Host Defense	No	No	Yes (CRISPR)
Virus-virus interactions	N/A	Yes (prophage, ssDNA)	Yes (prophage)

First, to test the potential influence of these factors on potential coinfection (the estimate of phage infecting each host) I conducted a factorial analysis of variance (ANOVA) on the potential coinfection (cross-infectivity) data set. The independent variables tested were the study type/source (natural, coevolution, artificial), bacterial taxon (roughly corresponding to Genus), habitat from which bacteria and phages were isolated, bacterial trophy (photosynthetic or

heterotrophic), and bacterial association (e.g. pathogen, free-living). Geographic origin and phage taxa were present in the metadata, but were largely incomplete; therefore they were not included in the analyses. Because the infection matrices were derived from different studies testing varying numbers of phages, the dependent variable was the proportion of phage tested that infected a given host. Model criticism suggested that ANOVA assumptions were reasonably met (Supplementary Figure 1), despite using proportion data (Warton and Hui, 2011). ANOVA on the arc-sine transform of the proportions and a binomial regression provided qualitatively similar results (data not shown, see associated code in FigShare repository).

Second, to test the factors influencing culture coinfection I conducted a negative binomial regression on the culture coinfection data set. The number of extrachromosomal viruses was the dependent variable, and the explanatory variables tested were the number of prophages, ssDNA virus presence, energy source (heterotrophic/autotrophic), taxonomic rank (Genus was selected as the best predictor over Phylum or Family, details in code in Figshare repository), and habitat (environmental, host-associated, or engineered). I conducted stepwise model selection with AIC, to arrive at a reduced model that minimized the deviance of the regression.

Third, to test the factors influencing single cell coinfection I conducted a Poisson regression on single cell data set. The number of actively infecting viruses was the dependent variable and the explanatory variables tested were phylogenetic cluster (based on small subunit rRNA amplicon sequencing), ocean depth, number of prophages, and number of CRISPR spacers. I conducted stepwise model selection with AIC, to arrive at a reduced model that minimized deviance.

Virus-virus interactions (prophages and ss/dsDNA) in culture and single cell coinfection

To obtain more detailed information on the influence of virus-virus interactions on the frequency and extent of coinfection, I analyzed the effect of prophages in culture and single-cell coinfections, and the influence of ssDNA and dsDNA viruses on culture coinfection.

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

To examine the effect of prophages on coinfection in the single-cell data set, I examined cells that harbored putative defective prophages in the genome and calculated the proportion of those that also had active infections. To determine the extent of coinfection in cells prophages, I calculated the average number of active viruses infecting these cells. I examined differences in the frequency of active 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.

To examine whether ssDNA and dsDNA viruses exhibited non-random patterns of culture coinfection, I compared the frequency of dsDNA-ssDNA mixed coinfections against ssDNA-

only coinfections among all host cultures coinfected with at least one ssDNA virus, using a binomial test.

Effect of CRISPR-Cas bacterial defense on single cell coinfection

In order to obtain a more detailed picture of the role of bacterial defense in coinfection, I investigated the effect of CRISPR spacers (representing past viral infections) on the frequency of cells undergoing current (active) infections. I examined cells that harbored CRISPR spacers in the genome and calculated the proportion of those that also had active infections. To determine the extent of coinfection in cells with 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 CRISPR spacers 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 ggplot2 package (Wickham, 2009). Means are presented as means ± standard deviation, unless otherwise stated. Data and code for analyses and figures are available in the Figshare data repository (FigShare doi:10.6084/m9.figshare.2072929).

Results

Host ecology and virus-virus interactions consistently explain variation in potential, culture, and single-cell coinfection

To test the factors that influence microbial coinfection, I conducted regression analyses on each of the three data sets (cross-infectivity/potential coinfection, culture coinfection, and single cell coinfection). The explanatory variables tested in each data set varied slightly (Table 2), but can be grouped into host ecology, virus-virus interactions, host taxonomic or phylogenetic group,

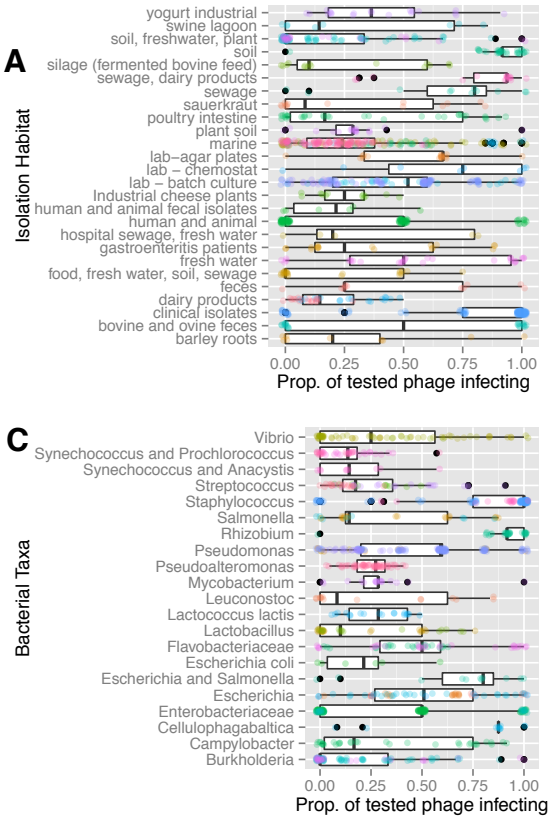
and host defense. Overall, host ecology and virus-virus interactions appeared as statistically significant factors that explained a substantial amount of variation in coinfection in every data set they were tested (see details for each data set below). Host taxonomy was a less consistent predictor across the data sets. The taxonomic group of the host explained little variation or was not a statistically significant predictor of potential (cross-infectivity) and single-cell coinfection, respectively. However, host taxonomy was the strongest predictor of the amount of culture coinfection, judging by the amount of deviance explained in the negative binomial regression. Host defense, as measured by CRISPR spacers, was tested only in the single cell data set and was a statistically significant and strong predictor of coinfection.

Potential for coinfection (cross-infectivity) is shaped by bacterial ecology and taxonomy

To test the viral and host factors that affected potential coinfection, I conducted an analysis of variance on the cross-infectivity data set, a compilation of 38 studies (Flores *et al.*, 2011) of phage host-range (499 phages; 1,005 bacterial hosts). Stepwise model selection with AIC, yielded a reduced model that explained 33.89% of the variance in potential coinfection using three factors: host association (e.g. free-living, pathogen), isolation habitat (e.g. soil, animal), and taxonomic grouping (Figure 1). The two ecological factors together explained >30% of the variance in potential coinfection, while taxonomy explained only ~3% (Figure 1D). Host association explained 8.41% of the variance in potential coinfection. Bacteria that were pathogenic to cows had the highest potential coinfection with more than 75% of tested phage infecting each bacterial strain (Figure 1B). In absolute terms, the average host that was a pathogenic to cows could be infected 15 phages on average. The isolation habitat explained 24.36% of the variation in potential coinfection with clinical isolates having 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) and, in absolute terms, the average host in each of these habitats could be infected by 3-15 different phages.

Bacterial trophy (heterotrophic, autotrophic) and the type of study (natural, coevolution, artificial) were not selected by AIC in final model. An alternative categorization of habitat that matched the culture coinfection data set (ecosystem: host-associated, engineered, environmental) was not a statistically significant predictor and was dropped from the model, but results for this variable are shown in Figure S1 and Table S2. Details of the full, reduced, and alternative models are provided in Figure S2 and associated code in the FigShare repository.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bacterial.Association	7	10.6497	1.5214	18.2389	0
Isolation_Habitat_new	22	30.8534	1.4024	16.8127	0
Bacteria_new	5	4.259	0.8518	10.2115	0
Residuals	970	80.9121	0.0834	NA	NA

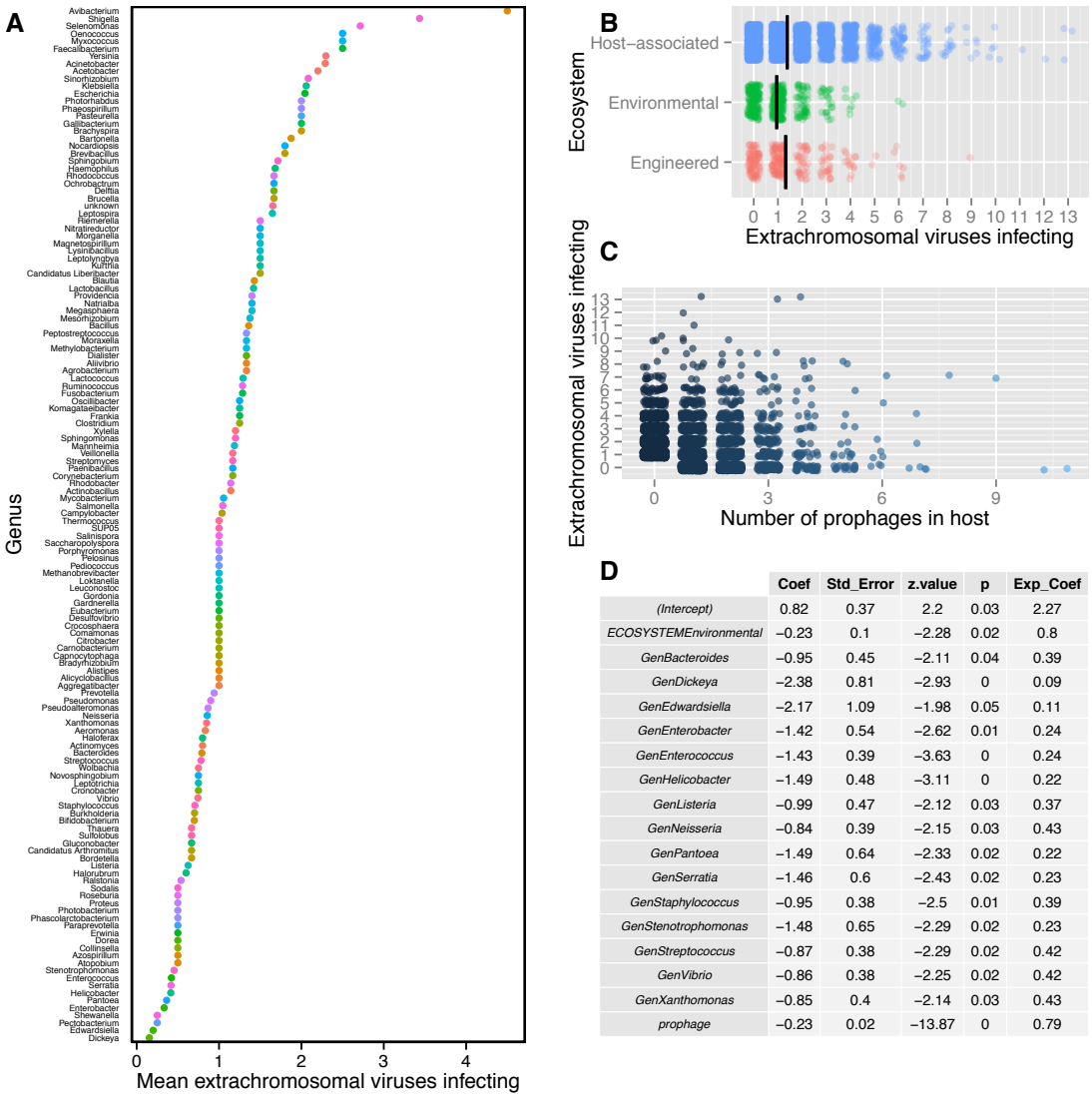
Figure 1. Bacterial-phage ecology and taxonomy explain most of the variation in potential coinfection. Potential coinfection is the number of phages that can infect a bacterial host, here measured as the proportion of tested phages infecting each host (represented by points). Points represent hosts; 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. Those factors selected after stepwise model selection using AIC are depicted in panels A-C. The ANOVA table is presented in panel D.

Culture coinfection is influenced by host ecology and taxonomy and virus-virus interactions

To test the viral and host factors that affected culture coinfection, I conducted a negative binomial regression on the culture coinfection data set, which documented 12,498 viral infections in 5,492 microbial hosts using NCBI-deposited sequenced genomes (Roux *et al.*, 2015) supplemented with metadata collected from GOLD database (Mukherjee *et al.*, 2016). Stepwise model selection with AIC resulted in a reduced model that used three factors to explain the number of extrachromosomal infections (representing lytic, chronic or carrier state infections): host taxonomy (Genus), number of prophages and, host ecosystem. The genera with the most coinfection (mean >2.5 extrachromosomal infections) were *Avibacterium*, *Shigella*, *Selenomonas*, *Faecalibacterium*, *Myxococcus*, and *Oenococcus*, whereas *Enterococcus*, *Serratia*, *Helicobacter*, *Pantoea*, *Enterobacter*, *Pectobacterium*, *Shewanella*, *Edwardsiella*, and *Dickeya* were rarely coinfected (mean < 0.45 extrachromosomal infections)(Figure 2A). Microbes that were host-associated had the highest mean extrachromosomal infections (1.39 ± 1.62, n=4,282), followed by engineered (1.32 ± 1.44, n=281) and environmental ecosystems (0.95 ± 0.97, n=450)(Figure 2B). The model predicts that each additional prophage reduces extrachromosomal infections by >79% (Figure 2C).

Host energy source (heterotrophic, autotrophic) and ssDNA virus presence were not statistically significant predictors as judged by AIC model selection. Details of the full, reduced, and alternative models are provided in the Supplementary Materials and all associated code and data is deposited in FigShare repository.



D

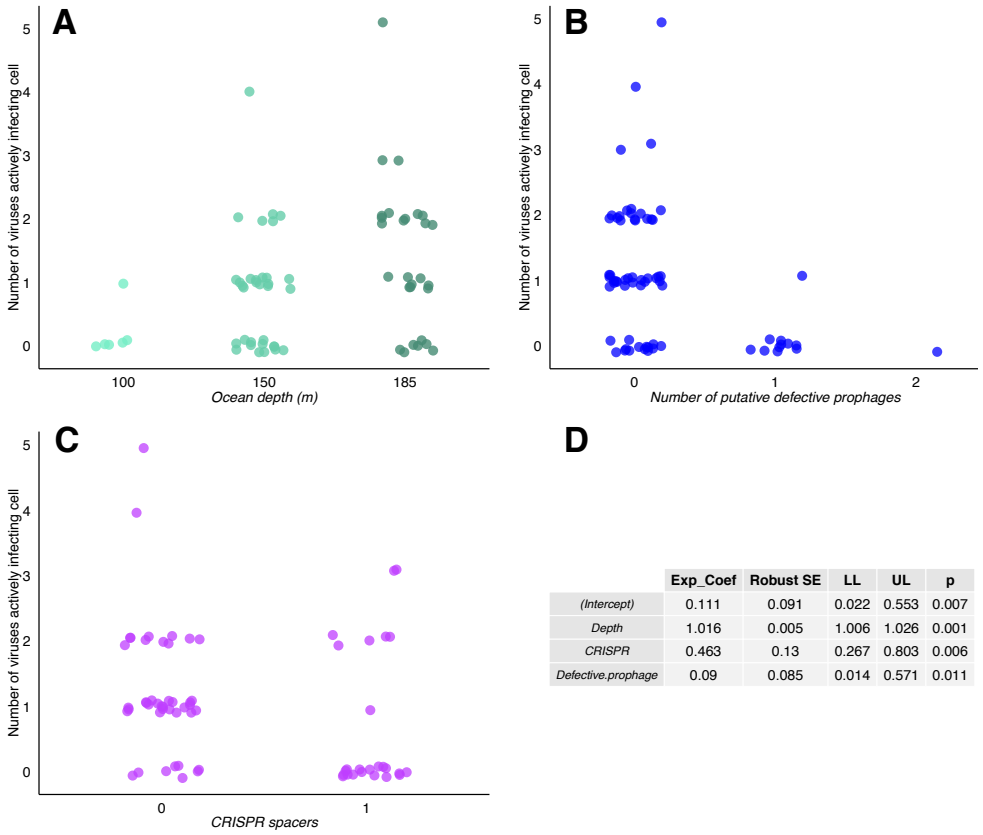
	Coef	Std_Error	z.value	p	Exp_Coef
(Intercept)	0.82	0.37	2.2	0.03	2.27
ECOSYSTEMEnvironmental	−0.23	0.1	−2.28	0.02	0.8
GenBacteroides	−0.95	0.45	−2.11	0.04	0.39
GenDickeya	−2.38	0.81	−2.93	0	0.09
GenEdwardsiella	−2.17	1.09	−1.98	0.05	0.11
GenEnterobacter	−1.42	0.54	−2.62	0.01	0.24
GenEnterococcus	−1.43	0.39	−3.63	0	0.24
GenHelicobacter	−1.49	0.48	−3.11	0	0.22
GenListeria	−0.99	0.47	−2.12	0.03	0.37
GenNeisseria	−0.84	0.39	−2.15	0.03	0.43
GenPantoea	−1.49	0.64	−2.33	0.02	0.22
GenSerratia	−1.46	0.6	−2.43	0.02	0.23
GenStaphylococcus	−0.95	0.38	−2.5	0.01	0.39
GenStenotrophomonas	−1.48	0.65	−2.29	0.02	0.23
GenStreptococcus	−0.87	0.38	−2.29	0.02	0.42
GenVibrio	−0.86	0.38	−2.25	0.02	0.42
GenXanthomonas	−0.85	0.4	−2.14	0.03	0.43
prophage	−0.23	0.02	−13.87	0	0.79

Figure 2. Host taxonomy, ecology, and number of prophages predict variation in culture coinfection. Plots depict all variables from a reduced model explaining the number of extrachromosomal infections (A-C). Panel A depicts the mean number of extrachromosomal infections in all microbial genera with >1 host sampled and nonzero means. Panel A and B have data points are offset by a random and small amount (jittered) to enhance visibility and reduce overplotting. The regression table is presented in panel D, and only includes variables with $p < 0.05$.

Single-cell coinfection is influenced by bacterial ecology, virus-virus interactions, and the CRISPR-Cas defense mechanism

To test the viral and host factors that affected viral coinfection of single cells, I constructed a Poisson regression on the single-cell coinfection data set, which characterized viral infections in SUP-05 bacteria isolated directly from their marine habitat (Roux *et al.*, 2014). Stepwise model

selection with AIC led to a model that included three factors explaining the number of actively infecting viruses in single cells: number of prophages, number of CRISPR spacers, and ocean depth where the cells were collected (Figure 3). The model estimated each additional prophage would result in a ~9% decrease in active infections, while each CRISPR spacer would result in an >46% decrease. Depth had a positive influence on coinfection: every 50 meter increase in depth was predicted to result in ~80% more active infections.



	Exp_Coef	Robust SE	LL	UL	p
(Intercept)	0.111	0.091	0.022	0.553	0.007
Depth	1.016	0.005	1.006	1.026	0.001
CRISPR	0.463	0.13	0.267	0.803	0.006
Defective.prophage	0.09	0.085	0.014	0.571	0.011

Figure 3. Host ecology, number of prophages, and CRISPR spacers predict variation in single-cell coinfection. Plots A-C depict all variables from a reduced model explaining the number of active infections in single cells of marine SUP05 bacteria. Each point represents a cell and is offset by a random and small amount (jittered) to enhance visibility and reduce overplotting. The regression table is presented in panel D.

The phylogenetic cluster of the SUP-05 bacterial cells (based on small subunit rRNA amplicon sequencing) was not a statistically significant predictor selected using AIC model criticism.

Details of the full, reduced, and alternative models are provided in the Supplementary Materials and all associated code and data is deposited in FigShare repository.

Prophages limit culture and single-cell coinfection

Based on the results of the regression analyses and aiming to test the phenomenon of superinfection exclusion in a large data set, I conducted more fine-grained analyses regarding the influence of prophages on coinfection in microbial cultures and single cells.

First, because virus-virus interactions can occur in cultures of cells, I tested whether prophage-infected host cultures reduced the probability of other viral infections in the entire culture coinfection data set. 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 4A 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 4A). 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

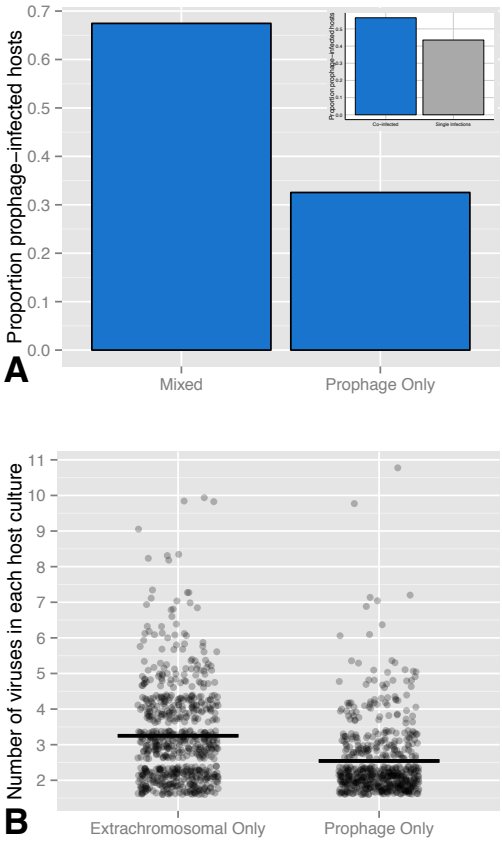


Figure 4. Host cultures infected with prophages limit coinfection by other prophages, but not extrachromosomal viruses. A slight, but statistically significant, majority of prophage-infected host cultures were coinfected (A-inset). Of these, host cultures containing multiple prophages were less frequent 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).

387 infected by 3.25 ± 1.34 viruses, compared to 2.54 ± 1.02 prophages (n=575); these quantities
388 showed a statistically significant difference (Wilcoxon Rank Sum: W = 125398.5, $p < 2.2e^{-16}$,
389 Figure 4B).
390
391 Second, to test whether past infections integrated into the host genome could affect coinfection
392 of single cells in a natural environment, I examined a single cell amplified genomics data set of
393 SUP05 marine bacteria. Cells with putative defective prophages were less likely to have current
394 infections: 9.09% of cells with prophages had current infections, compared to 74.55% of cells
395 that had no prophages (X-squared = 14.2607, df = 1, p-value = 0.00015). Bacteria with
396 prophages were currently infected by an average of 0.09 ± 0.30 phages, whereas bacteria without
397 prophages were infected by 1.22 ± 1.05 phages (Figure 3B). No host with prophages had current
398 coinfections (i.e., > 1 active virus infection).

399 ***Non-random coinfection of host cultures by ssDNA and dsDNA viruses suggests***
400 ***mechanisms enhancing coinfection***
401 The number of ssDNA viruses in culture coinfections was not a statistically significant predictor
402 of extrachromosomal infections in the regression analysis. However, to test the hypothesis that
403 ssDNA-dsDNA viral infections exhibit non-random coinfection patterns that increase the
404 likelihood of coinfection, I conducted a focused analysis on the culture coinfection data set.
405 Coinfected host cultures containing ssDNA viruses (n = 331), were more likely to have dsDNA
406 or unclassified viruses (70.69%), than multiple ssDNA infections (exact binomial: $p= 3.314e^{-14}$).
407 These coinfections were >2 times more likely to involve at least one dsDNA viruses than none
408 (exact binomial: $p = 2.559e^{-11}$, Figure 5).

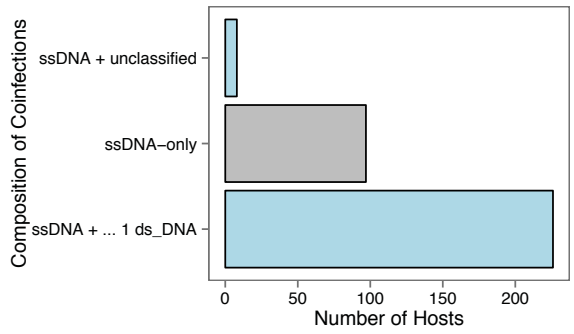


Figure 5. Culture coinfections between ssDNA-dsDNA viruses are more common than expected by chance. Host cultures with dsDNA-ssDNA coinfections, were more frequent than ssDNA-only coinfections (shaded in gray).

CRISPR spacers limit coinfection at single cell level without spacer matches

The regression analysis of the single-cell data set revealed the CRISPR bacterial defense mechanism had a significant overall effect on coinfection in SUP-05 marine bacteria, therefore I examined the effects of CRISPR spacers more closely. Cells with CRISPR spacers were 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 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 3C). In contrast to prophages in single cells, cells with CRISPR spacers could have active infections and coinfections with up to 3 phages. None of the CRISPR spacers matched any of the actively infecting viruses (Roux *et al.*, 2014).

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 host ecology and virus-virus interactions in shaping potential, culture, and single-cell coinfection. In the most comprehensive test of the

phenomenon of superinfection immunity conferred by prophages, I found that prophages limit coinfection of cultures by other prophages, but less strongly by extrachromosomal viruses. Furthermore, prophages completely excluded coinfection (by prophages or extrachromosomal viruses) in single cells of SUP-05 marine bacteria. In contrast, 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 revealed that CRISPR spacers limit coinfection of single cells in a natural environment, despite the absence of spacer matches in the infecting viruses. In light of the increasing awareness of the widespread occurrence of viral coinfection, this study provides the foundation for future work on the frequency, mechanisms, and dynamics of viral coinfection and its ecological and evolutionary consequences.

Host correlates of coinfection

Host ecology stood out as an important predictor of coinfection across all three datasets: potential, culture, and single cell coinfection. The specific ecological variables differed in the data sets (bacterial association, isolation habitat, and ocean depth), but ecological factors were retained as statistically significant predictors in all three models. Moreover, when ecological variables were standardized between the potential and culture coinfection data sets, the differences in coinfection were remarkably similar (Figure S5, Table S2). This result lends further support to the consistency of host ecology as a predictor of coinfection and suggests that the ecological drivers of *potential* coinfection might also drive *realized* coinfection. On the other hand, host taxonomy was a less consistent predictor. It was weak or absent in the potential coinfection and single-cell coinfection models, respectively, yet it was the strongest predictor of culture coinfection. This difference could be because the hosts in the potential coinfection and single-cell data sets varied predominantly at the strain level (Flores *et al.*, 2011; Roux *et al.*,

2014), whereas infection patterns are less variable at higher taxonomic scales (Flores *et al.*, 2013; Roux *et al.*, 2015), as seen with Genus in the culture coinfection model. Collectively, these findings suggest that the diverse and complex patterns of cross-infectivity (Holmfeldt *et al.*, 2007) and coinfection observed at the level of bacterial strains may be best explained by local ecological factors, while at higher taxonomic ranks the phylogenetic origin of hosts increases in importance (Flores *et al.*, 2011; 2013; Roux *et al.*, 2015). Particular bacterial lineages can exhibit dramatic differences in cross-infectivity (Koskella and Meaden, 2013; Liu *et al.*, 2015) and, as this study shows, coinfection. Thus, further studies with ecological data and multi-scale phylogenetic information will be necessary to test the relative influence of bacterial phylogeny on coinfection.

Bacterial defense was another important factor influencing coinfection patterns, but was only tested in the single-cell data set. 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 some of the first evidence from a natural environment that CRISPR's protective effects extend beyond viruses with exact matches to the particular spacers within the cell (Fineran *et al.*, 2014; Semenova *et al.*, 2011). Although very specific sequence matching is thought to be required for CRISPR-Cas-based immunity (Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Mojica *et al.*, 2005), the system can tolerate mismatches in protospacers (within and outside of the seed region: Semenova *et al.*, 2011; Fineran *et al.*, 2014), enabling protection (interference) against related phages by a mechanism of enhanced spacer acquisition termed priming (Fineran *et al.*, 2014). The seemingly broader protective effect of CRISPR-Cas beyond specific sequences may help explain continuing effectiveness of CRISPR-

Cas (Fineran *et al.*, 2014) in the face of rapid viral coevolution for escape (Heidelberg *et al.*, 2009; Tyson and Banfield, 2008; Andersson and Banfield, 2008). To elucidate the roles of bacterial defense systems in shaping coinfection, more data on CRISPR-Cas and other viral infection defense mechanisms will be required across different taxa and environments.

This study revealed the strong predictive power of several host factors in explaining viral coinfection, yet there was still substantial unexplained variation in the regression models (see Results). Thus, the host factors tested herein should be regarded as starting points for future experimental examinations. For instance, hetero- vs. autotrophy was not a statistically significant predictor in the potential coinfection and culture coinfection models, perhaps due to the much smaller sample sizes of autotrophic hosts. However, both data sets yield similar summary statistics, with heterotrophic hosts having 59% higher cross-infectivity and 77% higher culture coinfection (Figure S6). Moreover, other factors not examined, such as geography could plausibly affect coinfection. 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, as observed with 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.

The role of virus-virus interactions in coinfection

The results of this study suggest that virus-virus interactions play a role in limiting *and* enhancing coinfection. First, prophages limit coinfection in host cultures and single cells. In what is effectively the largest test of viral superinfection exclusion (n = 3,134 hosts), prophages limited coinfection of host cultures by other prophages, but not by extrachromosomal viruses.

Although this focused analysis did not find strong evidence of prophages limiting extrachromosomal infection, the regression analysis suggested a slight, but statistically significant ~9% reduction for each additional prophage. As these were culture coinfections and not necessarily single-cell coinfections, these results are consistent with a single-cell study of *Salmonella* cultures showing that lysogens can activate cell subpopulations to be transiently immune from viral infection (Cenens *et al.*, 2015). Prophages had a more dramatic impact at the single-cell level in SUP05 marine bacteria in a natural environment, severely limiting active viral infection and completely excluding coinfection. The results on culture-level and single-cell coinfection come from very different data sets, which should be examined carefully before drawing general patterns. First, the culture-level data set is composed of an analysis of all publicly available bacterial and archaeal genome sequences in NCBI databases that show evidence of a viral infection. These sequences show a bias towards particular taxonomic groups (e.g. Proteobacteria, model study species) and those that are easy to grow in pure culture. The single cell data set is limited to just one host type isolated in a particular environment, as opposed to the 5,492 hosts in the culture coinfection data set. This limitation prohibits taxonomic generalizations about the effects on prophages on single cells, but extends 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

(Dulbecco, 1952; Refardt, 2011), suggesting latent infections are a profitable avenue for future research on virus-virus interactions.

Second, another virus-virus interaction 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 bacteria from the SUP05_03 cluster. 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 across a diverse set of 331 bacterial and archaeal hosts. Thus, enhanced coinfection, perhaps due to the long replicative cycle of some ssDNA viruses (e.g. Innoviridae: Rakonjac *et al.*, 2011), might be a major factor explaining findings of phages with chimeric genomes composed of different types of nucleic acids (Roux *et al.*, 2013; Diemer and 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 and Duffy, 2009).

Implications and applications

Collectively, these results suggest microbial host 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,

543 superinfection avoidance may actually highlight the selective force of virus-virus interactions. In
544 an evolutionary sense, this viral trait exists precisely because the potential for coinfection is high.
545 If this is correct, variability in potential and realized coinfection, as found in this study, suggests
546 that the manifestation of superinfection exclusion will vary across viral groups according to their
547 ecological context. Accordingly, some viral mechanisms will promote coinfection, as found in
548 this study with ssDNA/dsDNA coinfections and in other studies (Dang *et al.*, 2004; Cicin-Sain *et*
549 *al.*, 2005; Turner *et al.*, 1999). I found substantial variation in potential, culture, and single-cell
550 coinfection and, in the analyses herein, ecology was always a statistically significant and strong
551 predictor of coinfection, suggesting that the selective pressure for coinfection is going to vary
552 across local ecologies. This is in agreement with observations of variation in viral genetic
553 exchange (which requires coinfection) rates across different geographic localities in a variety of
554 viruses (Díaz-Muñoz *et al.*, 2013; Trifonov *et al.*, 2009; Held and Whitaker, 2009).
555
556 These results have clear implications, not only for the study of viral ecology in general, but for
557 practical biomedical and agricultural applications of phages and bacteria/archaea. Phage therapy
558 is often predicated on the high host specificity of phages, but intentional coinfection could be an
559 important part of the arsenal as part of combined or cocktail phage therapy. This study also
560 suggests that viral coinfection in the microbiome should be examined, as part of the influence of
561 the virome on the larger microbiome (Pride *et al.*, 2012; Minot *et al.*, 2011; Reyes *et al.*, 2010).
562 Finally, if these results apply in eukaryotic viruses and their hosts, variation in viral coinfection
563 rates should be considered in the context of treating and preventing infections, as coinfection
564 likely represents the default condition of human hosts (Wylie *et al.*, 2014). Coinfection and
565 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. T.B.K. Reddy kindly provided database records from Joint Genome Institute's Genomes Online Database (GOLD). I am indebted to Joshua Weitz, Britt Koskella, and Jay Lennon 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.

Data availability

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

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.

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

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

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

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

600 Birger RB, Kouyos RD, Cohen T, Griffiths EC, Huijben S, Mina M, *et al.* (2015). The potential
601 impact of coinfection on antimicrobial chemotherapy and drug resistance. *Trends Microbiol* **23**:
602 537–544.

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

605 Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, *et al.* (2008).
606 Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960–964.

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

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

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

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

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

619 Diemer GS, Stedman KM. (2012). A novel virus genome discovered in an extreme environment
620 suggests recombination between unrelated groups of RNA and DNA viruses. *Biology Direct* **7**:
621 13.

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

624 Díaz-Muñoz SL, Tenaillon O, Goldhill D, Brao K, Turner PE, Chao L. (2013). Electrophoretic
625 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*. e-pub ahead of print, doi: 10.1371/journal.pone.0004169.s006.

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, Suttle CA. (2013). Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Front Microbiol*. e-pub ahead of print, doi: 10.3389/fmicb.2013.00404/abstract.

Labrie SJ, Samson JE, Moineau S. (2010). Bacteriophage resistance mechanisms. *Nature Publishing Group* **8**: 317–327.

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*. e-pub ahead of print, doi: 10.1038/ismej.2015.47.

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.

Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, Isbandi M, *et al.* (2016). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Research*. e-pub ahead of print, doi: 10.1093/nar/gkw992.

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.

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

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

698 Roux S, Hallam SJ, Woyke T, Sullivan MB. (2015). Viral dark matter and virus-host interactions
699 resolved from publicly available microbial genomes. *eLife* **4**. e-pub ahead of print, doi:
700 10.7554/eLife.08490.

701 Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, *et al.*
702 (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by
703 single-cell- and meta- genomics. *eLife* **3**. e-pub ahead of print, doi: 10.7554/eLife.03125.

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

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

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

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

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

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

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

719 Turner PE, Duffy S. (2009). Evolutionary ecology of multiple phage adsorption and infection.
720 In: Abedon ST (ed). *Bacteriophage Ecology: Population Growth, Evolution, and Impact of*
721 *Bacterial Viruses*. Cambridge University Press: Cambridge. e-pub ahead of print, doi:
722 10.1017/CBO9780511541483.

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

725 Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. (2006). Quasispecies diversity
726 determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**: 344–
727 348.

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.

Wickham H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag: New York.

Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF, Fuhrman JA, *et al.* (2016). Wiggington_etal_2016. *Nature Microbiology* 1–8.

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.