Review

Cell PRESS

Phage–bacteria infection networks

Joshua S. Weitz[1,2], Timothée Poisot[3,4], Justin R. Meyer[5*], Cesar O. Flores[2], Sergi Valverde[6], Matthew B. Sullivan[7], and Michael E. Hochberg[8,9]

[1] School of Biology, Georgia Institute of Technology, Atlanta, GA, USA
[2] School of Physics, Georgia Institute of Technology, Atlanta, GA, USA
[3] Département de Biologie, Chimie et Géographie, Université du Québec à Rimouski, Québec, Canada
[4] Québec Center for Biodiversity Sciences, Montréal, Québec, Canada
[5] BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI, USA
[6] ICREA-Complex Systems Lab and CSIC-Institute of Evolutionary Biology, Universitat Pompeu Fabra, Barcelona, Spain
[7] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA
[8] Institut des Sciences de l'Evolution, University of Montpellier II, Montpellier, France
[9] Santa Fe Institute, Santa Fe, NM, USA

Phage and their bacterial hosts are the most abundant and genetically diverse group of organisms on the planet. Given their dominance, it is no wonder that many recent studies have found that phage–bacteria interactions strongly influence global biogeochemical cycles, incidence of human diseases, productivity of industrial microbial commodities, and patterns of microbial genome diversity. Unfortunately, given the extreme diversity and complexity of microbial communities, traditional analyses fail to characterize interaction patterns and underlying processes. Here, we review emerging systems approaches that combine empirical data with rigorous theoretical analysis to study phage–bacterial interactions as networks rather than as coupled interactions in isolation.

Phages: key components of complex microbial communities

Historically, the study of phages facilitated important advances in molecular biology [1]. More recently, the discovery of high levels of viral abundance and diversity in natural environments has sparked the burgeoning field of 'viral ecology' [2–4]. Viruses, including phages and viruses of microeukaryotes, are thought to have key effects on microbial ecosystems. For example, viruses are responsible for an estimated 20–40% of marine microbial mortality [3]. Viral-induced lysis of microbes redirects organic matter back into the microbial loop and away from zooplankton and fish [5,6]. Viral infection also facilitates gene transfer both among phages and their hosts, thereby contributing to the webbed architecture of the tree of life [7]. In addition, phages may also impact human health, for example, by altering the fate of pathogenic bacteria such as *Vibrio cholerae* [8], or by interacting with bacteria found in the human microbiome [9,10].

However, despite the increasing recognition that phages play a significant role in shaping microbial ecosystems, fundamental questions remain unanswered. For example: which phages infect and exploit which hosts in complex communities? Quantifying who infects whom is essential to understand how infections at the cellular level (such as changes to metabolic rates, gene transfer, and the fate of cells) scale-up to influence ecosystem function in complex environments. Here, we synthesize approaches and findings from multiple disciplines to clarify the roles of ecological and evolutionary processes in structuring phage–bacteria infection networks (PBINs). In so doing, we highlight the ways in which a better understanding of PBINs will further predictive models of viral effects on microbial communities, from microbiomes to the Earth system.

Defining PBINs

Cross-infection of phage isolates against a panel of bacterial isolates is a microbiological tool used for many purposes, including the identification of pathogenic strains, analysis of strain-specific lysis in complex communities, and characterization of coevolutionary dynamics. In practice, the host range of a given phage type is determined using infection tests such as spot assays. In a spot assay, a small sample of phages from pure culture is added to a bacterial lawn and infection is marked as positive if a clearing is observed[†].

Phages can infect individual hosts from different species and even different genera, as in the case of cyanophages infecting *Prochlorococcus* and *Synechococcus* [11]. In some cases, a given phage may infect only one strain of bacteria from a given sample. However, cross-infection is commonly observed and the emergent patterns resulting from the cross-infection of multiple phages and multiple hosts can be challenging to interpret. The difficulty in interpretation arises, in part, due to the many possible ways to represent cross-infection studies. For example, a cross-infection matrix involving 20 hosts, one per row, and 20 viruses, one per column, could be represented in approximately 6×10^{36} ways [there are 20 hosts to choose from for the first row, then 19 for the second row, and so on; the same holds for the

Corresponding author: Weitz, J.S. (jsweitz@gatech.edu)
* *Current address:* Department of Systems Biology, Harvard Medical School, Boston, MA, USA.
Keywords: microbial ecology; ecosystem; biodiversity; functional diversity; complex networks; theory.

† Note that the majority of studies utilizing spot assays have focused, as we have, on antagonistic relationships. However, the lysogenic mode of interaction warrants further attention using the methods reviewed and synthesized here.

Box 1. Key types of PBINs

The four key types of PBINs are: random, one-to-one, nested, and modular (Figure I).

Random: the pattern of who infects whom is not statistically different than what would be expected if cross-infection occurred by chance.

One-to-one: an infection network with elevated specialization, such that each phage can only infect one host, and each host is only infected by one phage.

Nested: a PBIN that contains interactions that form a hierarchy for both phages and hosts. Consider the case of a maximally nested PBIN with *S* host and *S* phage types (see Figure I for an example in the case of nine hosts and nine phage types). A perfectly nested PBIN is one in which both bacteria and phages can be ordered from 1 to *S*: in this ordering, bacterium 1 is the type most difficult to infect (i.e., only one of the *S* phages can infect it) whereas bacteria *S* is the most permissive to infection (i.e., all of the *S* phages can infect it). The next most permissive bacteria can be infected by all but one phage, and so on. Similarly, in this ordering, phage *S* is a generalist and can infect all *S* bacteria whereas phage 1 is a specialist and can only infect one bacterial type. The next most specialized phage infects the two most permissive bacteria, and so on.

Modular: a PBIN that contains interactions that tend to occur among distinct groups of phages and hosts. Consider the case of a maximally modular PBIN with *S* host and *S* phage types (see Figure I for an example). In this case, all infections occur between phages and bacteria in the same group (i.e., 'module') rather than between groups. There are three modules in the example in Figure I. A module is a group of phages and hosts for which the phages in the set are more likely to infect the hosts in the group, and likewise, that the hosts in the group are more likely to be infected by phages from within the group. For any given size of a PBIN, the maximally modular configuration would be one in which the infections occur between phages and hosts within the same module.

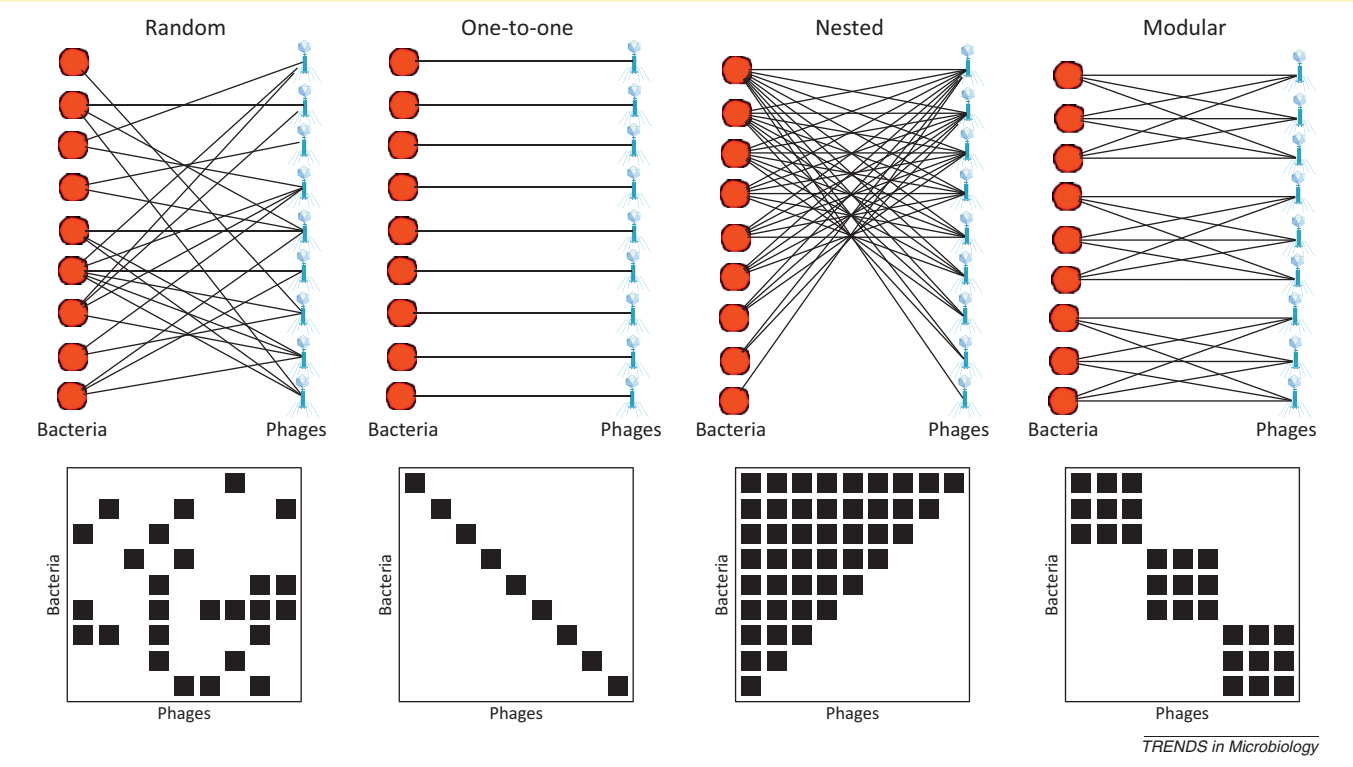

Figure I. Paired schematics of four PBINs. Top, network representation; bottom, matrix representation. Infections between phages and bacteria are represented as black lines (top) and as black cells (bottom).

placement of viruses in columns, that is, the total is equal to $(20 \times 19 \times 18 \times \ldots \times 3 \times 2 \times 1)^2$]. Some of these representations may be more informative than others for revealing regularities and patterns in who infects whom.

Network-based approaches have recently been proposed to help unify and shed light on the quantitative analysis of the cross-infection of multiple phages with multiple bacteria [12,13]. Within a PBIN, phages and bacteria are represented as nodes. An edge between a phage node and a bacterium node indicates the ability of that phage strain to infect and lyse that host strain (Box 1). Similar networks have been used to describe interactions in other branches of ecology, for example, food webs [14] and plant–pollinator networks [15]. Networks facilitate the representation and analysis of diverse systems, particularly when there are far fewer realized interactions than potential interactions [16]. Moreover, a diversity of methods are available to analyze network structure and its consequences, many of which we highlight in this review (see [17] for a broad overview of network science). In doing so, one of the primary objectives of network analysis is to identify patterns within a network that are not necessarily expected by chance.

Nestedness and modularity: hypothesized signatures of coevolutionary mechanisms

The two most frequently examined patterns in ecological networks are nestedness and modularity [18]. Nestedness is characteristic of PBINs that have a hierarchy of resistance among hosts and infection ability among phages. Likewise, modularity is characteristic of PBINs in which bacteria and phages preferentially cross-infect within

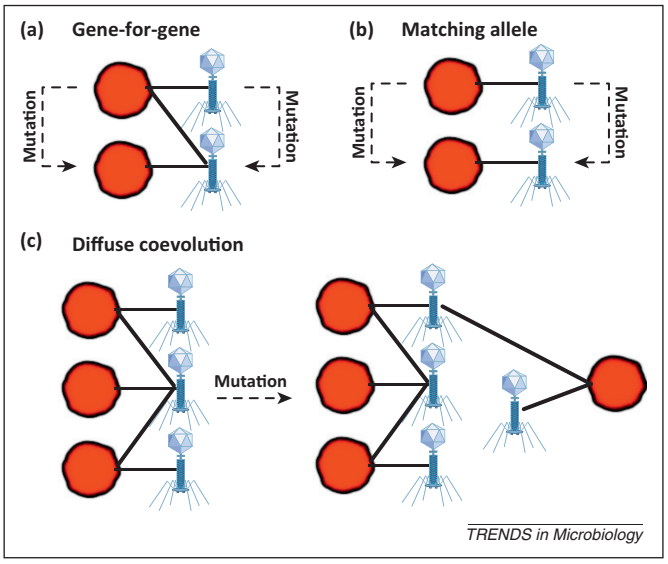

Figure 1. Schematic examples of three modes of coevolution and their effect on cross-infection. In all cases, edges (black lines) represent infections between phage and host types (circles). The three modes are **(a)** gene-for-gene coevolution; **(b)** matching allele coevolution; **(c)** diffuse coevolution. Note that in all cases evolved strains can differ in their host range (in the case of viruses) and phage susceptibility (in the case of hosts).

groups or 'modules'. In Box 1, we elaborate further on the definition and interpretation of these concepts.

The nested pattern in evolutionary PBINs is hypothesized to result from a sequence of adaptations that are arbitrated by gene-for-gene processes [19,20]. In a dynamic gene-for-gene coevolutionary sequence, new bacterial mutations confer bacterial resistance to recently evolved phages while maintaining resistance to past phages (Figure 1a). Likewise, mutations for host range expansion among phages evolve without losing the ability to infect ancestral host genotypes. Hence, the set of hosts that a phage can infect are 'nested' across a sequence, that is, the host range of phages are subsets of each other. The same nesting applies to hosts, that is, the phage susceptibility of hosts are subsets of each other.

An alternative model of phage–host coevolution posits that phages must have alleles that facilitate infection against specific bacterial defensive alleles (Figure 1b). Hence, in the simplest matching allele model [20], bacteria evolve resistance to a single phage genotype and lose any evolved resistance to other phages, whereas phage mutations confer infection against single strains of bacteria at the cost of entire loss of infectivity against ancestral strains. Specialization is promoted by the interaction genetics and selection in this coevolutionary model, possibly in a successive fashion in time resembling a Red Queen dynamic [21]. In a Red Queen sequence, each interacting organism (in this case a phage and a host) evolves, leading to lineages whose genomes are changing but yet never able to escape from interactions with their evolving partner.

Both the gene-for-gene and matching allele models are idealizations and intermediate mechanisms are possible by which phages evolve the ability to infect new hosts and (partially) lose the ability to infect existing hosts [20]. For example, there is evidence of trade-offs in which evolved phages suffer fitness costs in terms of the productivity of infecting the ancestral host [22]. Moreover, neither the gene-for-gene nor matching allele models account for larger pre-existing variation in a community, corresponding to coevolution in multi-species and/or multi-strain communities (Figure 1c). This process has been termed diffuse coevolution [23]. Regardless of mechanism, these coevolutionary 'steps' affect the dynamical emergence of cross-infection networks. Observed cross-infection networks are complex and do not, in general, have a perfectly nested or modular structure [24].

The analysis of modularity and nestedness requires (i) methods to estimate the importance of such patterns and (ii) a critical assessment of statistical significance, that is, do these patterns signal ecological and evolutionary drivers or could they have resulted by chance? We provide a description of how to estimate nestedness and modularity in Figures 2 and 3, respectively. In Box 2, we describe how the statistical significance of these patterns is calculated so as to provide background for potential users. Next, we evaluate patterns found in PBINs derived from ecological and experimental studies.

Cross-infections in the environment

Cross-infections between viruses and hosts in natural environments have been investigated for decades, without a consensus on patterns and associated mechanisms. Recently, 38 published PBINs spanning 20 years of research and nearly 12 000 individual phage–bacterial strain infection trials were aggregated and re-analyzed [12]. The majority of ecological studies assembled in this re-analysis included phages and hosts collected at different sites from within similar environments. Hence, PBINs measured in this way provide information on interactions within 'metacommunities' [25], that is, a set of communities potentially linked by dispersal. The re-analysis identified many PBINs that have a previously unrecognized nested structure (see Figure 4a,b for two examples). What are the drivers of these patterns?

First, taxonomy is known to be strongly associated with infection outcome. In instances where taxonomic identity is available, phage isolates from different clades can differ in host range and specificity. For example, Wichels and coworkers [26] found that phages of the clade *Myoviridae* have broader host ranges than *Siphoviridae*, which have broader ranges than *Podoviridae*. The host ranges of these viruses were further found to be nested. From the host perspective, variation in susceptibility to infection can occur within a clade. Despite being highly related, environmental isolates of the heterotrophic *Cellulophaga baltica*, a *Flavobacteriaceae* from the *Bacteriodes* phylum, show variable phage-susceptibility patterns consistent with a nested structure [27]. However, differences in susceptibility can also facilitate the grouping of hosts and phages into modules (e.g., Figure 4c,d). For example, Sullivan and colleagues [11] studied host isolates from multiple cyanobacteria genera, including *Synechococcus* and *Prochlorococcus*, and found that phages preferentially infected one of the two genera. Hence, the infection pattern will be, in part, a consequence of taxonomic diversity in the sample set. Further aggregation of studies will help to define patterns characteristic of cross-infection among

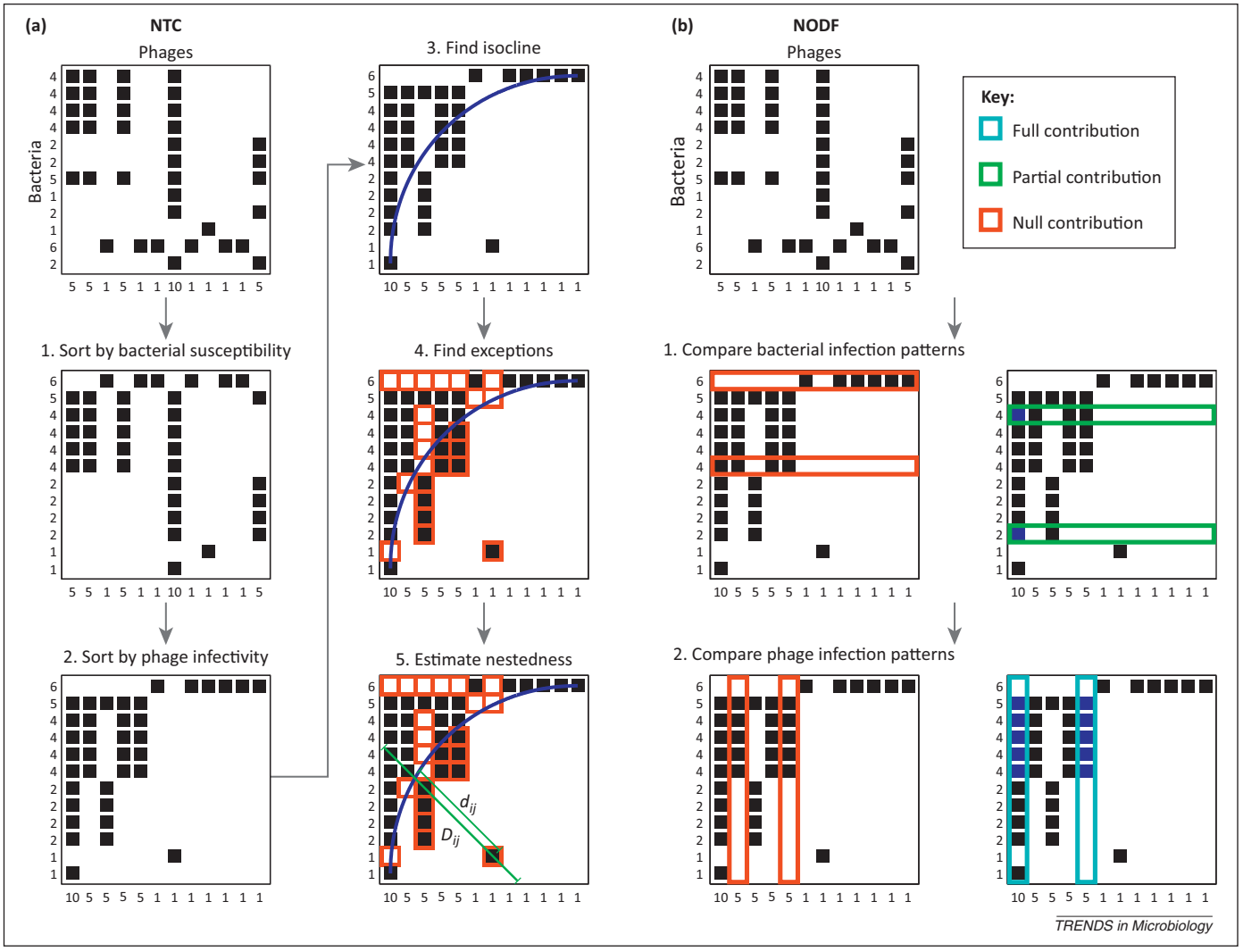

Figure 2. Methods to calculate nestedness. Two widely used methods are the temperature calculator (NTC) [74] and the overlapping fill method (NODF) [75]. We illustrate these methods, in panels (a) and (b) respectively, using a phage–bacteria infection network (PBIN) derived from interactions between *Streptococcus thermophiles* and associated phages [54]. Nestedness algorithms such as NTC and NODF take a PBIN as input and return a nestedness value between $N = 0$ (non-nested, i.e., there is not an ordering of infection and resistance) to $N = 1$ (perfectly nested, i.e., there is a perfect ordering of infection and resistance). **(a)** In NTC, the steps are: (1) rows and (2) columns are sorted in descending order of host susceptibility (the numbers alongside rows denote the number of phages that can infect a host) and phage infectivity (the numbers below columns denote the number of hosts that a phage can infect). Then, (3) an isocline is found corresponding to the arrangement of infections if the matrix were perfectly nested – in such a case, all infections would be in the upper left region, above the blue line. (4) Interactions that deviate from perfect nestedness are identified, including infections that do not occur, but are expected given perfect nestedness (see upper left red squares with white centers) and infections that do occur but are not expected given perfect nestedness (see lower right red squares with black centers). (5) A numerical weight is assigned to each unexpected interaction, such that they are weighted based on the relative distance of exceptions to the isocline d_{ij} versus the off-diagonal distance D_{ij}. More details can be found in [74]. **(b)** In NODF, rows and columns are compared pairwise for all pairs of hosts and phages, respectively. The overall nestedness is the normalized sum of the contribution to nestedness made by each pairwise comparison. The pairwise nestedness contribution can be 0, 1, or some intermediate value. The contribution is 0 when there is no overlap between sets of infections (middle left) or when the sets of infections have the same degree (bottom left). The contribution is 1 when one set of infections is a subset of the other set (bottom right). The contribution is at some intermediate value when there is partial overlap between sets of infections, so long as the sets of infections each have a different degree (middle right). These scenarios are noted as null, full, and partial contributions, respectively. More details can be found in [75].

phages infecting hosts of different taxa. We might expect that large taxonomic diversity would ultimately promote modularity. However, taxonomic diversity is, alone, insufficient to explain network patterns. Eco-evolutionary factors governing selection for cross-infection may result in different emergent patterns.

Second, PBINs are associated with the spatiotemporal scale over which they are collected. Recent studies involving systematic cross-infection of phages and hosts from multiple sites, for example, soil and tree-associated communities, suggest that phages preferentially infect hosts from the same site more so than hosts isolated from similar, but 'distant' sites [28,29]. This finding is consistent with biogeographic studies that find that the species composition of communities grows more dissimilar with increasing geographic and environmental distance [30]. Nonetheless, many studies (including [28,29]) also find cross-infections that transcend site and time of isolation. As but one example, a study in marine waters identified positive cross-infection between a phage isolate ϕ19:2 (unknown family, isolated in February 2000 in Øresund surface water, 56°2′N, 12°37′E) with an isolate of *Cellulophaga baltica* denoted OL12A (isolated in April 2005 in Baltic Sea surface water, 56°37′N, 16°45′E). Hence, phage ϕ19:2 is able to infect a host isolated >200 km and 5 years after the initial phage isolation [27]. Many similar examples can be

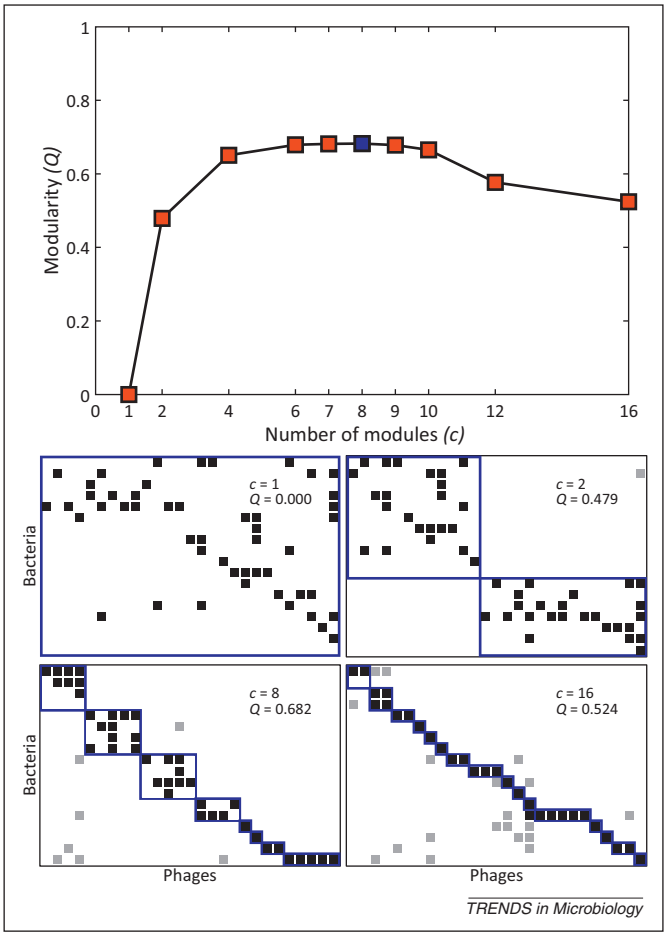

Figure 3. Method to calculate modularity. We utilize the Bipartite, Recursively Induced Modules (BRIM) method for modularity of 'bipartite' networks (i.e., those in which there are two types of agents, like phages and bacteria; [76,77]). BRIM attempts to decompose a matrix into a certain number of modules, *c*. We illustrate how BRIM works using the phage–bacteria infection network (PBIN) of [56] (middle left panel). The optimal number of modules is one in which interactions occur frequently inside the module and infrequently outside the module. We evaluated alternative possibilities, and find the maximal modularity ($Q = 0.682$) occurs when $c = 8$ (top panel). In so doing, we reject $c = 1$ (middle left panel) and $c = 2$ (middle right panel) because interactions within a module occur too infrequently. Similarly, we reject $c = 16$ (bottom right panel) because interactions outside of modules occur too frequently. The finding of $c = 8$ modules (lower right panel) maximizes the extent to which interactions frequently occur within modules and rarely occur outside of modules. In all four matrix panels, black squares denote interactions within a module and gray squares denote interactions outside of a module. In general, the modularity Q spans cases where interactions occur more frequently within modules than expected by chance ($0 < Q \leq 1$), no differently than what would be expected by chance ($Q = 0$), to less frequently within modules than what would have been expected by chance ($-1 \leq Q < 0$).

identified. Infections of hosts via phages that span large spatiotemporal separations could imply strong stabilizing selection of host defense mechanisms and/or the loss of resistance mechanisms [31]. Such infections could also represent instances of fluctuating selection where phages and hosts recur in communities after undergoing severe drops in population abundance or even local extinction [32].

In addition, environmental factors such as nutrient stress, host physiological state, and ambient temperature can affect individual virus–host infections [33]. A prominent example can be found in an analysis of phages infecting strains of *Vibrio* spp. in the Georgia Strait [34]. In this study, Comeau and colleagues showed that the similarity of ecological conditions in which hosts were sampled (e.g., depth, presence of cultivated oysters, and temperature) predicted the variation in susceptibility to infection by phages more so than did distance between sampling sites. Similarly, an experiment using lytic phages of soil *Pseudomonas* [13] showed that most network metrics changed following an artificial alteration of the environment (in this case, changes in resource availability). Hence, environmental conditions affect individual phage–host interactions as well as the overall mode of interaction at the community scale. In the future, it will be important to disentangle the effect that environmental conditions have on community composition, thereby affecting network patterns, from the effect that environmental conditions have on network patterns given the same community composition.

Tracking infection structure in experimental evolution

Complex cross-infection patterns emerge via the interplay of ecological and biogeographic factors (as discussed in the previous section) as well as via coevolutionary processes. Coevolution is a potent driver of taxonomic and functional diversity in natural communities [35]. The ease of use of phage–bacteria systems [36,37] has made them increasingly popular as experimental models for coevolution [29,37–42] and as the basis for theoretical models of the interplay between ecology and evolution [24,43–45].

Indeed, classic genetic studies, including those conducted by Luria, revealed the first glimpses of how phage and bacterial interactions coevolve. Multiple serial transfers of bacteria and phages led to what we would denote as small infection networks ranging from two phages by two hosts to 18 phages by three hosts [46–48]. What was striking is that all of the networks identified by these researchers possessed the same highly nested pattern, despite being generated with different *Escherichia coli* phages (T2, T4, and λ). Phages evolved expanded host ranges and were able to infect past hosts. Likewise, bacteria evolved resistance to the newly evolved phages and maintained resistance to past phages leading to a nested pattern. It is important to note that the nested pattern is most often observed when evaluating cross-infection of hosts and viral strains isolated at different points in an experiment.

This early work has been confirmed by recent studies using serial transfer coevolutionary experiments with *Pseudomonas fluorescens* and associated phage [41] as well as *E. coli* and its phage [42]. However, even though nested evolutionary PBINs are commonly observed in controlled environments that tend to have abundant resources, they may not emerge in low-resource settings. For example, it has been hypothesized that mutations that cause ever-increasing host ranges and defenses would not be selected in depauperate culture conditions if broad-spectrum resistance or host ranges are costly, as classic trade-off theory predicts [49,50]. Instead, one would expect more specialized interactions to evolve since both resistance to past phages (by hosts) and the ability to infect extinct or rare hosts (by phages) would be costly.

Fortunately, many studies on coevolution in more natural continuous culture have been performed on a variety of bacteria-phage pairs (*E. coli* and phages T2, T3, T4, T5, T7, PP01, and λ as well as *P. fluorescens* and φ2) and produced PBINs of a range of sizes (including a few isolates

Box 2. Methods to find significant patterns

The analysis of modularity and nestedness requires a critical assessment of statistical significance, that is, are these patterns signaling ecological and evolutionary drivers or are they dominated by chance? To do so requires comparing the pattern found in a PBIN (as seen in Figure IIa, taken from [80]) to the range of patterns found in a suitable null model. The PBIN depicted here includes $H = 18$ hosts, $P = 8$ phages, and $I = 54$ cross-infections. Two widely utilized null models to measure the statistical significance of patterns in a PBIN are the Bernoulli random network and the probabilistic degree network. Both null models have the same size as the original PBIN and both null models have the same number of interactions, on average, as the original PBIN. However, the null models differ in the assumptions they make in randomizing the location of positive interactions.

The Bernoulli random network null model takes the original matrix and randomly reshuffles positive infections so that the total number of infections is conserved (Figure IIb). The probabilistic degree network null model takes the original matrix and reshuffles positive infections while maintaining the same 'degree distribution', on average, of both phages and hosts while reshuffling (Figure IIc). The 'degree' of a host is the number of phages that infect it (see the numbers associated with each row in panels a–c), whereas the degree of a phage is the number of hosts it infects (see the numbers associated with each column in Figure II). To contrast the two methods, consider the red and blue outlined matrix elements in the original PBIN on the left. In a Bernoulli random network, the probability of a 1 being assigned is: $I/(HP) = 54/(18*8) = 0.375$. Hence, the probability of an infection event in a Bernoulli random network will be the same for all possible interactions. In a probabilistic degree network, the probability of a 1 being assigned to an interaction between host i and phage j is $k_i k_j/(HP)$, where k_i and k_j are the number of phages that infect host i and the number of hosts that phage j can infect, respectively. Hence, the red square in the matrix Figure IIc has a probability of $(5*8)/(18*8) \sim 0.28$, and, the blue square in the matrix in Figure Ic has a probability of $(5*2)/(18*8) \sim 0.07$.

Given this, the significance of the measured value of nestedness, N, and modularity, Q, in the original PBIN can be compared to either of these null models as follows: (i) generate a large number of matrices following the rules of the null model, the inverse of the number of matrices determines the lowest possible calculated significance value; (ii) calculate the nestedness and modularity of the resulting matrices; (iii) compare the true value of N and Q to the distribution of nestedness and modularity values measured from the null models. The choice of the appropriate null model affects whether nestedness and/or modularity will be considered significant. We recommend, as best practice, reporting the algorithm used, significance values, and the choice of the null model.

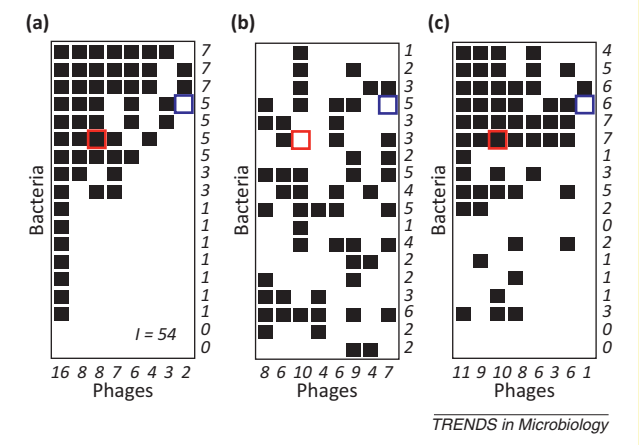

Figure II. Empirical infection compared to null models. **(a)** PBIN. **(b)** Bernoulli random network null model. **(c)** Probabilistic degree network null model.

to a few dozen isolates) to test the idea that specialization is linked to resource supply [12,19,24,40,51]. Even though increased resistance and host range are often found to have fitness costs, virtually all of these studies were at odds with the prediction of specialization, and produced nested PBINs as observed in the earlier studies. One notable exception is *P. fluorescens* coevolving with the phage φ2 that produces both nested and modular PBINs depending on growth conditions [39]. Consistent with the hypothesis that specialization is linked to resource-limited environments, modularity evolves in resource-poor environments where pleiotropic costs of host range resistance mutations are expected to be higher [39]. Further investigation of the link between genetic constraints and environmental factors is warranted.

Resistance mechanisms and environmental drivers act synergistically to determine patterns of cross-infection

Cross-infection at the community scale depends on the underlying genetics of defense and counter-defense mechanisms, but also on the ecological context in which evolution unfolds. The central difficulty in making the link between experimental and ecological studies is the relative paucity of biotic and abiotic diversity in laboratory evolutionary studies versus that commonly found in natural communities.

Host-switching experiments have been proposed to directly evaluate the effect of including larger taxonomic diversity on evolutionary PBINs. For example, mutations within phages are found to expand the viral host range to include new species when phages are cultured on alternating species (φ6 from *Pseudomonas syringe* to *Pseudomonas pseudoalcaligenes* and φ×174 between *Salmonella enterica* and *E. coli*) [22,52,53]. These mutations have negative pleiotropic effects on infecting the original host, though the underlying genetic mechanisms remain unclear. The negative pleiotropic effects drive the evolution of specialist phages that preferentially infect one species over another, that is, they lead to modular PBINs [22,52,53]. Whether or not such a result extends to even greater taxonomic diversity, as would be expected in natural systems, remains unknown.

Finally, it is important to keep in mind that environmental context matters even when a given resistance mechanism can be associated with a certain class of pattern. For example, a number of PBINs with elevated modularity identified in a recent meta-analysis come from communities dominated by *Streptococcus thermophilus* hosts [54–56] (see Figure 4 for an example). *S. thermophilus* possesses the phage-resistance mechanism of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas defense [57–59]. This defense mechanism is predicated on the homology of short sequences in the host genome with corresponding sequences in the infecting virus; hence modularity might be possible if groups of hosts possess distinct sequences that occur exclusively in subsets of sampled phages. Despite examples for which coevolution is apparently driven by selection both for and against CRISPR resistance [60–62], the conditions under which CRISPR-Cas defense predominates in the environment remains unknown. Hence, multiple explanations exist for similar

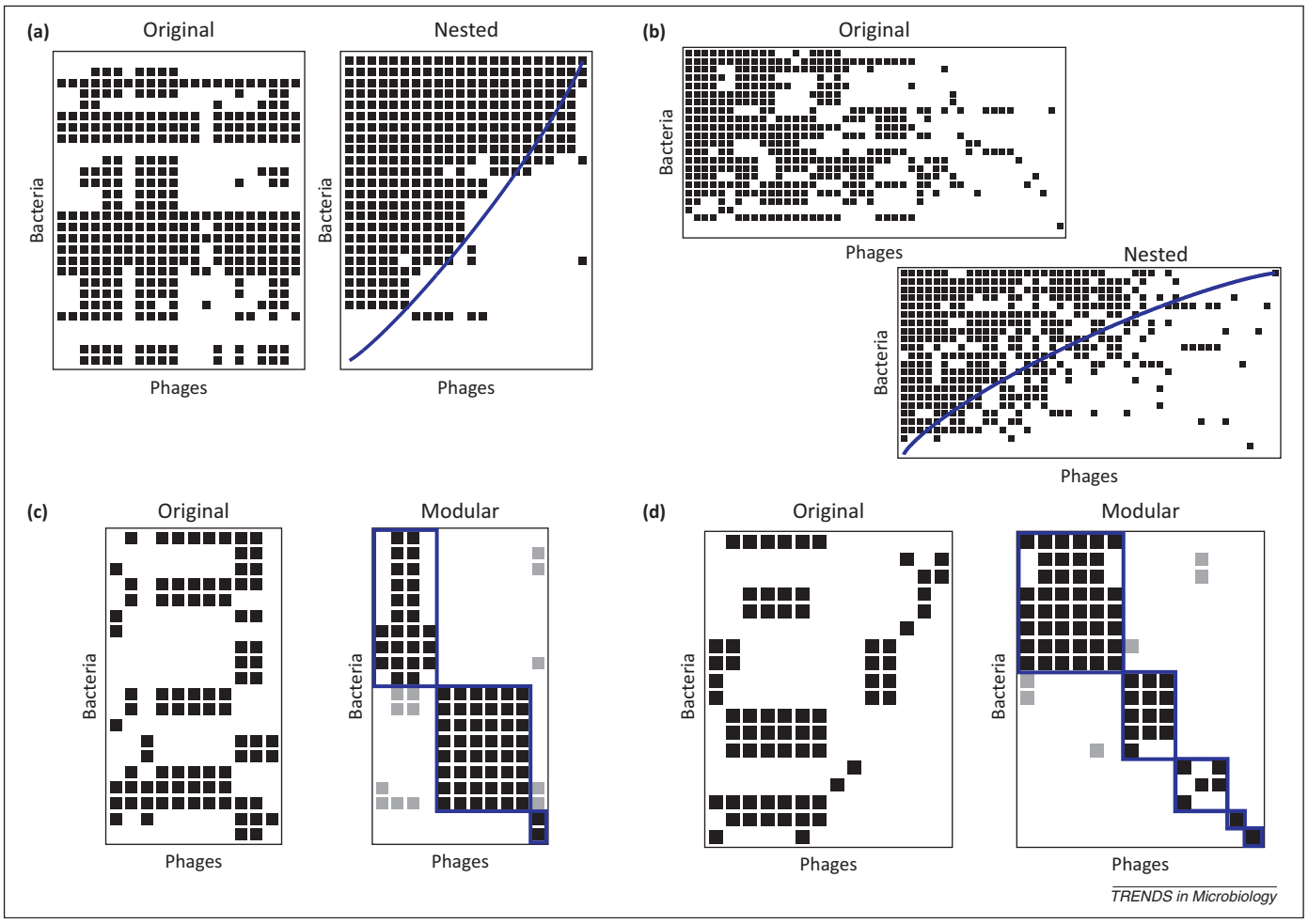

Figure 4. Phage–bacteria infection network (PBIN) structure revealed through network analysis. Two examples of PBINs that have significantly elevated nestedness, not immediately apparent from the original matrix with **(a)** from [78] and **(b)** from [27]. Two examples of PBINs that have significantly elevated modularity, not immediately apparent from the original matrix with **(c)** from [55] and **(d)** from [79]. In each case, the left panel represents the original format and the right panel represents the structure revealed by network analysis.

patterns – in this case modularity – highlighting the challenge and need to disentangle genetic and environmental drivers when considering the basis for and effects of complex cross-infection.

Directions for future research

As we have shown, the use of network approaches has enabled the discovery of more complex structures within PBINs than previously appreciated. Non-random patterns have been identified within interaction networks in other biological domains including food webs, plant–pollinator networks, and metabolic networks [14–17,63]. However, as in these other domains, the question remains: how does measuring and quantifying a PBIN help advance our understanding of key biological questions? We discuss three directions for future research.

Predicting who infects whom in microbial communities

There are many factors that influence which phages can infect and exploit which bacterial hosts in a given environment. Thus far, the analysis of PBINs has focused largely on groups of bacteria and phage that are relatively closely related, for example, most of the PBINs that have been analyzed include congeneric bacteria. We are not aware of studies providing definitive evidence for whether there are critical scales in phylogenetic distance that lead to rapid drop-offs in cross-infection, for example, can the same phage infect bacteria from different families, orders, or classes? A reductionist view would build from the bottom up, that is, investigating the links between the molecular details of multiple classes of phage-resistance mechanisms [21,64]. This is worth pursuing but necessarily limited by our lack of genetic models for all but a few systems. Moreover, we should anticipate that environmental drivers of cross-infection can manifest themselves at the community scale, as noted above. Unlike many laboratory systems, bacterial hosts interact with multiple strains of phages in natural environments. These phages may be intracellular competitors [65] and/or be beneficial to the host [66]. The bacteria–phage relationship may itself be intertwined with other ecological interactions, such as competition and defense against grazing. We suggest that the integration of molecular models of resistance with evolutionary models of selective pressures will be necessary to predict the outcome of infection, whether at the cellular or community scale. Such integrated approaches are also likely to shed light on taxonomic or other signatures that determine the likelihood of cross-infection.

Effects of PBINs on microbial community structure and function

There is increasing evidence that PBIN structure is associated with ecological properties of microbial communities and ecosystems. A combined theoretical and experimental study considered how nested versus modular networks yield different patterns of biodiversity across a gradient of resource availability [24]. PBINs of modular patterns yielded an increase in bacterial diversity with resource availability, while nested patterns show a unimodal relationship with resource availability. Interestingly, with little awareness of this network approach, cheese and yogurt manufacturers have favored the creation of modular microbial communities rather than nested communities in order to create more stable and productive bacterial cultures. This practice, known as 'phage unrelated' culturing, involves co-culturing bacteria that have distinct phage sensitivity profiles in order to create robust bacterial communities [67]. Along these lines, laboratory experiments have shown that by introducing resistant hosts into microbial host–parasite systems, the population growth of the host is improved [68], and population dynamics and environmental stoichiometry of the systems are altered [69]. While some progress has been made in understanding how PBINs are associated with the ecology of microbial communities, more work is required where PBINs are directly manipulated and suites of ecological properties monitored to identify possible causative mechanisms and/or feedbacks between structure and function. Ecosystem properties to examine should include, for example, the storage, export and regeneration of carbon and nutrients as mediated by viruses, that is, the 'viral shunt' [5,6]. Doing so will further our understanding of how cross-infection patterns can lead to differential levels of antagonism or even benefits to bacteria in complex communities.

Technological innovations to study phage–bacteria interactions

Existing PBINs from microbial communities include cross-infections of tens of phage and host strains. However these communities have many orders of magnitude more individuals and, likely, orders of magnitude more strain types. Isolation-based approaches will, by necessity, provide incomplete information on the interactions between environmental viruses and their hosts. Several emerging, new methods promise to help us delve deeper into various aspects of virus–host interactions in myriad ways. First, whole genome re-sequencing now allows the identification of resistance mutations in phage–host systems whose genetics is poorly understood, a common issue for ecologically important systems [70]. Second, single virus genomics promises high-throughput exploration of environmental phage genome sequence space without shearing community DNA into shotgun fragments such that we can reassemble significant chunks of genomic sequences if not whole genomes [71]. Third, where sequence data are already available for a given environment, researchers can now develop primers or probes to match uncultured viruses to their hosts in wild populations by screening for colocalized virus and host signals using microfluidic digital PCR [72]. Fourth, the genomic analysis of CRISPR-associated sequences may provide a direct method to identify prior interactions with viruses in complex communities [60–62]. Finally, a new 'viral-tagging' method uses a fluorescent stain to label wild viruses in combination with flow cytometry to provide a sequence-independent means for experimentally linking wild viruses *en masse* (i.e., tens of thousands at a time) to a cultured host cell [73]. Hence, viral tagging provides a means to examine the PBIN of a microbial community in a given environment, in contrast to prior efforts that aggregate cultured isolates of bacterial and viral isolates from a metacommunity. Together these methods promise rapid and significant advances in our ability to generate virus–host interaction data, including moving from PBINs with qualitative information to PBINs with (semi)quantitative information. These developments necessitate new analytical and theoretical frameworks to maximize their potential. A subset of the current authors (T.P., C.O.F., S.V., and J.S.W.) are part of a consortium of developers of BiWeb (https://github.com/tpoisot/BiWeb), a multi-platform software tool for the analysis of bipartite networks, of which PBINs are one prominent example. This is but one example of the type of theoretical advance that will be required to enable microbiologists to examine and analyze functional interaction data.

Concluding remarks

We have synthesized and reviewed the study of complex patterns of infection between phages and bacteria via the unifying concept of a PBIN. We find that non-random patterns, in particular, nestedness and modularity are repeatedly observed. Although genetic models of coevolution exist to explain both classes of patterns in simple, low-diversity communities, there is no unified theory of coevolution that can explain the emergence of complex interaction patterns in diverse communities. In this review, we have focused on the antagonistic mode of interaction between phages and bacteria. Looking forward, there is a need to extend network methodologies and tools to the study of viruses of currently underrepresented bacterial hosts, viruses of algae and of archaea, as well as viruses with alternate infection strategies (e.g., lysogeny). Novel culture-independent methods and novel computational tools will be essential to connect genomic, functional,

Box 3. Outstanding questions

- What properties of PBINs influence microbial community assembly and ecosystem functioning?
- Can properties of PBINs be used to predict the fate of bacteria and phage populations?
- How do environmental perturbations, such as climate change, alter PBIN structure?
- Are some forms of interaction networks more resilient than others to biotic and abiotic perturbations?
- What underlies the finding of nestedness in PBINs? Are nested patterns a byproduct of the coevolutionary process? Is there an ecological force selecting for this structure, or is nestedness a biophysical necessity?
- What forms do PBINs take when sampled over larger phylogenetic and geographic scales?
- What is the structure of interactions among bacteria and phages using culture-independent methods? Do they differ from patterns identified using culture-based approaches?

and ecological information in predicting cross-infection at the community scale (Box 3). Improving the ability to quantify and predict viral infection of hosts constitutes a key challenge in assessing the functional effects of viruses in the environment.

Acknowledgments

This research was supported by grants from the James S. McDonnell Foundation and the Burroughs Wellcome Fund (to J.S.W.), the MELS and FQRNT (to T.P.), the Gordon and Betty Moore Foundation (to M.B.S.), and the Agence National de Recherche ('EvolStress') and the James S. McDonnell Foundation (to M.E.H.). The authors declare no competing interests.

References

1 Cairns, J. *et al.* (2000) *Phage and the Origins of Molecular Biology*, Cold Spring Harbor Laboratory Press
2 Rohwer, F. and Thurber, R.V. (2009) Viruses manipulate the marine environment. *Nature* 459, 207–212
3 Suttle, C.A. (2007) Marine viruses – major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812
4 Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181
5 Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548
6 Wilhelm, S.W. and Suttle, C.A. (1999) Viruses and nutrient cycles in the sea. *Bioscience* 49, 781–788
7 Sano, E. *et al.* (2004) Movement of viruses between biomes. *Appl. Environ. Microbiol.* 70, 5842–5846
8 Nelson, E.J. *et al.* (2009) Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat. Rev. Micro* 7, 693–702
9 Reyes, A. *et al.* (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338
10 Rho, M. *et al.* (2012) Diverse CRISPRs evolving in human microbiomes. *PLoS Genet.* 8, e1002441
11 Sullivan, M.B. *et al.* (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* 424, 1047–1051
12 Flores, C.O. *et al.* (2011) Statistical structure of host–phage interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, E288–E297
13 Poisot, T. *et al.* (2011) Resource availability affects the structure of a natural bacteria–bacteriophage community. *Biol. Lett.* 7, 201–204
14 Pascual, M. and Dunne, J.A. (2006) *Ecological Networks: Linking Structure to Dynamics in Food Webs*, Oxford University Press
15 Bascompte, J. *et al.* (2003) The nested assembly of plant–animal mutualistic networks. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9383–9387
16 Bascompte, J. (2009) Disentangling the web of life. *Science* 325, 416–419
17 Newman, M. (2010) *Networks: An Introduction*, Oxford University Press
18 Fortuna, M.A. *et al.* (2010) Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* 79, 811–817
19 Lenski, R.E. and Levin, B.R. (1985) Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *Am. Nat.* 125, 585–602
20 Agrawal, A.F. and Lively, C.M. (2003) Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. *Proc. R. Soc. Lond. B: Biol. Sci.* 270, 323–334
21 Stern, A. and Sorek, R. (2011) The phage–host arms race: shaping the evolution of microbes. *Bioessays* 33, 43–51
22 Duffy, S. *et al.* (2006) Pleiotropic costs of niche expansion in the RNA bacteriophage Φ6. *Genetics* 172, 751–757
23 Inouye, B. and Stinchcombe, J.R. (2001) Relationships between ecological interaction modifications and diffuse coevolution: similarities, differences, and causal links. *Oikos* 95, 353–360
24 Forde, S.E. *et al.* (2008) Understanding the limits to generalizability of experimental evolutionary models. *Nature* 455, 220–223
25 Leibold, M.A. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7, 601–613
26 Wichels, A. *et al.* (1998) Bacteriophage diversity in the North Sea. *Appl. Environ. Microbiol.* 64, 4128–4133
27 Holmfeldt, K. *et al.* (2007) Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. *Appl. Environ. Microbiol.* 73, 6730–6739
28 Koskella, B. *et al.* (2011) Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. *Am. Nat.* 177, 440–451
29 Vos, M. *et al.* (2009) Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* 325, 833
30 Hanson, C.A. *et al.* (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10, 497–506
31 Waterbury, J. and Valois, F. (1993) Resistance to co-occurring phages enables marine Synechococcus communities to coexist with cyanophages abundant in seawater. *Appl. Environ. Microb.* 59, 3393–3399
32 Hall, A.R. *et al.* (2011) Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecol. Lett.* 14, 635–642
33 Wilson, W.H. *et al.* (1996) The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp WH7803. *J. Phycol.* 32, 506–516
34 Comeau, A.M. *et al.* (2006) Genetic richness of vibriophages isolated in a coastal environment. *Environ. Microbiol.* 8, 1164–1176
35 Thompson, J.N. (2005) *The Geographic Mosaic of Coevolution*, University of Chicago Press
36 Jessup, C.M. *et al.* (2004) Big questions, small worlds: microbial model systems in ecology. *Trends Ecol. Evol.* 19, 189–197
37 Buckling, A. *et al.* (2009) The *Beagle* in a bottle. *Nature* 457, 824–829
38 Buckling, A. and Rainey, P.B. (2002) Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. R. Soc. Lond. B: Biol. Sci.* 269, 931–936
39 Gómez, P. and Buckling, A. (2011) Bacteria–phage antagonistic coevolution in soil. *Science* 332, 106–109
40 Paterson, S. *et al.* (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464, 275–278
41 Poullain, V. *et al.* (2008) The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. *Evolution* 62, 1–11
42 Meyer, J.R. *et al.* (2012) Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335, 428–432
43 Weitz, J.S. *et al.* (2005) Coevolutionary arms races between bacteria and bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9535–9540
44 Childs, L.M. *et al.* (2012) Multi-scale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. *Evolution* 66, 2015–2029
45 Weinberger, A.D. *et al.* (2012) Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* 8, e1002475
46 Appleyard, R. *et al.* (1956) Mutation to extended host range and the occurrence of phenotypic mixing in the temperate coliphage lambda. *Virology* 2, 565–574
47 Luria, S.E. (1945) Mutations of bacterial viruses affecting their host range. *Genetics* 30, 84–99
48 Hershey, A.D. (1946) Spontaneous mutations in bacterial viruses. *Cold Spring Harb. Symp. Quant. Biol.* 11, 67–77
49 Bowen, L. and Vuren, D.V. (1997) Insular endemic plants lack defenses against herbivores. *Conserv. Biol.* 11, 1249–1254
50 Brodie, E.D. (1991) Evolutionary response of predators to dangerous prey: reduction of toxicity of newts and resistance of garter snakes in island populations. *Evolution* 45, 221–224
51 Mizoguchi, K. *et al.* (2003) Coevolution of bacteriophage PP01 and *Escherichia coli* O157:H7 in continuous culture. *Appl. Environ. Microbiol.* 69, 170–176
52 Crill, W.D. *et al.* (2000) Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154, 27–37
53 Ferris, M.T. *et al.* (2007) High frequency of mutations that expand the host range of an RNA virus. *Genetics* 176, 1013–1022
54 Duplessis, M. and Moineau, S. (2001) Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. *Mol. Microbiol.* 41, 325–336
55 Quiberoni, A. *et al.* (2003) Comparative analysis of *Streptococcus thermophilus* bacteriophages isolated from a yogurt industrial plant. *Food Microbiol.* 20, 461–469

56 Zinno, P. *et al.* (2010) Characterization of *Streptococcus thermophilus* lytic bacteriophages from mozzarella cheese plants. *Int. J. Food Microbiol.* 138, 137–144
57 Deveau, H. *et al.* (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* 64, 475–493
58 Marraffini, L.A. and Sontheimer, E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11, 181–190
59 Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170
60 Andersson, A.F. and Banfield, J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050
61 Heidelberg, J.F. *et al.* (2009) Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS ONE* 4, e4169
62 Held, N.L. and Whitaker, R.J. (2009) Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol.* 11, 457–466
63 Guimerá, R. and Amaral, L.A.N. (2005) Functional cartography of complex metabolic networks. *Nature* 433, 895–900
64 Labrie, S.J. *et al.* (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327
65 Refardt, D. (2011) Within-host competition determines reproductive success of temperate bacteriophages. *ISME J.* 5, 1451–1460
66 Oliver, K.M. *et al.* (2009) Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* 325, 992–994
67 Sturino, J.M. and Klaenhammer, T.R. (2006) Engineered bacteriophage-defence systems in bioprocessing. *Nat. Rev. Microbiol.* 4, 395–404
68 Bohannan, B.J.M. and Lenski, R.E. (1999) Effect of prey heterogeneity on the response of a model food chain to resource enrichment. *Am. Nat.* 153, 73–82
69 Lennon, J.T. and Martiny, J.B.H. (2008) Rapid evolution buffers ecosystem impacts of viruses in a microbial food web. *Ecol. Lett.* 11, 1178–1188
70 Avrani, S. *et al.* (2011) Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474, 604–608
71 Allen, L.Z. *et al.* (2011) Single virus genomics: a new tool for virus discovery. *PLoS ONE* 6, e17722
72 Tadmor, A.D. *et al.* (2011) Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* 333, 58–62
73 Deng, L. *et al.* (2012) Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral-tagging. *MBio* 3, e00373–12
74 Atmar, W. and Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96, 373–382
75 Almeida-Neto, M. *et al.* (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117, 1–13
76 Barber, M. (2007) Modularity and community detection in bipartite networks. *Phys. Rev. E* 76, 066102
77 Guimerà, R. *et al.* (2007) Module identification in bipartite and directed networks. *Phys. Rev. E* 76, 036102
78 Stenholm, A.R. *et al.* (2008) Isolation and characterization of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*. *Appl. Environ. Microbiol.* 74, 4070–4078
79 Miklič, A. and Rogelj, I. (2003) Characterization of lactococcal bacteriophages isolated from Slovenian dairies. *Int. J. Food Sci. Technol.* 38, 305–311
80 Capparelli, R. *et al.* (2010) Bacteriophage therapy of *Salmonella enterica*: a fresh appraisal of bacteriophage therapy. *J. Infect. Dis.* 201, 52–61