Received: 14 September 2023 | Revised: 15 November 2023 | Accepted: 21 December 2023

DOI: 10.1111/ele.14383

LETTER

ECOLOGY LETTERS cnrs WILEY

The dimensionality of infection networks among viruses infecting microbial eukaryotes and bacteria

Kyle F. Edwards | **Colleen Hayward**

Department of Oceanography, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA

Correspondence
Kyle F. Edwards, University of Hawai'i at Mānoa, 1000 Pope Road, Honolulu, HI 96822, USA.
Email: kfe@hawaii.edu

Funding information
Division of Ocean Sciences, Grant/Award Number: 2129697; Simons Foundation; Office of Experimental Program to Stimulate Competitive Research, Grant/Award Number: RII Track-2 FEC 1736030

Editor: Ulrich Brose

Abstract

Diverse viruses and their hosts are interconnected through complex networks of infection, which are thought to influence ecological and evolutionary processes, but the principles underlying infection network structure are not well understood. Here we focus on network dimensionality and how it varies across 37 networks of viruses infecting eukaryotic phytoplankton and bacteria. We find that dimensionality is often strikingly low, with most networks being one- or two-dimensional, although dimensionality increases with network richness, suggesting that the true dimensionality of natural systems is higher. Low-dimensional networks generally exhibit a mixture of host partitioning among viruses and nestedness of host ranges. Networks of bacteria-infecting and eukaryote-infecting viruses possess comparable distributions of dimensionality and prevalence of nestedness, indicating that fundamentals of network structure are similar among domains of life and different viral lineages. The relative simplicity of many infection networks suggests that coevolutionary dynamics are often driven by a modest number of underlying mechanisms.

KEYWORDS

coevolution, host range, modularity, nestedness, phage, phytoplankton, prasinovirus, resource partitioning

INTRODUCTION

Viruses are pervasive molecular parasites that influence the ecology and evolution of all life forms. Studies of diverse viruses infecting multiple strains of a host organism often document complex patterns of cross-infection (e.g., Baudoux et al., 2015; Clerissi et al., 2012). The distinct but overlapping host ranges among a set of viruses can be analysed as an infection network, where the virus and host strains are the network nodes and the network edges represent successful infection of a host strain by a viral strain (Weitz et al., 2013). The structure of infection networks is predicted to affect the population dynamics and diversity of virus and host genotypes (Beckett & Williams, 2013; Jover et al., 2013; Pilosof et al., 2020; Poisot et al., 2011). In addition, the structure of these networks presumably derives from the (co)evolutionary processes between viruses and their hosts, such as arms races or fluctuating selection dynamics, and analysis of network structure should help reveal how these processes operate (Beckett & Williams, 2013; Pilosof et al., 2020; Weitz et al., 2013). At the same time, multiple types of coevolutionary dynamics may be associated with similar network structures (Gupta et al., 2022; Gurney et al., 2017), which challenges our ability to derive mechanistic insight from network patterns. Nonetheless, better understanding network structure and how it varies across different virus and host taxa should help illuminate virus-host coevolution and the ecological consequences.

Studies of virus-bacteria infection networks have quantified statistical evidence for network nestedness and modularity (e.g., Flores et al., 2011; Gurney et al., 2017; Pilosof et al., 2020; Poisot et al., 2013). Nestedness can emerge when the host ranges of viruses with narrower ranges are a subset of the host ranges of viruses with broader ranges, and statistically significant nestedness is common in virus-bacteria networks (Flores et al., 2011). A variety of mechanisms could produce nestedness (Mariani et al. 2019), such as tradeoffs between viral host range breadth and infection efficiency, which can permit stable coexistence of viruses with nested host ranges (Jover et al., 2013), and coevolutionary arms races

that cause successively broader host ranges to evolve, as earlier viral lineages with narrower host ranges are driven extinct (Brockhurst & Koskella, 2013). Infection networks can also possess signals of modularity, where nodes in modules of the network interact with each other at a higher frequency than they interact with nodes in other modules. Modularity may be more common when the network being analysed includes a relatively broad phylogenetic diversity of hosts, such that the modules represent different host clades and the sets of viruses capable of infecting them (Beckett & Williams, 2013; Flores et al., 2013).

A measure of network structure that has not yet been explored in viral infection networks is dimensionality. Eklöf et al. (2013) quantified the dimensionality of ecological networks in terms of a multivariate embedding space, where dimensionality is the minimum number of dimensions needed to order nodes in the network such that pairs of nodes are linked if they fall within a certain volume of this space. Infection networks can be analysed in terms of dimensionality from either the host or virus point of view, because infection networks are bipartite: each edge in the network connects a node in one group (the viruses) to a node in the other group (the hosts). For example, a network could be one-dimensional from the host direction, if hosts can be ordered along an axis such that the host range of each virus consists of a continuous interval along the axis (Figure 1a,c,e,g,h). Likewise, a network could be one-dimensional from the virus direction, if viruses can be ordered along an axis such that each host is infected by a contiguous set of viruses. A two-dimensional network would need two dimensions to order the hosts, such that the host range of each virus includes all hosts within a box in those dimensions (Figure 1b,d,f), and so on. Even if an observed infection network is relatively high dimensional, it may be possible to arrange the network in 1–2 dimensions with relatively few errors, and visualization/analysis of the network in this low-dimensional space may yield significant insights, analogous to other dimensionality reduction techniques such as principal component analysis.

Eklöf et al. (2013) analysed the dimensionality of 200 ecological networks, including food webs, plant-pollinator networks, and host–parasite networks. They found that dimensionality increased with the number of links, but on average dimensionality was fairly low (mean 2.7). Low or moderate dimensionality of ecological networks may occur if the interactions, such as who-eats-whom in a food web, are determined by a small number of traits (Eklöf et al., 2013). For example, if predator–prey interactions are determined solely by body size then all feeding links can be ordered along a dimension representing size (e.g., Ward et al., 2012). An implication of low dimensionality is that explaining and modelling network structure and its dynamics may be relatively tractable, if the key structuring traits can be identified. Eklöf et al. (2013) did not analyse viral infection networks, and for viruses, it may be less likely that a small number of traits determines network patterns, due to the complexity of host-virus interactions. For example, in *Escherichia coli* a variety of single nucleotide changes can confer host resistance to phage λ (Chatterjee & Rothenberg, 2012), and hosts can employ multiple resistance mechanisms, leading to multifaceted coevolutionary arms races (Burmeister et al., 2021). However, whether the potential complexity of host-virus coevolution leads to complex, high-dimensional networks in natural populations is currently unknown. Quantifying the dimensionality of infection networks, and how viral host ranges are organized along different dimensions, should yield important insights into how these networks are structured and guide investigations into the underlying mechanisms.

Combining analyses of dimensionality with other metrics such as nestedness and modularity may be particularly helpful for understanding the mechanisms underlying dimensionality. For example, strong nestedness may lead to low dimensionality (Figure 1a,b), but alternative patterns of network structure could also underlie low-dimensional infection networks. In particular, competition among viruses for hosts could select for resource partitioning of hosts based on the host traits that affect successful viral infection; we will refer to this as 'host partitioning'. Host partitioning could lead to low-dimensional network structure without strong patterns of nestedness or modularity, if the key axes of host phenotypic variation are small in number (Figure 1c,d). This form of host range evolution has been modelled as an imperfect lock-and-key, intended to represent the evolution of specificity in receptor-binding proteins (Beckett & Williams, 2013; Weitz et al., 2005), although host ranges could also be partitioned based on infection processes downstream of receptor binding. Host partitioning among viruses could occur simultaneously with other forms of viral diversity, such as generalist/specialist host range tradeoffs, or transient diversity associated with arms race dynamics, leading to host range patterns that combine nestedness with host partitioning (Figure 1e,f). Modularity may also emerge along axes of host partitioning as initially overlapping clusters of hosts and viruses diverge over time (Figure 1g,h).

Viruses are extremely diverse, with genome types and replication modes not found in cellular life, and many major viral lineages likely arose independently, some before the last universal cellular ancestor (Koonin et al., 2023). It is not clear whether viruses that infect different domains of life, or viruses from different ancient lineages, tend to have distinct eco-evolutionary features. For example, infection network structure may vary across host domains and the distinct virus lineages infecting them, reflecting different mechanisms of host-virus interaction, or network structure may be generally similar, reflecting common constraints that guide diverse viruses down comparable evolutionary paths. In this study, we compile and analyse

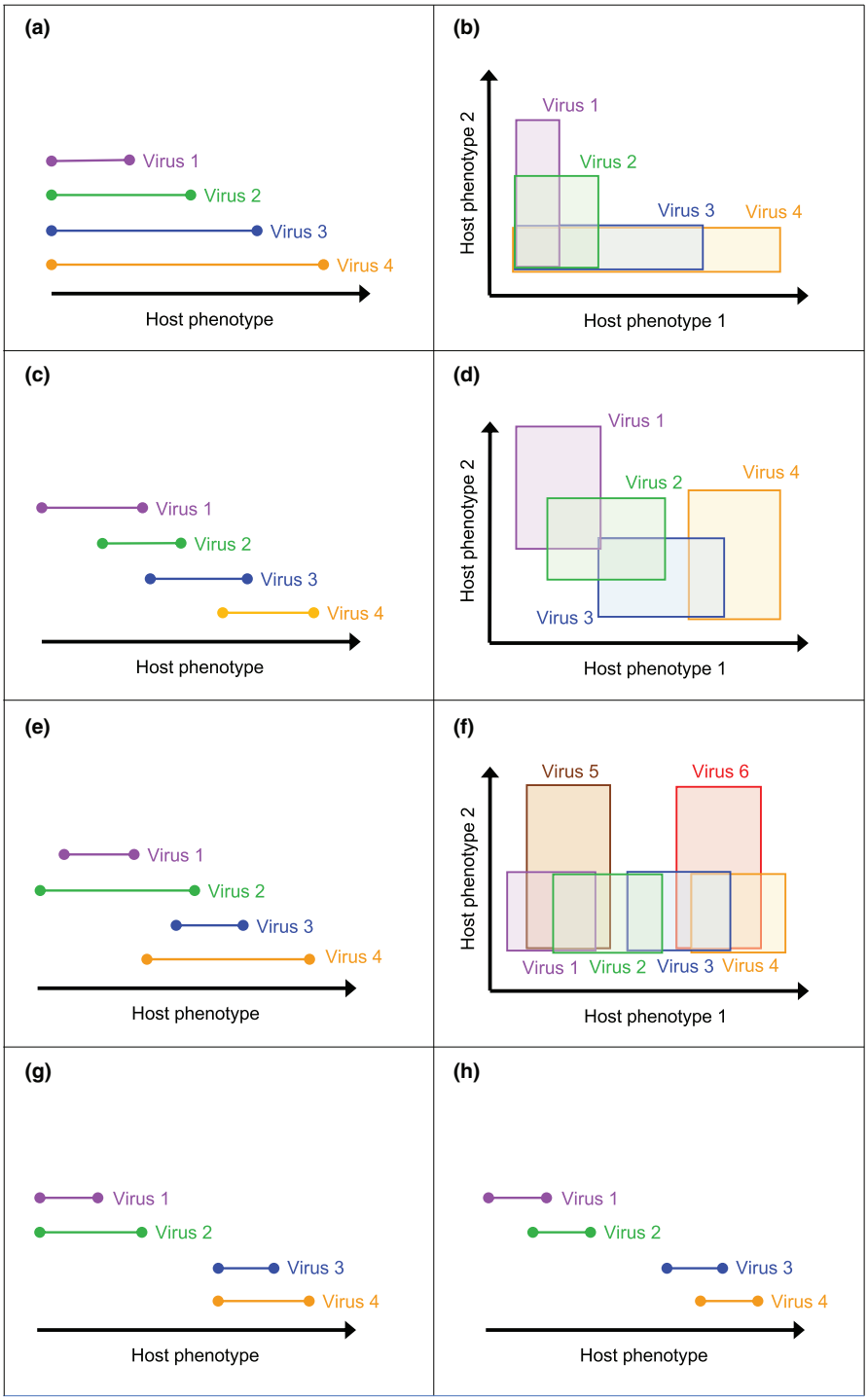

FIGURE 1 Examples of how different viral host range patterns can lead to one- or two-dimensional infection networks. In the 1D cases, the axis represents host phenotypic differences that are ordered in one dimension, and the coloured segments represent viral host ranges, such that a host strain with a phenotype that falls within a particular segment will be infected by that virus. In the 2D cases, the axes represent two distinct aspects of the host phenotype, and the coloured boxes represent viral host ranges, such that a host strain with a phenotype that falls within a particular box can be infected by that virus. (a) Host ranges exhibiting nestedness in 1D, such that smaller host ranges are always nested within broader host ranges. (b) Host ranges in 2D where range positions along each axis are nested. (c) Host partitioning in 1D, such that host ranges are distinct and may overlap, but do not exhibit nestedness. (d) Host partitioning in 2D. (e) Host ranges that can be arranged in 1D and exhibit a combination of partitioning and nestedness. (f) Combined host partitioning and nestedness in 2D. (g) Host ranges in 1D that combine modularity (two non-overlapping pairs of host ranges) and nestedness (within modules, the narrower host range is nested within the broader). (h) Host ranges in 1D that combine modularity and partitioning (within modules, host ranges are distinct but not nested).

11 infection networks of viruses infecting marine eukaryotic phytoplankton, including dsDNA, ssDNA, and ssRNA viruses. Eukaryotic phytoplankton are the dominant primary producers in the ocean, and their viruses greatly influence mortality, diversity and biogeochemistry (Suttle, 2005). We also reanalyse 26

infection networks of phage-infecting bacteria from diverse environments (Flores et al., 2011). Our analysis has several goals: assess the dimensionality of viral infection networks; consider what insights can be gleaned by comparing dimensionality, nestedness, and modularity; and compare network structure between virus-bacteria and virus-eukaryote networks. In total, these comparisons allow us to ask whether network dimensionality yields new insights into infection network structure and whether networks are structured in similar ways across viral lineages and domains of cellular life.

METHODS

Network compilation

We surveyed the literature for cross-infection matrices reported for viruses of eukaryotic phytoplankton. We found 11 matrices, including 3 involving prasinophytes (1 *Micromonas*, 1 *Ostreococcus*, and 1 both), 3 involving the raphidophyte *Heterosigma*, 3 involving diatoms (2 *Chaetoceros*, 1 *Rhizosolenia*), and 2 involving haptophytes (1 *Emiliania*, 1 *Phaeocystis*) (File S2; Table S1). Most of the viruses in these studies were dsDNA viruses, but the three diatom studies included one on ssDNA viruses (Tomaru et al. 2008), one on ssRNA viruses (Nagasaki et al., 2004), and one with both ssDNA and ssRNA viruses (Tomaru et al., 2011). The number of host strains in these studies varied from 6 to 79, the number of virus strains varied from 9 to 44, and the number of virus-host links varied from 29 to 784. The strain diversity in all of these studies represents natural diversity, i.e., the strains did not result from laboratory (co)evolution experiments.

To compare network structure between eukaryote-infecting and bacteria-infecting viruses we re-analysed the phage-bacteria infection matrices previously compiled by Flores et al. (2011). We excluded matrices that did not have at least 15 total strains (virus + host), or that did not have at least 4 strains of both virus and host. These were excluded because preliminary analyses indicated that low strain diversity led to low statistical power for distinguishing network patterns from a null model. The number of included matrices from Flores et al. (2011) is 26. For all matrices we coded whether the matrix included multiple kinds of virus—for the phage networks this meant multiple of the traditionally defined 'families' Myoviridae, Siphoviridae, and Podoviridae; for the eukaryote-infecting viruses, this meant the one study that included both ssDNA and ssRNA viruses. Finally, all networks were coded as to whether the strain diversity represented natural diversity, or experimentally generated diversity (6 of the phage networks included strains that were experimentally evolved).

Dimensionality algorithm

Following Eklöf et al. (2013), we used an iterative algorithm to sort the 37 infection matrices, with the goal of defining the minimum dimensionality of each infection network, and quantifying how well each network can be described by a single dimension. Infection networks are bipartite, because each edge in the matrix connects one part of the network (the set of viruses) to the second part of the network (the set of hosts). Therefore, when analysing network dimensionality the network can be sorted from the host direction or from the virus direction. To describe the algorithm, we will focus on the one-dimensional case where hosts are sorted. At the start, an infection matrix possesses an initial arrangement such that there are N total hosts arranged at positions 1:N along the host axis of the matrix. For each virus, we can define an interval [i,j] such that all hosts infected by that virus are between positions i and j (including hosts i and j). However, there may be hosts between i and j that are *not* infected by the virus, and we refer to the non-infected hosts in [i,j] as *errors*. The goal of the algorithm is to rearrange the positions of the hosts to minimize the number of errors. The algorithm proceeds by choosing two hosts and swapping their positions on the host axis. If the new host arrangement results in fewer total errors, the new arrangement is kept; otherwise, the previous arrangement is retained. The algorithm alternates between choosing two adjacent hosts to swap, and choosing two hosts at random positions to swap—this provides both small and large changes in overall host arrangement. If an arrangement with zero errors is found, the algorithm stops. If there are zero errors, then the interval [i,j] of each virus includes all hosts in its host range and no hosts that are not in its host range. A network that can be perfectly sorted in one dimension is referred to as an 'interval' network in graph theory (Eklöf et al., 2013).

The algorithm proceeds for 10,000 steps (if a zero-error solution is not found) before stopping. The procedure is repeated, with 25 total initial host configurations. The first configuration sorts host by degree (the total number of viruses that can infect the host), while the remaining 24 configurations are chosen at random. Extensive preliminary tests with diverse networks found that using more than 25 initial states, or more than 10,000 steps, did not improve the best solution (i.e., the host arrangement with the fewest number of errors). We also tested a stochastic algorithm that occasionally accepts worse arrangements (i.e., an increase in errors), in order to avoid local optima, but this did not lead to better solutions. Therefore, although the nature of the algorithm precludes us from being certain that the best solution in one dimension has been found, it is likely that the solutions we report are the best possible. It should be noted that there may be more than one solution with the same number of errors, particularly if the network contains true modules. For example, if there is a set A of hosts that can

only be infected by set α of viruses and a set B of hosts that can only be infected by a set β of viruses, then host set A could be placed to the left or the right of host set B without changing the number of errors.

Each network was sorted from both the host and virus directions, with results recorded separately. If the network is sorted from the virus direction the goal is to obtain an arrangement of viruses such that each host is infected by all viruses in its interval $[i,j]$. In other words, each host has a *susceptibility range* that is represented by a contiguous set of viruses along the sorted virus axis. An infection network can be one-dimensional (interval) from the host direction, the virus direction, or both.

If a network could not be perfectly sorted in one dimension, from either the host or virus direction, we repeated the algorithm using two dimensions, and so on with an increasing number of dimensions until the minimum dimensionality was found. To illustrate the procedure in >1 dimension, we describe host sorting in two dimensions. At the start, each host is given a location from 1:N along two axes; these coordinates can be denoted (x_n, y_n) for host n along the two axes x and y. Each virus possesses a box, $[x_i,y_i]\times[x_j,y_j]$, the smallest box containing all hosts that a virus can infect. Any hosts within this box that the virus cannot infect are defined as errors. The algorithm proceeds by randomly picking the first or second axis, swapping host positions along the chosen axis, and calculating the # errors of the new arrangement, as described above for a single dimension. To conceptualize the meaning of >1 dimensional network structure, it can be helpful to think of the axes as putative host traits (or virus traits, when sorting viruses). If a network can be sorted in two host dimensions with zero errors, it means that each virus has a box such that it only infects hosts with traits that meet two conditions: trait values between x_i and x_j on the first trait axis, and values between y_i and y_j on the second trait axis. R code for finding the minimum dimensionality of an infection matrix is provided in File S1.

Null model analysis

Networks that are relatively small or have low connectance may be sortable into one dimension even if the edges are randomly arranged. Therefore, we used a null model to ask whether the 'fit' of each network to 1D is better than expected by chance. We used a probabilistic degree network (Flores et al., 2013) where the ability of virus m to infect host n is assigned with probability $p_{mn} = \frac{1}{2}\left(\frac{k_n}{V} + \frac{d_m}{H}\right)$, where the degree k_n is the number of viruses that infect host n, the degree d_m is the number of hosts infected by virus m, V is the total number of viruses in the network, and H is the total number of hosts in the network. This model maintains, on average, the degree for each host and virus in the network, while allowing the arrangement of host-virus interactions to vary randomly within those constraints. For each empirical network, 500 random networks were generated, and the sorting algorithm defined above was run for each network. This generated a null distribution of the # of errors when sorting random networks in one dimension. The null distribution was used to calculate the probability of observing the # of errors in the empirical network, as well as the mean, SD, and 95% interval of # errors in the null distribution. We also implemented the 'EE' null model that maintains the total number of species occurrences in the matrix, but allows both row and column totals to vary freely (Ulrich & Gotelli, 2007). This model produced qualitatively similar results to the probabilistic degree network, but we report only the results from the latter model here because the EE model is known to have inflated type I error when testing nestedness on random matrices (Ulrich & Gotelli, 2007). Finally, we did not implement the commonly used 'FF' model that maintains fixed row and column totals, because some of the networks in our compilation are perfectly nested, while others are nearly perfectedly nested, and it is not possible to permute perfectly nested matrices using swap operations (Kallio, 2016).

Nestedness and modularity

To compare network dimensionality to other aspects of network structure, and compare network structure between eukaryote-infecting and bacteria-infecting viruses, we quantified nestedness and modularity for each network. Nestedness was quantified using stable NODF (sNODF), which is a modification of the common NODF metric that avoids penalizing rows or columns with the same degree, making the metric more stable to small perturbations of the data (Mariani et al. 2019). We also calculated nestedness using network temperature (Atmar and Patterson 1993), but we report sNODF results here, as the results were qualitatively highly similar for the two metrics. Modularity was quantified using Newman's modularity measure as implemented by the function computeModules in the R package 'bipartite' v2.17 (Dormann et al. 2008). Nestedness and modularity were compared to the same probabilistic degree null model used for dimensionality analysis.

Comparative network analysis

To investigate potential drivers of dimensionality across networks, we used multiple regression, where network dimensionality was the response variable and the predictors included host taxonomy (bacteria or eukaryote), whether the network included multiple viral taxa (yes/no), whether the strains in the network represent natural diversity or experimentally derived diversity, network connectance, total richness (total number of virus and host strains), minimum richness (defined as min(virus

richness, host richness)), and number of links in the network. We fit separate models using total richness, minimum richness, or number of links as predictors, to compare their explanatory power, because these metrics are correlated with one another.

RESULTS

We first discuss some qualitative insights gained from sorting infection networks to minimize dimensionality, before reporting quantitative comparisons of all networks. Throughout the results, we will focus on sorting networks from the host direction, but highly similar results are obtained when sorting from the virus direction (Table S1). Many networks can be sorted along a single dimension with zero or few errors, where an error means that the host range of a virus includes a host that it cannot actually infect (Figure 2). Some of these (nearly) one-dimensional networks exhibit a strong signal of host partitioning (similar to Figure 1c,h), such that viral host ranges are spread out along the axis of sorted hosts, with partial overlap in host range among neighbouring viruses (e.g., Figure 2a-e). Many networks display substantial evidence of both host partitioning and nestedness (similar to Figure 1e,g), such that the host ranges of viruses with narrower ranges are often included within the host ranges of viruses with broader ranges, while at the same time the nestedness is not perfect and there is clear segregation of host ranges along the sorted host axis (e.g., Figure 2f-o). Finally, three networks are perfectly nested along a single host axis (similar to Figures 1a and 2p-r).

Sorting networks to minimize dimensionality can help reveal multiscale network patterns, such as modules and partitioning or nestedness within modules (Figure 1g,h), which is exemplified by the network from Baudoux et al. (2015), which incorporates three clades of the picoeukaryote *Micromonas* and viruses that infect them. The sorted host axis separates clade C from clades A + B, as these two groups are infected by distinct sets of viruses (Figure 3). The axis also captures host partitioning between clades A and B, with these two clades separated along the axis, although some viruses can infect both clades (Figure 3). There is also evidence of partitioning within clade C and some nested host ranges in both modules (Figure 3). Remarkably, the phylogeny of host isolates based on 18S rRNA sequences maps very closely to the sorted host axis, with only one pair of isolates swapped in terms of order between the two arrangements (Figure 3). This network also illustrates the utility of a null model for interpreting dimensionality. The one-dimensional sort of this network contains seven errors, and two dimensions are required to perfectly sort the network with zero errors (Table S1). At the same time, a null model of this network contains 130 errors on average when sorted in one dimension, with a 95% interval of [108,152], indicating that the one-dimensional sort of the observed network is ca. 95% closer to perfect than expected by chance (i.e., 7/130 = 0.053). It should also be noted that networks can show a significant 'fit' to one dimension while still containing many errors. As described further below, nearly all networks in the dataset can be fit to one dimension better than expected by chance, but in ~half of the networks many errors are still present (File S3). In addition, many networks that are not one-dimensional can be perfectly sorted in two dimensions. Host ranges in two or more dimensions are more challenging to visualize, but could help reveal multiple mechanisms driving host range evolution. For example, the network reported by Flores et al. (2011), derived from a coevolution experiment with *Escherichia coli* and phage λ, is suggestive of an arms race causing host ranges to expand in two dimensions over time (Figure S1).

When comparing all 37 compiled networks, 11 can be perfectly sorted in one dimension, while 13 are two-dimensional, 9 are three-dimensional, 3 are four-dimensional, and 1 is five-dimensional (Table S1; Figure 4a). Nine of the two-dimensional networks can be sorted in one dimension with 5 or fewer errors. Therefore, roughly two-thirds of the networks are one- or two-dimensional, and roughly half are either perfectly one-dimensional or nearly so. Thirty-three of the networks can be fit to one dimension better than expected by chance, and the median fit of these networks to one dimension is 86% closer to perfect than expected by chance (Table S1). Two of the non-significant networks can actually be perfectly sorted in one dimension, indicating that a one-dimensional network is not significantly different than the null expectation for these networks, likely because these are smaller networks with lower statistical power (Table S1). A multiple regression with dimensionality as the response variable shows that dimensionality tends to increase with network richness (i.e., the total number of virus and host strains; $p = 0.006$, $F_{1,30} = 8.8$, partial $R^2 = 0.23$). However, there is a stronger relationship when the total richness predictor is replaced with the minimum richness, defined as min(virus richness, host richness) ($p < 10^{-4}$, $F_{1,30} = 22$, partial $R^2 = 0.36$; Figure 4b). Dimensionality is also correlated with number of links in the network ($p < 10^{-3}$, $F_{1,30} = 16$, partial $R^2 = 0.26$). Dimensionality is not affected by network connectance ($p = 0.19$, $F_{1,30} = 1.8$), the presence of multiple viral clades ($p = 0.54$, $F_{2,30} = 0.62$), or whether the strains represent natural or experimentally derived diversity ($p = 0.18$, $F_{1,30} = 1.9$). Dimensionality does not differ between networks with bacterial and eukaryotic hosts ($p = 0.86$, $F_{1,30} = 0.03$). There is also no evidence that dimensionality differs between bacterial and eukaryotic networks when comparing the distribution of dimensionalities between these groups (Fisher's exact

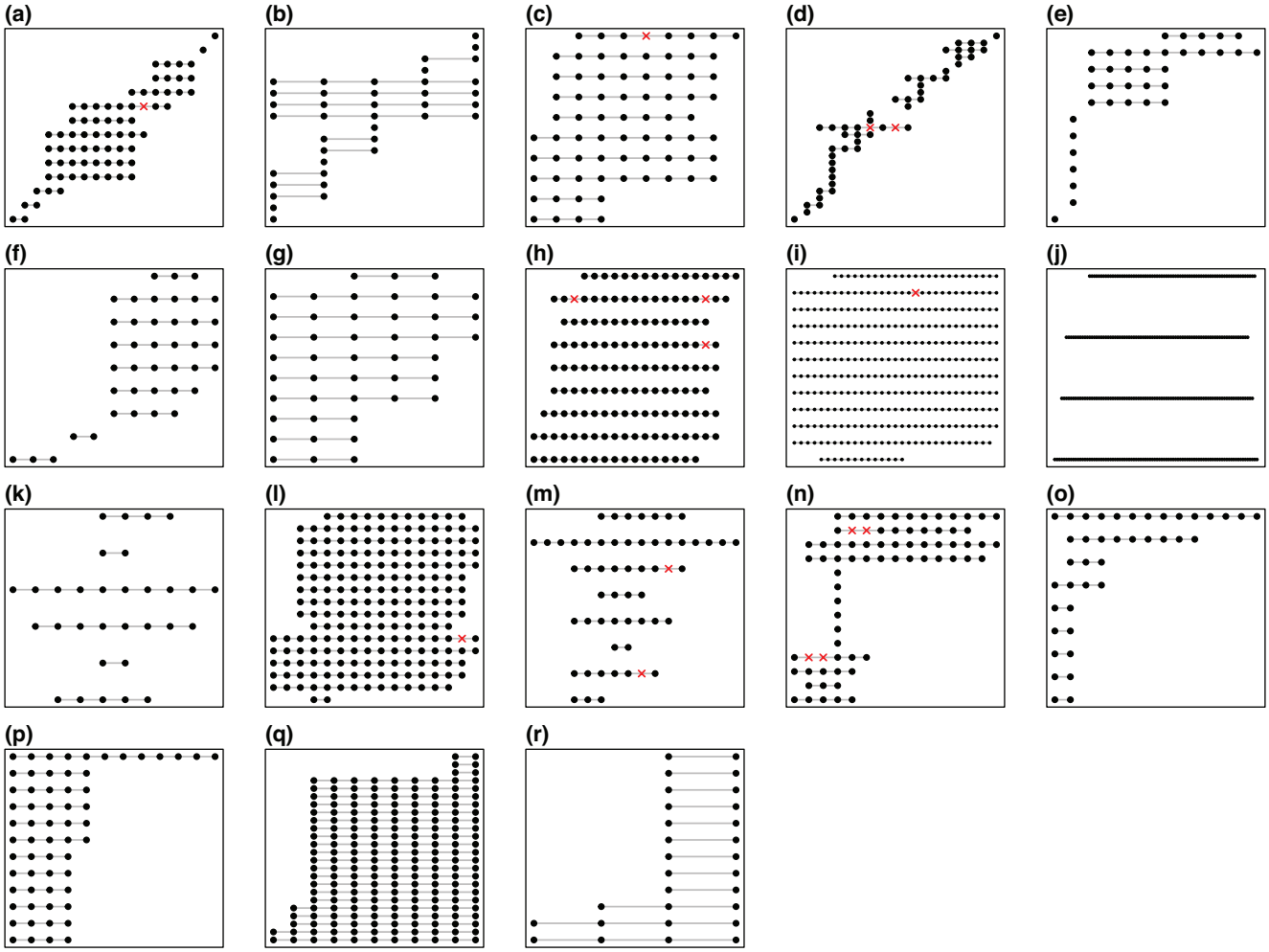

FIGURE 2 Empirical infection networks that can be sorted in one dimension with 0–4 errors. In each panel, hosts are arranged along the x-axis and viruses along the *y*-axis. Filled circles represent viruses that can successfully infect hosts at that position. Viral host ranges are represented by series of filled circles connected with lines, and red crosses represent errors (hosts within a viral host range that cannot be infected by that virus). Circles are not visible in panel (j) because of the large number of hosts in that study. Source publications for the networks: (a) Miklič & Rogelj, 2003; (b) DePaola et al., 1998; (c) Krylov et al. 2006; (d) Zinno et al., 2010; (e) Duplessis & Moineau, 2001; (f) Doi et al., 2003; (g) Nagasaki et al., 2005; (h) Nagasaki et al., 2004; (i) Kankila & Lindström, 1994; (j) Pantuček et al., 1998; (k) Seed & Dennis, 2005; (l) Synnott et al., 2009; (m) Capparelli et al. 2010; (n) Rybniker et al., 2006; (o) Tomaru et al., 2011; (p) Baudoux & Brussaard, 2005; (q) Middelboe et al., 2009; (r) Ceyssens et al., 2009.

test, $p=0.94$), or the average number of errors when sorted in one dimension (permutation test, $p=0.55$). Figure 4c,d compares dimensionality vs. minimum richness and vs. number of links.

Statistically significant nestedness is common in the compiled infection networks, consistent with the previous analysis of phage-bacteria networks by Flores et al. (2011). Of the 37 total networks, 26 are significantly nested, while 2 are significantly anti-nested (Table S1). These results were obtained using sNODF as the metric of nestedness, but the results are very similar when using network temperature (Table S1). 17 out of 26 bacterial networks are nested and 2 are anti-nested, while 9 out of 11 eukaryote networks are nested and 0 are anti-nested. The frequency of nested, non-nested, and anti-nested networks does not differ significantly between networks with bacterial and eukaryotic hosts (Fisher's exact test, $p=0.85$). Although most of the 37 networks are significantly nested, and most are relatively low-dimensional, there is no evidence for a relationship between nestedness and dimensionality. Networks that are significantly nested do not tend to have lower dimensionality than those that are not (permutation test, $p=0.40$), and do not tend to have fewer errors when sorted in one dimension (permutation test, $p=0.36$). Finally, in contrast to the commonness of nestedness, only a minority of the networks (8) are significantly modular, while 2 are significantly anti-modular (Table S1).

DISCUSSION

Our results show that viral infection networks are often low-dimensional, and the success in sorting these

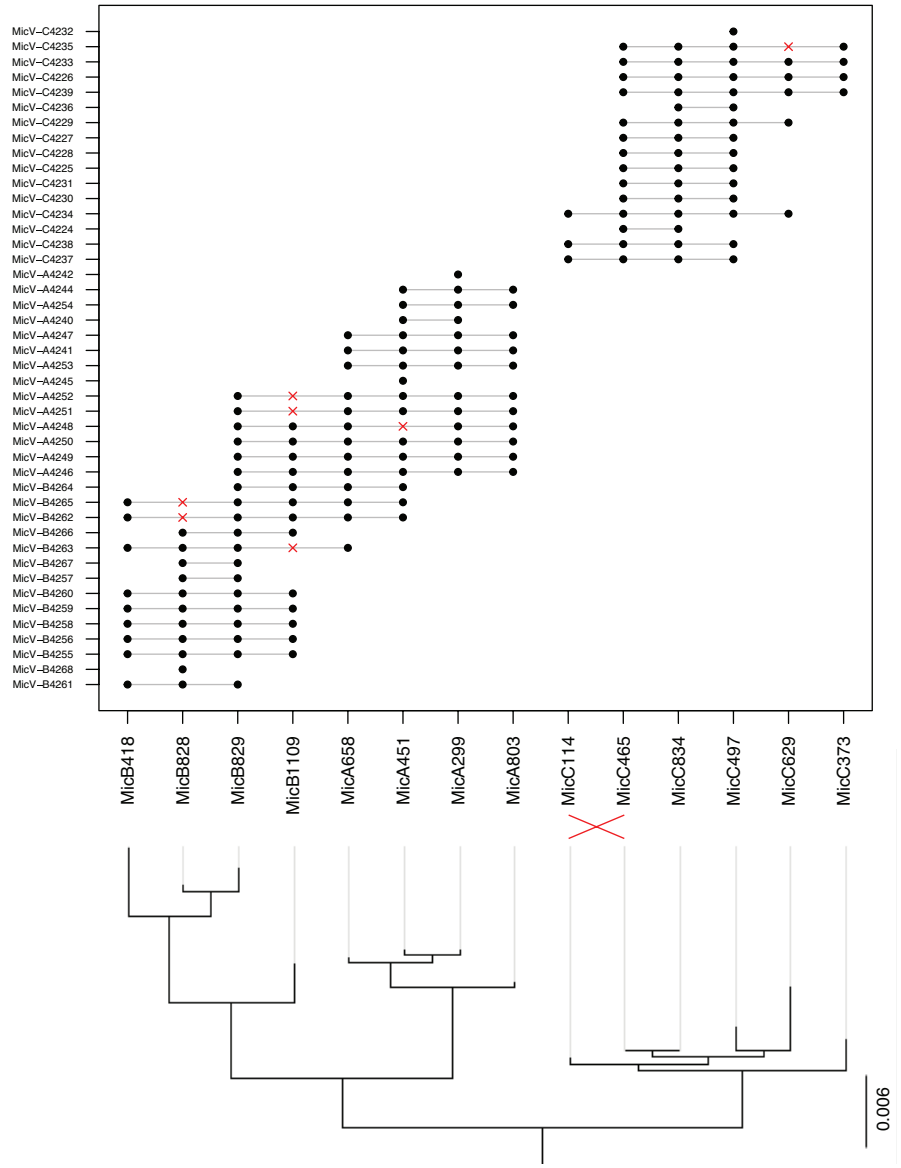

FIGURE 3 Infection network of 14 *Micromonas* and 44 virus strains, from Baudoux et al. (2015). The network is sorted from the host direction in one dimension with 7 errors. Host ranges and errors are represented in Figure 2. Below the host strain labels on the *x*-axis is a phylogeny of the host strains, estimated using FastTree v2.1.11 in Geneious v11.1.5, from a 1493 bp alignment (MAFFT v7.45) of 18S rRNA sequences. Support values for nodes are omitted for clarity. Red lines indicate two strains whose position on the phylogeny is flipped relative to their position in the 1D sorted network.

networks along a single dimension is generally much better than expected by chance. This simplicity in network structure is noteworthy, considering that the intricacy and complexity of host-virus interactions could conceivably lead to high-dimensional coevolutionary dynamics. The finding that dimensionality increases with network richness (Figure 4b) suggests that low dimensionality of observed networks may be driven in part by undersampling of the full diversity of host and virus populations. Nonetheless, even if true network dimensionality is underestimated, the relatively low dimensionality of the observed networks suggests it may be possible to explain most of network structure with a small number of underlying mechanisms that determine which viruses can infect which hosts. Likewise, host-virus coevolution may often play out primarily along a small number of dimensions. Testing the validity of these speculations will require further research that integrates analyses of network structure with detailed study of the mechanisms determining viral host range and host resistance.

The similar distributions of dimensionality and nestedness among bacterial and eukaryotic infection networks suggest that the processes underlying network structure are comparable across domains of life and distinct viral lineages that infect them. Because most (8/11) eukaryotic networks in the dataset include known or suspected dsDNA viruses, and all 26 bacterial networks

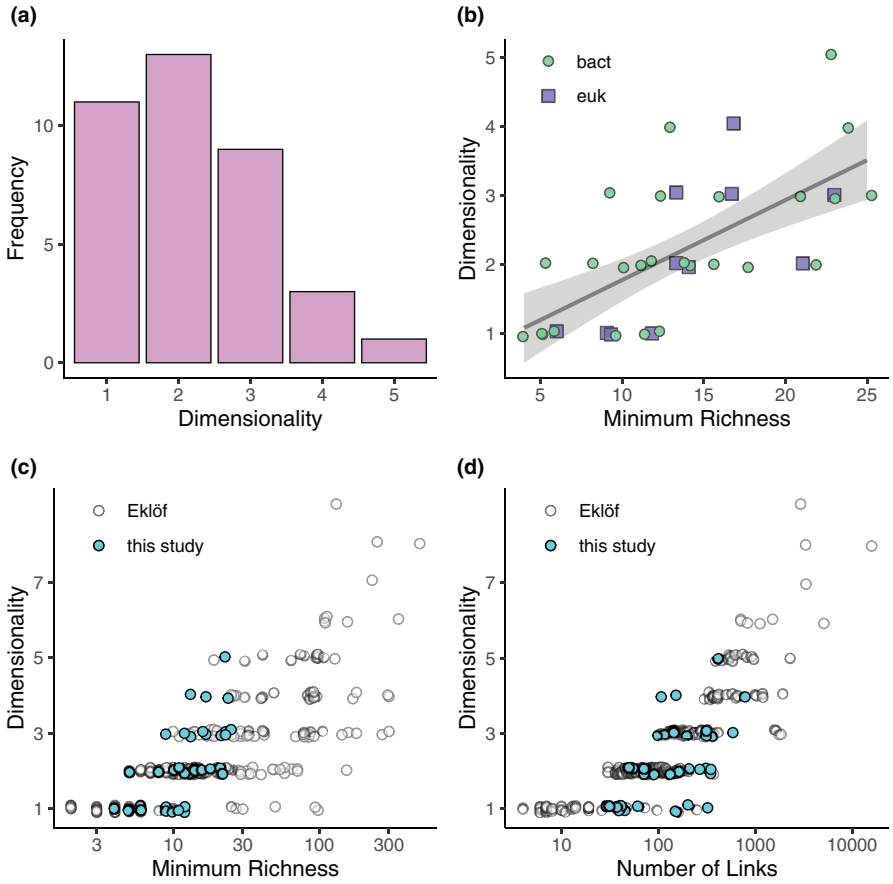

FIGURE 4 (a) Distribution of dimensionalities among the 37 compiled infection networks. (b) Relationship between dimensionality and minimum network richness (i.e., min(virus richness, host richness)). 'bact'=networks with bacterial hosts, 'euk'=networks with eukaryotic hosts. (c) Comparison of dimensionality vs. minimum richness in virus-host networks ('this study') and other ecological networks ('Eklöf'; data from Eklöf et al., 2013). (d) Comparison of dimensionality vs. number of links in virus-host networks ('this study') and other ecological networks ('Eklöf'; data from Eklöf et al., 2013). In (b-d) points are slightly jittered on the y-axis to better display overlapping points.

include known or suspected dsDNA viruses, it will be important to study network structure of viruses with other genome types, to assess whether other genome types are associated with alternative network patterns. It is noteworthy that the networks in the current study exhibit similar scaling of dimensionality vs. minimum richness and dimensionality vs. number of links, when compared to the diverse ecological networks compiled by Eklöf et al. (2013; Figure 4c,d). This suggests that networks composed of different kinds of ecological interactions are constrained to similar distributions of dimensionality, conditional on network diversity. It may be the case that common processes such as competitive resource partitioning and coevolution among interactors ultimately lead to similar dimensionalities for different network types.

One of the most useful outcomes of analysing network dimensionality is that the networks are sorted in a way that better reveals patterns of host partitioning (Figures 2, 3). The resulting axis (or axes) could be utilized in further analyses to test underlying mechanisms. For example, one can ask whether viral genotypes or traits are correlated with the location of their host ranges along the axis of sorted hosts. Likewise, host genotypes or traits can be compared to the location of host susceptibility ranges along an axis of sorted viruses, to test causes of susceptibility and resistance. We have focused on one-dimensional patterns here because they are easiest to visualize, but sorting networks in two or more dimensions would allow one to test whether different network dimensions are associated with different host and virus traits. One or more network axes could also be compared to phylogenies, to test whether there is a strong phylogenetic signal to network structure (as is evident in Figure 3), and axes could also be compared to environmental data describing the niches of isolates, to test whether host partitioning is associated with host niche differentiation.

Low dimensionality and nestedness are both frequent in the compiled networks, and nestedness is a network pattern that should be associated with low dimensionality. At the same time, many low-dimensional networks in this compilation do not show strongly nested structures, and there is no statistical relationship between dimensionality and nestedness. This suggests that a combination of host partitioning and nestedness, with their relative importance

varying among networks, maybe the most accurate paradigm for how infection networks are typically structured (Figure 1e,f), at least when focusing on networks that do not have great enough phylogenetic diversity to encompass true modules (Figure 1g,h). This conclusion has implications for understanding and modelling the dynamics of coevolution and the tradeoffs that constrain it. Host partitioning is most consistent with 'matching alleles' or 'imperfect lock and key' conceptual models of host–parasite coevolution (Beckett & Williams, 2013;Dennehy, 2012; Weitz et al., 2013), and its occurrence implies that viral diversity is structured by tradeoffs that cause viruses to specialize on different host phenotypes. It has also been noted that modularity in network structure may be driven by matching alleles coevolution (Weitz et al., 2013). Host partitioning and modularity are related patterns potentially driven by the same mechanism, because a continuous host axis with overlapping viral host ranges (Figure 1c) can ultimately form distinct modules, given sufficient host and virus divergence (Figure 1g,h) (Beckett & Williams, 2013). In contrast, nested network structure may be more consistent with 'gene for gene' coevolution, whereby a virus evolves a broader host range by possessing a set of genes needed to infect each host (Dennehy, 2012; Weitz et al., 2013). The apparent mix of host partitioning and nestedness in most low-dimensional networks (Figure 2) suggests that host-virus coevolution generally involves a mix of coevolutionary mechanisms (Agrawal & Lively, 2002). Combining analyses of dimensionality, nestedness, and modularity with genetic and phenotypic data will promote a better understanding of how this mixture of mechanisms operates in different microbial systems. Furthermore, the number of eukaryote-virus systems for which cross-infection data from diverse isolates has been collected is relatively modest, and additional research in this area will aid understanding of the distribution and drivers of network structure. Finally, understanding the causes of low or high dimensionality will benefit from in-depth studies of particular systems that attempt to disentangle the potential roles of strain richness, phylogenetic diversity, and spatiotemporal scale.

AUTHOR CONTRIBUTIONS

KFE designed the study, analysed data, and wrote the manuscript. CH contributed to data synthesis and analysis.

ACKNOWLEDGEMENTS

KFE was supported by a Simons Foundation Early Career Award in Marine Microbial Ecology and Evolution and NSF grants RII Track-2 FEC 1736030 and OCE 2129697.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ele.14383.

DATA AVAILABILITY STATEMENT

The data analysed in this study, and R code for analysing network dimensionality, are included as supplementary files and deposited on figshare (https://doi.org/10.6084/m9.figshare.24991041).

ORCID

Kyle F. Edwards https://orcid.org/0000-0002-0661-3903

REFERENCES

Agrawal, A. & Lively, C.M. (2002) Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research*, 4, 91–107.

Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, 96(3), 373–382. Available from: https://doi.org/10.1007/bf00317508

Baudoux, A.-C. & Brussaard, C.P. (2005) Characterization of different viruses infecting the marine harmful algal bloom species *Phaeocystis globosa*. *Virology*, 341, 80–90.

Baudoux, A.-C., Lebredonchel, H., Dehmer, H., Latimier, M., Edern, R., Rigaut-Jalabert, F. et al. (2015) Interplay between the genetic clades of *micromonas* and their viruses in the Western English Channel. *Environmental Microbiology Reports*, 7, 765–773.

Beckett, S.J. & Williams, H.T. (2013) Coevolutionary diversification creates nested-modular structure in phage–bacteria interaction networks. *Interface Focus*, 3, 20130033.

Brockhurst, M.A. & Koskella, B. (2013) Experimental coevolution of species interactions. *Trends in Ecology & Evolution*, 28, 367–375.

Burmeister, A.R., Sullivan, R.M., Gallie, J. & Lenski, R.E. (2021) Sustained coevolution of phage lambda and *Escherichia coli* involves inner-as well as outer-membrane defences and counter-defences. *Microbiology*, 167.

Capparelli, R., Nocerino, N., Iannaccone, M., Ercolini, D., Parlato, M., Chiara, M. et al. (2010) Bacteriophage therapy ofsalmonella enterica:a fresh appraisal of bacteriophage therapy. *The Journal of Infectious Diseases*, 201(1), 52–61. Available from: https://doi.org/10.1086/648478

Ceyssens, P.-J., Noben, J.-P., Ackermann, H.-W., Verhaegen, J., De Vos, D., Pirnay, J.-P. et al. (2009) Survey of *Pseudomonas aeruginosa* and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. *Environmental Microbiology*, 11, 1303–1313.

Chatterjee, S. & Rothenberg, E. (2012) Interaction of bacteriophage λ with its *E. Coli* receptor, LamB. *Viruses*, 4, 3162–3178.

Clerissi, C., Desdevises, Y. & Grimsley, N. (2012) Prasinoviruses of the marine green alga *Ostreococcus tauri* are mainly species specific. *Journal of Virology*, 86, 4611–4619.

Dennehy, J.J. (2012) What can phages tell us about host-pathogen coevolution? *International Journal of Evolutionary Biology*, 2012, 1–12.

DePaola, A., Motes, M.L., Chan, A.M. & Suttle, C.A. (1998) Phages infecting *Vibrio vulnificus* are abundant and diverse in oysters (*Crassostrea virginica*) collected from the Gulf of Mexico. *Applied and Environmental Microbiology*, 64, 346–351.

Doi, K., Zhang, Y., Nishizaki, Y., Umeda, A., Ohmomo, S. & Ogata, S. (2003) A comparative study and phage typing of silage-making *lactobacillus* bacteriophages. *Journal of Bioscience and Bioengineering*, 95, 518–525.

Dormann, C.F., Gruber, B. & Fruend, J. (2008) Introducing the bipartite package: analysing ecological networks. *R News*, 8(2), 8–11.

Duplessis, M. & Moineau, S. (2001) Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. *Molecular Microbiology*, 41, 325–336.

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P. et al. (2013) The dimensionality of ecological networks. *Ecology Letters*, 16, 577–583.

Flores, C.O., Meyer, J.R., Valverde, S., Farr, L. & Weitz, J.S. (2011) Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences*, 108, E288–E297.
Flores, C.O., Valverde, S. & Weitz, J.S. (2013) Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal*, 7, 520–532.
Gupta, A., Peng, S., Leung, C.Y., Borin, J.M., Medina, S.J., Weitz, J.S. et al. (2022) Leapfrog dynamics in phage-bacteria coevolution revealed by joint analysis of cross-infection phenotypes and whole genome sequencing. *Ecology Letters*, 25, 876–888.
Gurney, J., Aldakak, L., Betts, A., Gougat-Barbera, C., Poisot, T., Kaltz, O. et al. (2017) Network structure and local adaptation in co-evolving bacteria–phage interactions. *Molecular Ecology*, 26, 1764–1777.
Jover, L.F., Cortez, M.H. & Weitz, J.S. (2013) Mechanisms of multi-strain coexistence in host–phage systems with nested infection networks. *Journal of Theoretical Biology*, 332, 65–77.
Kallio, A. (2016) Properties of fixed-fixed models and alternatives in presence-absence data analysis. *PLoS One*, 11, e0165456.
Kankila, J. & Lindström, K. (1994) Host range, morphology and DNA restriction patterns of bacteriophage isolates infecting *rhizobium leguminosarum* bv. *Trifolii*. *Soil Biology and Biochemistry*, 26, 429–437.
Koonin, E.V., Krupovic, M. & Dolja, V.V. (2023) The global virome: how much diversity and how many independent origins? *Environmental Microbiology*, 25, 40–44.
Krylov, V.N., Miller, S., Rachel, R., Biebl, M., Pleteneva, E.A., Schuetz, M. et al. (2006) Ambivalent bacteriophages of different species active on escherichia coli K12 and salmonella sp. strains. *Russian Journal of Genetics*, 42(2), 106–114. Available from: https://doi.org/10.1134/s1022795406020025
Mariani, M.S., Ren, Z.-M., Bascompte, J. & Tessone, C.J. (2019) Nestedness in complex networks: observation, emergence, and implications. *Physics Reports*, 813, 1–90. Available from: https://doi.org/10.1016/j.physrep.2019.04.001
Middelboe, M., Holmfeldt, K., Riemann, L., Nybroe, O. & Haaber, J. (2009) Bacteriophages drive strain diversification in a marine *flavobacterium*: implications for phage resistance and physiological properties. *Environmental Microbiology*, 11, 1971–1982.
Miklič, A. & Rogelj, I. (2003) Characterization of lactococcal bacteriophages isolated from Slovenian dairies. *International Journal of Food Science & Technology*, 38, 305–311.
Nagasaki, K., Shirai, Y., Tomaru, Y., Nishida, K. & Pietrokovski, S. (2005) Algal viruses with distinct intraspecies host specificities include identical intein elements. *Applied and Environmental Microbiology*, 71, 3599–3607.
Nagasaki, K., Tomaru, Y., Katanozaka, N., Shirai, Y., Nishida, K., Itakura, S. et al. (2004) Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Applied and Environmental Microbiology*, 70, 704–711.
Pantuček, R., Rosypalová, A., Doškař, J., Kailerová, J., Ružičková, V., Borecká, P. et al. (1998) The polyvalent staphylococcal phage φ812: its host-range mutants and related phages. *Virology*, 246, 241–252.
Pilosof, S., Alcala-Corona, S.A., Wang, T., Kim, T., Maslov, S., Whitaker, R. et al. (2020) The network structure and eco-evolutionary dynamics of CRISPR-induced immune diversification. *Nature Ecology & Evolution*, 4, 1650–1660.
Poisot, T., Lepennetier, G., Martinez, E., Ramsayer, J. & Hochberg, M.E. (2011) Resource availability affects the structure of a natural bacteria–bacteriophage community. *Biology Letters*, 7, 201–204.
Poisot, T., Lounnas, M. & Hochberg, M.E. (2013) The structure of natural microbial enemy-victim networks. *Ecological Processes*, 2, 1–9.
Rybniker, J., Kramme, S. & Small, P.L. (2006) Host range of 14 mycobacteriophages in *mycobacterium ulcerans* and seven other mycobacteria including *mycobacterium tuberculosis*–application for identification and susceptibility testing. *Journal of Medical Microbiology*, 55, 37–42.
Seed, K.D. & Dennis, J.J. (2005) Isolation and characterization of bacteriophages of the *Burkholderia cepacia* complex. *FEMS Microbiology Letters*, 251, 273–280.
Suttle, C.A. (2005) Viruses in the sea. *Nature*, 437, 356–361.
Synnott, A.J., Kuang, Y., Kurimoto, M., Yamamichi, K., Iwano, H. & Tanji, Y. (2009) Isolation from sewage influent and characterization of novel *Staphylococcus aureus* bacteriophages with wide host ranges and potent lytic capabilities. *Applied and Environmental Microbiology*, 75, 4483–4490.
Tomaru, Y., Shirai, Y., Suzuki, H., Nagasaki, T. & Nagumo, T. (2008) Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom chaetoceros debilis. *Aquatic Microbial Ecology*, 50, 103–112. Available from: https://doi.org/10.3354/ame01170
Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., Koike, K. & Nagasaki, K. (2011) Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* Grunow. *Applied and Environmental Microbiology*, 77, 5285–5293.
Ulrich, W. & Gotelli, N.J. (2007) Null model analysis of species nestedness patterns. *Ecology*, 88, 1824–1831.
Ward, B.A., Dutkiewicz, S., Jahn, O. & Follows, M.J. (2012) A size-structured food-web model for the global ocean. *Limnology and Oceanography*, 57, 1877–1891.
Weitz, J.S., Hartman, H. & Levin, S.A. (2005) Coevolutionary arms races between bacteria and bacteriophage. *Proceedings of the National Academy of Sciences*, 102, 9535–9540.
Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B. et al. (2013) Phage–bacteria infection networks. *Trends in Microbiology*, 21, 82–91.
Zinno, P., Janzen, T., Bennedsen, M., Ercolini, D. & Mauriello, G. (2010) Characterization of *Streptococcus thermophilus* lytic bacteriophages from mozzarella cheese plants. *International Journal of Food Microbiology*, 138, 137–144.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Edwards, K.F. & Hayward, C. (2024) The dimensionality of infection networks among viruses infecting microbial eukaryotes and bacteria. *Ecology Letters*, 27, e14383. Available from: https://doi.org/10.1111/ele.14383