Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks

Stephen J. Beckett* and Hywel T.P. Williams

Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

S.J.Beckett@exeter.ac.uk , H.T.P.Williams@exeter.ac.uk

Phage and their bacterial hosts are the most diverse and abundant biological entities in the oceans, where their interactions have a major impact on marine ecology and ecosystem function. The structure of interaction networks for natural phage-bacteria communities offers insight into their coevolutionary origin. At small phylogenetic scales, observed communities typically show a nested structure, in which both hosts and phage can be ranked by their range of resistance and infectivity respectively. A qualitatively different multiscale structure is seen at larger phylogenetic scales; a natural assemblage sampled from the Atlantic Ocean displays large-scale modularity and local nestedness within each module. Here we show that such "nested-modular" interaction networks can be produced by a simple model of host-phage coevolution in which infection depends on genetic matching. Negative frequency-dependent selection causes diversification of hosts (to escape phage) and phage (to track their evolving hosts). This creates a diverse community of bacteria and phage, maintained by kill-the-winner ecological dynamics. When the resulting communities are visualised as bipartite networks of who-infects-whom, they show the nested-modular structure characteristic of the Atlantic sample. The statistical significance and strength of this observation varies depending on whether the interaction networks take into account the density of the interacting strains, with implications for interpretation of interaction networks constructed by different methods. Our results suggest that the apparently complex community structures associated with marine bacteria and phage may arise from relatively simple coevolutionary origins.

*Corresponding author.

1 Introduction

Bacteriophage and their bacterial hosts are the most abundant and diverse replicating entities in the oceans, playing central roles in marine ecology and ecosystem processes [1, 2, 3, 4, 5, 6, 7]. Fast replication and high mutation rates mean that bacteria and phage can evolve – and coevolve – rapidly [8, 9, 10, 11], suggesting that coevolution will influence both ecological dynamics and ecosystem processes. Yet the basic mode of bacteria-phage coevolution is unclear. Experimental studies have demonstrated adaptation of resistance and infectivity ranges over just a few generations of laboratory coevolution [8, 9, 12, 13], often interpreted as a coevolutionary 'arms-race' in which hosts evolve to expand their range of resistance, while phages evolve to expand their host range. Unconstrained arms races are predicted to result in low diversity [14], with a single dominant host/phage strain, or perhaps two dominant host types if there is a trade-off between resistance and resource competition [8]. However, the short time intervals involved mean that the experimental coevolution data can be ambiguous and may sometimes also be consistent with a 'fluctuating selection' mode of coevolution in which infection is highly specific, so that hosts are more resistant to contemporary phage than ancestral or future strains [15]. Fluctuating selection dynamics are consistent with aquatic viral ecology models predicting kill-the winner dynamics [16, 17], whereby the most successful hosts (in terms of resource competition) are prevented from becoming dominant by increased viral predation. In kill-the-winner ecological dynamics, density-dependent predation by specialised viruses imposes negative frequency-dependent selection pressure on hosts, favouring rare phenotypes. Such dynamics are believed to support the maintenance of diverse communities of marine bacteria and phage [18, 19]. Recent genomic studies give empirical support for high natural diversity of marine bacteria and phage, with high specificity of infection and rapid coevolution [18, 20, 11, 21]. Thus despite much progress in experimental coevolution and marine microbial genomics, substantial uncertainties remain about the basic mode of phage-bacteria coevolution.

A complementary source of data about coevolution lies in the structure of natural phage-bacteria communities. Some recent data compilations have represented phage-bacteria communities as bipartite networks representing which phage strains were observed to infect which bacteria strains [22, 23, 24]. Statistical analyses of these 'phage-bacteria infection networks' (most often given in the form of binary matrices of presence(1)-absence(0) of pairwise infection) has so far focused on the matrix metrics of "nestedness" and "modularity" (see inset to Figure 1). Nestedness is a measure of the extent to which the non-zero elements of each row (or column) in the

matrix are a subset of the non-zero elements in the subsequent rows (or columns). In a perfectly nested matrix, the entries in each row (column) are a strict subset of the entries in the next row (column); thus each row (column) is nested inside the next row (column). In terms of phage-bacteria interaction, nestedness relates to the differentiation of strains along a gradient from specialist (small range) to generalist (large range). Here "range" refers, for bacteria, to the number of phage strains against which it is resistant, and for phage, to the number of bacterial strains it can infect. A perfectly nested pattern is one where the hosts and phages are each ranked along the specialist-generalist gradient such that the specialist strategies are subsets of the more generalised strategies. A modular network structure occurs when nodes can be partitioned into subsets such that most connections occur within rather than between the different subsets. For bacteria and phages modularity can be interpreted as a specialised interaction structure, without transitivity (i.e. where strains cannot be ranked by increasing range), in which distinct clusters of phage strains preferentially infect distinct clusters of bacterial strains.

Assuming that natural interactions between phage and bacteria are ultimately the product of coevolution, phage-bacteria interaction network structures may offer insight into the coevolutionary processes that produced them. At small phylogenetic scales these networks typically show higher-than-expected nestedness and lower-than-expected modularity [22]. High nestedness is consistent with an arms-race mode of coevolution, where hosts and phages evolve to increase their range of resistance/infectivity. However, the largest reported cross-infection assay involved 774 bacterial strains and 298 phage strains isolated from multiple geographically dispersed sites across the Atlantic Ocean [25]. Although no explicit genotyping was conducted in this study, it is likely that this dataset spans a broad phylogenetic scale. Reanalysis of interactions between 286 host strains and 215 phage strains from this dataset [23] found that the resulting network showed large-scale modularity, with local nestedness within each module. This is visualised in Figure 1, in which bacteria (rows) and phage (columns) were ordered (following [23]) to maximise statistical modularity and within-module nestedness. Here we highlight the identified modules by colour; interactions falling outside any identified module are shown in white. We also add two inset schematics illustrating perfectly modular and perfectly nested matrix structures. It is difficult to explain this "nested-modular" pattern as the result solely of arms race coevolution; the lack of global nestedness and the presence of distinct modules suggests that some additional mechanism is needed. While models of coevolution based on high specificity of infection often predict diversification [14] – and might thus be invoked to explain formation of distinct modules – such models do not explain the presence of within-module nestedness.

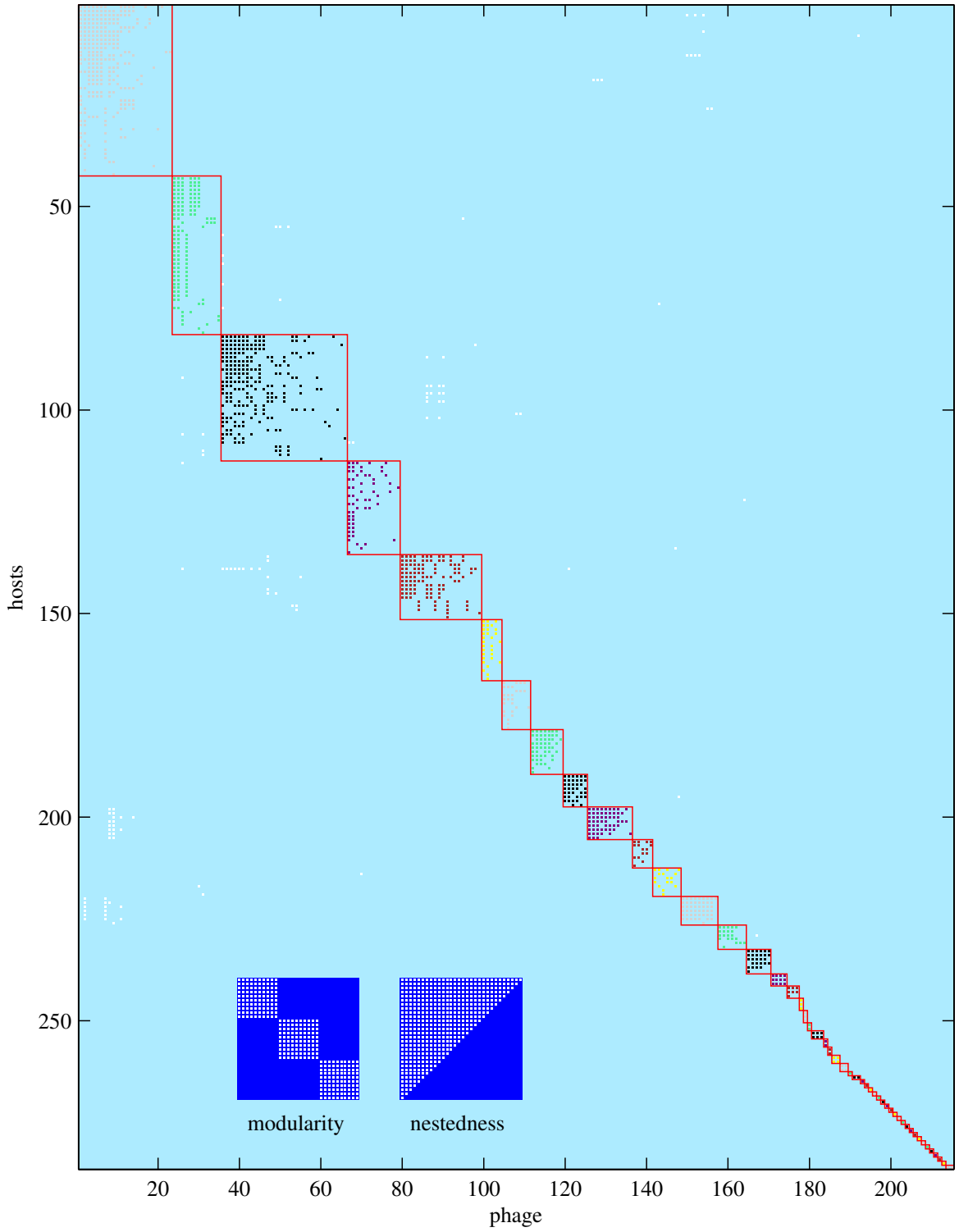

Figure 1: Nested-modular interaction structure of 215 phage strains and 286 bacteria strains sampled from the Atlantic Ocean (adapted from [23]). The plot shows which phage strains can infect which host strains from the dataset presented by Moebus and Nattkemper [25]. Flores *et al.* [23] re-sorted the interaction matrix to maximise modularity and within-module nestedness. Here we additionally highlight identified modules by shading; interactions falling outside any module are shown in white. We also add two inset schematics illustrating perfectly modular and perfectly nested matrix structures.

Here we explore a simple model of coevolution based on genetic matching [26, 27, 28], which we dub the 'relaxed lock-and-key model'. The model is mechanistically justified by reference to coadaptation of (e.g.) phage tail-fibres and host surface receptors [29]. We show that relaxed lock-and-key coevolution is sufficient to produce the core structural features of observed phage-bacteria communities: stable high diversity of bacteria and phage, modularity at large phylogenetic scales, and nestedness at small phylogenetic scales. Furthermore, we show that the strength and statistical significance of the observed nested-modular pattern depends on how the interaction networks are formed. Here we contrast interaction networks based on the potential adsorption rate of each phage strain on each host strain with interaction networks based on actual infection rate measured in an ecological context. Our findings highlight difficulties with comparison of interaction networks constructed by these different methods.

In the next section, we present the relaxed lock-and-key coevolution model in the ecological context of a multi-strain chemostat. This is followed by presentation of results showing the co-diversification of bacteria and phage, the construction of associated adsorption rate and infection rate interaction networks, and analyses of network properties over time. Finally we discuss the relevance of the relaxed lock-and-key model for understanding natural phage-bacteria communities.

2 Model

We model bacteria-phage coevolution by adding mutation to numerical simulations of a multi-strain chemostat. Below we describe the ecological model, the coevolutionary model (including how infection rate is calculated for a given pair of bacteria and phage), and methods used to analyse phage-bacteria interaction networks. A description of the parameters used is given in Table 1. The model has been analysed previously [27, 28]; model sensitivity to key parameters is given in [28]. It is derived from a similar model [26] (with the principal difference being in how the evolutionary dynamics are evaluated), which in turn derives from an earlier single-strain ecological model of bacteria-phage growth in a chemostat [30].

Symbol	Description	Value	Unit
R	Resource concentration	Variable	$\mu g\, ml^{-1}$
N_i	Density of host strain i	Variable	$cells\, ml^{-1}$
V_j	Density of phage strain j	Variable	$virions\, ml^{-1}$
N_{init}	Initial host density	4.6×10^4	$cells\, ml^{-1}$
V_{init}	Initial phage density	8.1×10^5	$virions\, ml^{-1}$
ω	Chemostat dilution rate	0.0033	min^{-1}
R_0	Resource supply concentration	2.2	$\mu g\, ml^{-1}$
ε	Resource conversion rate	2.6×10^{-6}	$\mu g\, cell^{-1}$
γ	Maximum resource uptake rate	0.0123	$\mu g\, min^{-1}$
K	Half-saturation constant	4	$\mu g\, ml^{-1}$
δ_i	Growth scaling for host h_i	Range $[\delta_{min}, \delta_{max}]$	$scalar$
δ_{min}	Min. growth scaling factor	0.8	$scalar$
δ_{max}	Max. growth scaling factor	1.2	$scalar$
ϕ	Maximum adsorption rate	0.104×10^{-8}	$ml(min\, virion)^{-1}$
θ_{ij}	Ads. scaling for v_j on h_i	Range $[0, \phi]$	$scalar$
β	Burst size	71	$virions\, cell^{-1}$
h_i	Genotype of bacteria i	Range $[0, 1]$	$scalar$
v_j	Genotype of phage j	Range $[0, 1]$	$scalar$
$\hat{h}_i$	Resistance phenotype of bacteria i	Range $[0, 1]$	$scalar$
$\hat{v}_j$	Infection phenotype of phage j	Range $[0, 1]$	$scalar$
h_{init}	Initial bacteria genotype	0.2	$scalar$
v_{init}	Initial phage genotype	0.2	$scalar$
S	Specificity of phage	100	$scalar$
μ_N	Host mutation rate	10^{-6}	$cell^{-1}$
μ_V	Phage mutation rate	10^{-5}	$virion^{-1}$
σ_N	Std. dev. of host mut. range	0.01	$scalar$
σ_V	Std. dev. of phage mut. range	0.01	$scalar$
M_N	Bacterial mutation size	Random variable	$scalar$
M_V	Phage mutation size	Random variable	$scalar$
Δt	Integration timestep	10	min
T	Simulation duration	5×10^7	min
ρ	Resolution of genotype diversity	0.001	$scalar$
L	Chemostat volume	1	ml

Table 1: **Model parameters and variable definitions**. Variables can change during a simulation. Numerical values are parameters fixed for the duration of a simulation. Range values are deterministically calculated from other variables/parameters.

2.1 Multistrain chemostat model

The ecological model represents the interactions between multiple strains of bacteria and phage in a single-resource chemostat. Resource concentration R and densities of the i^{th} strain of bacteria N_i and the j^{th} strain of phage V_j are governed by the system of equations below:

$$\frac{dR}{dt} = -\omega(R - R_0) - \sum_i \varepsilon \frac{\gamma \delta_i R N_i}{R + K}$$

$$\frac{dN_i}{dt} = -\omega N_i + \frac{\gamma \delta_i R N_i}{R + K} - \sum_j \phi \theta_{ij} N_i V_j$$

$$\frac{dV_j}{dt} = -\omega V_j + \sum_i \beta \phi \theta_{ij} N_i V_j \quad (1)$$

Resource concentration is affected by the chemostat washout rate ω, the supply concentration R_0, and uptake by all bacterial strains. Bacterial resource uptake is governed by Monod kinetics [31] with half-saturation rate K and maximum uptake rate γ, adjusted for each bacterial strain i by a genetically encoded scaling coefficient δ_i. Bacterial strain density is a function of washout, population growth and lysis. Resource uptake is converted directly into bacterial population growth via resource conversion constant ε. Each bacterial strain is potentially susceptible to infection by every strain of phage, depending on genetic match. Phage strain density is determined by washout and the sum of production on all available hosts. Adsorption of phage j to host i is the product of the maximum adsorption rate ϕ and a scaling coefficient θ_{ij}. Every adsorption event leads to infection and instantaneous cell lysis (we assume no latent period) creating new phage with burst size β.

2.2 Relaxed lock-and-key coevolution model

We model evolution in the multi-strain ecological model by adding a mutation process that introduces new variants of existing bacteria and phage strains. Uncompetitive strains are eventually removed by chemostat dilution. Thus we have a simple model in which bacteria and phage phenotypes can evolve by natural selection. We do not separate the evolutionary and ecological timescales, but instead assume a fixed probability of mutation per new cell/phage and allow evolutionary dynamics to play out in our numerical simulations. We ran our simulations for 5×10^7 minutes, sufficient for the dynamics to reach a quasi-stable equilibrium. This timescale is chosen to allow a relatively slow evolutionary dynamic in the context of faster ecological dynamics, and is not intended to accurately reflect analogous timescales in natural systems.

Bacterial h and phage v genotypes are modelled as single values in the range $[0, 1]$ (binned at resolution $\rho = 0.001$). Bacteria and phage have mutation rates μ_N and μ_V respectively, applied stochastically for each new cell/phage. Mutation cre-

ates a single cell/phage with a genotype created by adding a normal deviate to the parental genotype, with standard deviation (mutation range) of σ_N or σ_V for bacteria and phage respectively. If the density of any population falls below $1\,cell\,ml^{-1}$ or $1\,virion\,ml^{-1}$ (possible due to the continuous nature of the mathematical abstraction), that population is assumed to be lost and is removed from the system. Any simulations where genotypes reached the edges of the permitted range $[0, 1]$ were discarded; however parameters were chosen to ensure this did not occur.

The relaxed lock-and-key coevolution model is created by enforcing a dependence of adsorption rate on the genetic similarity of host and phage. The relative adsorption rate θ_{ij} of phage j on host i is given by:

$$\theta_{ij} = e^{-S\left(\hat{h}_i - \hat{v}_j\right)^2} \tag{2}$$

where $\hat{h}_i$ is the bacterial resistance strategy for the i^{th} host encoded by genotype h_i, $\hat{v}_j$ is the phage infectivity strategy for the j^{th} phage encoded by genotype v_j and S represents infection specificity. This function has Gaussian form, with specificity governing the width of the infection curve, i.e. high specificity indicates narrow host range, whilst low specificity indicates wide host range. Hence the closer the numerical values of strategies $\hat{h}_i$ and $\hat{v}_j$ are, the greater the rate of adsorption between host i and phage j.

Bacterial genotype also specifies a growth-rate scaling trait δ ; where each bacterial genotype h_i maps to a growth rate δ_i. Here we impose a simple growth rate fitness landscape with a singular peak of $\delta_{max} = 1.2$ at $h = 0.5$, falling linearly to a minimum of $\delta_{min} = 0.8$ at the edges of the range (i.e. for $h = 0$ and $h = 1$). The growth rate trait δ is a coefficient that affects population growth rate by scaling the rate of resource uptake by bacteria (see Equation 1). It is important to note that there are no inherent differences in bacterial resistance (that is, all bacteria have the same size of range of resistance); thus there are no costs of resistance and no explicit trade-offs between growth rate and infection rate. Ecological trade-offs between bacterial growth and susceptibility to infection can emerge [28], but these depend on the density and composition of the contemporary phage and bacterial communities.

2.3 Infection network analysis

Interactions between bacteria and phage strains (i.e. who-infects-whom) can be represented as a network [24]. Such networks are bipartite graphs with interactions between two types of node (bacteria and phage). We use two forms of interaction network to visualise our model phage-bacteria communities. The first type is formed from the adsorption rate $\phi\theta_{ij}$ of phage j on host i , which gives a pairwise interaction

matrix for all strains present in the community irrespective of their abundance. The second type of interaction network is formed from the actual infection rates for phage j on host i in the current ecological context, calculated as $\phi\theta_{ij}N_iV_j$ from the adsorption rate and the current densities of each strain, giving a more ecologically relevant measure of interaction.

Since both adsorption rate and infection rate are quantitative metrics, we create binary interaction matrices by applying threshold filters. The resulting binary matrices consist of 1s where the pairwise interaction is strong and 0s elsewhere. The binary matrices can be analysed using standard metrics for nestedness and modularity. They can also be compared to reported interaction networks, which are typically given in binary form. We used the package BiWeb [32], which utilises the LP-BRIM algorithm to find the partition that best maximises Barber's modularity (Q_b) for bipartite networks [33, 34]. Since the LP-BRIM sorting algorithm is sensitive to initialisation, we repeated the modularity assessment 5 times for each measurement, taking the maximum score returned. We measured nestedness using the deterministic NODF (nestedness metric based on overlap and decreasing fill) algorithm [35], which returns a score in the range $[0, 100]$ (where 100 indicates a perfectly nested structure). NODF normalises for matrix size, allowing matrices of differing sizes to be compared.

To quantify the statistical significance of the nestedness (NODF) and modularity (Barber's Qb) scores measured for a particular binary matrix, we calculate statistical significance p as the likelihood of achieving a score greater than or equal to the score for the input matrix in a sample of M matrices from a null distribution of random matrices. Where this method returns $p = 0$ we conservatively assign $p < \frac{1}{M}$. We use a null model in which matrix size (number of rows and columns) and fill (number of 1s) are conserved, but where adjacency (position of 1s) is randomly reassigned. Thus we maintain the numbers of host and phage strains, and the total number of host-phage interactions, but reassign the pattern of who-infects-whom. We estimate the null distribution using a sample of M matrices generated using a stochastic algorithm with pseudo-random numbers. For significance of NODF scores, we create the sample in two sets, adding matrices to both sets until the mean NODF scores for both sets converge within a tolerance bound of 0.01. The sample used to estimate the null distribution is then formed as the union of both sets. Thus for NODF, the sample size M varies according to how many null matrices are needed to achieve convergence of the means, with a minimum sample of 500 null matrices computed in each case. This method gives reliable estimation of the underlying distribution while retaining computational efficiency. For the modularity scores, we used $M = 100$, assigning each null matrix the highest Barber's modularity score from 5 random initialisations of the LP-BRIM algorithm. A free software package including these methods is currently

in preparation [36].

3 Results

In previous work [28] we have shown that in the absence of phage, resource limitation leads to competitive exclusion of slow-growing bacteria by fast-growing bacteria. Faster-growing populations draw down resource concentration to a limiting level at which slower-growing populations cannot be sustained against losses from washout and are lost from the community. Thus in the absence of phage, bacteria evolve to the fastest-growing genotype permitted by the simple unimodal growth rate fitness landscape (here located at $h = 0.5$ - see Section 2.2). In the presence of phage, host evolution is affected by the additional coevolutionary selection pressures imposed by phage predation.

The relaxed lock-and-key coevolution model robustly produces diversification of bacteria and phage. Figure 2 shows a simulation run initialised with a single bacterial strain and perfect-match phage strain (with $h = v = 0.2$). In the first stage (until $t \approx 0.5 \times 10^7 min$) bacteria evolve to increase growth rates, while phage evolve to track their hosts through genotypic space. Once the bacteria reach the maximum growth rate genotype at $h = 0.5$, they can no longer improve fitness by increasing growth rate (and would remain at this fitness peak indefinitely in the absence of phage [28]). However, phage create a strong selective pressure for host diversification due to density-dependent predation, which favours host mutants with a lower genetic match to dominant phage strains. This causes an evolutionary branching event to occur between $t \approx 0.6 \times 10^7 min$ and $t \approx 1.1 \times 10^7 min$. Further evolutionary branching events occur until $t \approx 2.5 \times 10^7 min$. After this period of coevolutionary diversification the distribution of strains settles down to a quasi-stable state for the remainder of the simulation. At this stage there are 5 clearly identifiable clusters of similar genotypes for both bacteria and phage, where each cluster (hereafter "species") represents an ecologically similar (but genetically diverse) sub-population. Each phage species is attracted towards its two flanking host species, while each host species is repelled by its flanking phage species; this process of attraction and repulsion sometimes results in transient oscillatory dynamics (this effect is clearly seen for $t \approx 2 - 3 \times 10^7 mins$).

Figure 3 visualises community structure for a timeslice from the simulation taken at $t = 3.5 \times 10^7 min$. There are 64 bacterial strains and 97 phage strains present in the community. Densities of the different strains vary widely and are unevenly distributed, but 5 clearly identifiable species of similar genotypes in both the bacteria (lower left) and phage (lower right) populations are visible. For ease of reference we label the host species H1-H5 and the phage species P1-P5.

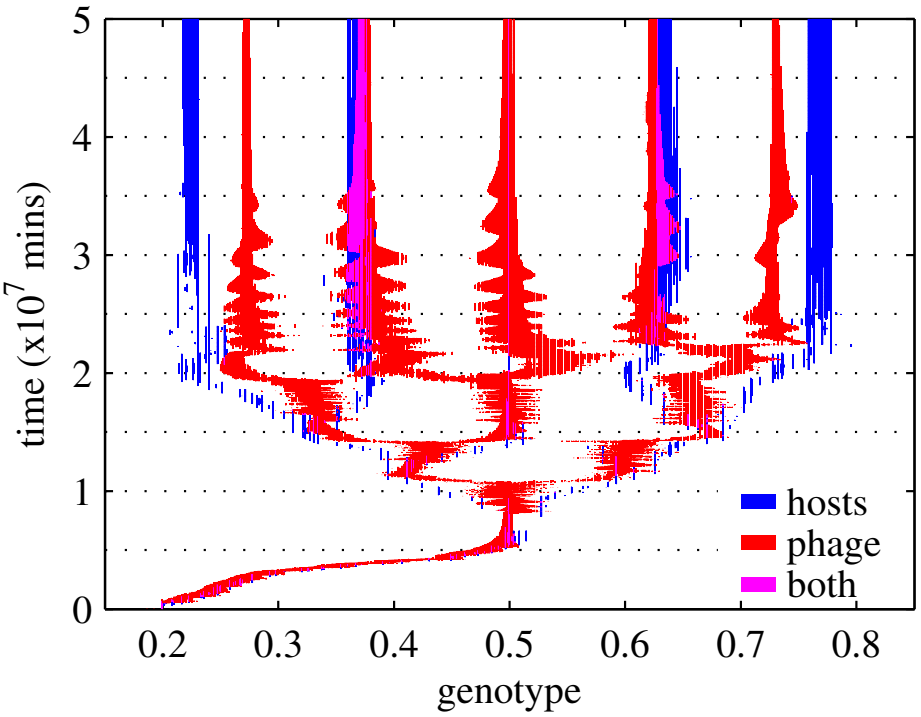

Figure 2: Coevolutionary diversification of bacteria and phage strains with relaxed lock-and-key model. Plot shows phage-bacteria community composition over time for a simulation run initialised with a single host strain and perfectly matched phage (initial genotypes $h = v = 0.2$). Genotypes with non-zero density for bacteria, phage, or both bacteria and phage are highlighted. The fastest-growing bacteria genotype is located at $h = 0.5$. The cross-section of the community at time $t = 3.5 \times 10^7$ is examined in figure 3. Horizontal dotted lines indicate time points for which interaction matrices are presented in figures 4 and 5.

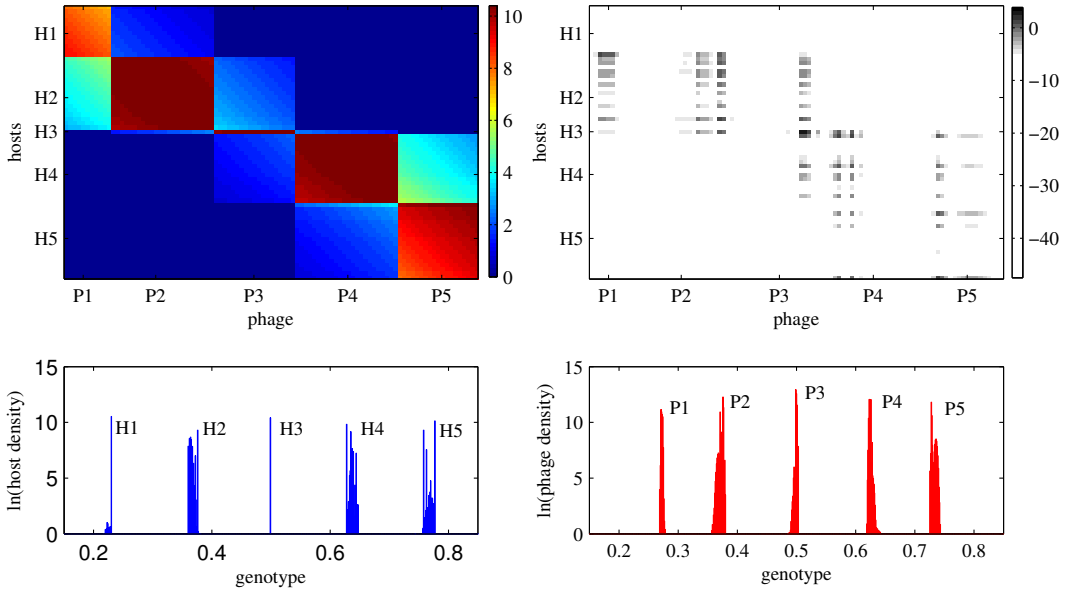

Figure 3: Snapshot of phage-bacteria community structure and interaction during the case study simulation (measured at $t = 3.5 \times 10^7 min$). *Upper left:* Adsorption rate matrix plot showing pairwise adsorption rate ($\phi\theta_{ij}$, values scaled$\times 10^{-10}$) for each host (i) and phage (j) strain present in the current community. *Upper right:* Infection rate matrix plot showing actual infection rate in the context of the community (ln $\mid \phi\theta_{ij}N_iV_j \mid$) for each host-phage pair. *Lower left:* Current density of all bacterial genotypes (ln $\mid N_i \mid$). *Lower right:* Current density of all phage genotypes (ln $\mid V_j \mid$). For ease of reference, labels H1-H5 and P1-P5 are manually attached to identify bacteria and phage "species".

We visualise interactions between phage and bacterial strains in Figure 3 by plotting the (density-independent) adsorption rate matrix (upper left) and the (density-weighted) infection rate matrix (upper right). These matrices represent a snapshot of the coevolving interactions between phage and bacteria - note that the matrices include all strains that are present, but do not cover the whole genetic space (i.e. absent strains are not plotted). Each host species interacts strongly with a single phage species, as indicated by the modular structure apparent in the adsorption rate matrix (e.g. P1 is specialised on H1, P2 on H2, and so on). However, there are also weaker interactions with adjacent phage species (e.g. H2 is also weakly affected by P1 and P3). The infection rate matrix gives a different view of the community; whereas the adsorption rate shows a potential interaction, the infection rate shows the interaction in terms of actual mortality of the host caused by the phage in ecological interaction. The infection rate matrix shows that most of the potential interactions shown in the adsorption rate matrix are ecologically insignificant, since only a few strains in each species are present at sufficient density for a strong interaction to occur. Also there is a many-many interaction structure, with each bacterial strain infected by multiple phage strains, and each phage infecting multiple bacteria. Interestingly, the infection rate matrix also suggests a degree of modularity, though module membership appears different to that seen in the adsorption rate matrix.

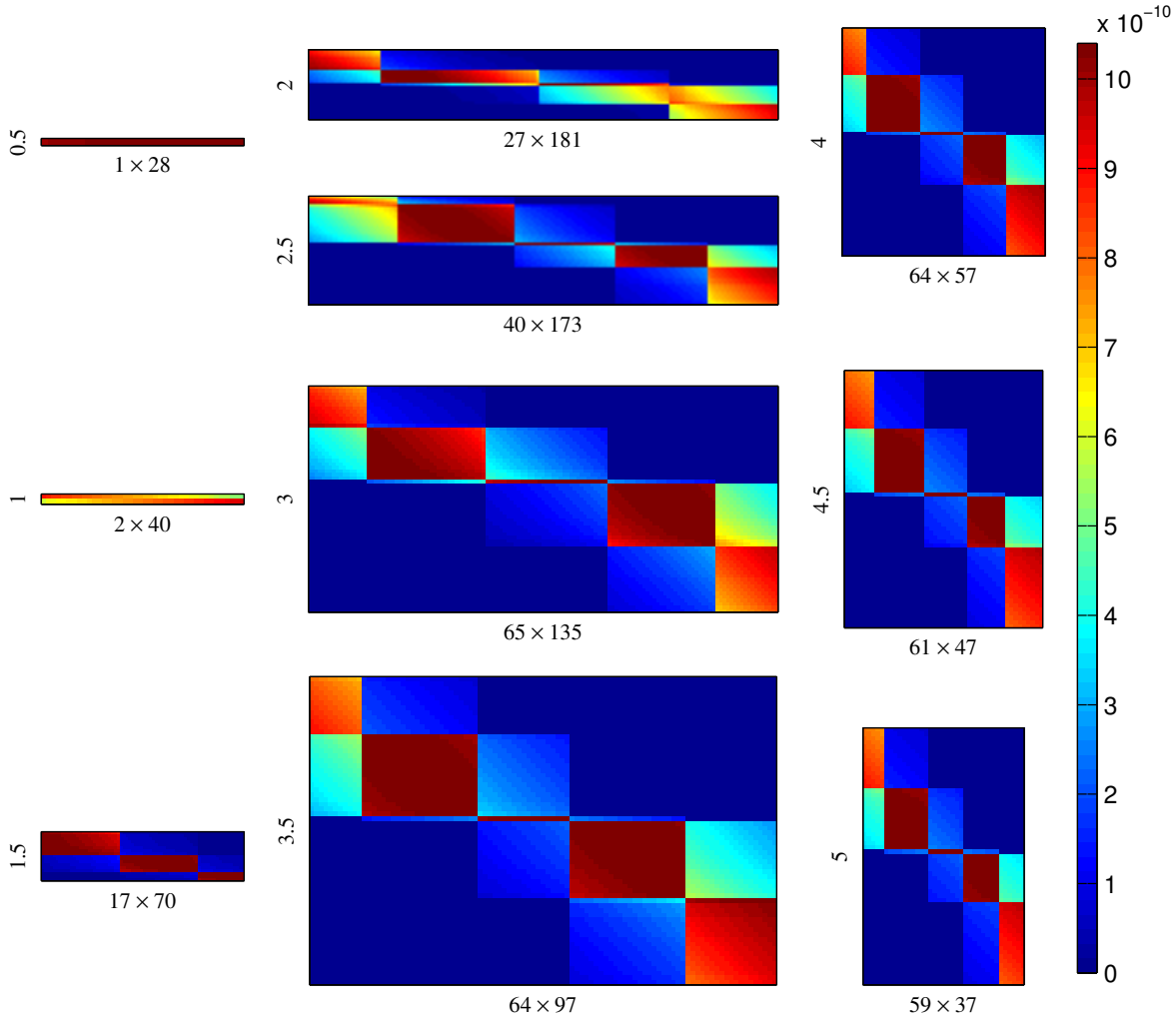

Figure 4: Coevolutionary formation of the adsorption rate interaction network over time. Colour shows pairwise interaction strength ($\phi\theta_{ij}$) for bacterial strains (rows) and phage strains (columns). Network size (number of bacterial strains x number of phage strains) is shown as the x-axis label for each plot, timepoint ($\times 10^7 min$) is shown as the y-axis label. Plots show timepoints corresponding to the dotted grid-lines shown in Figure 2, with time increasing down each column, from top-left to bottom-right.

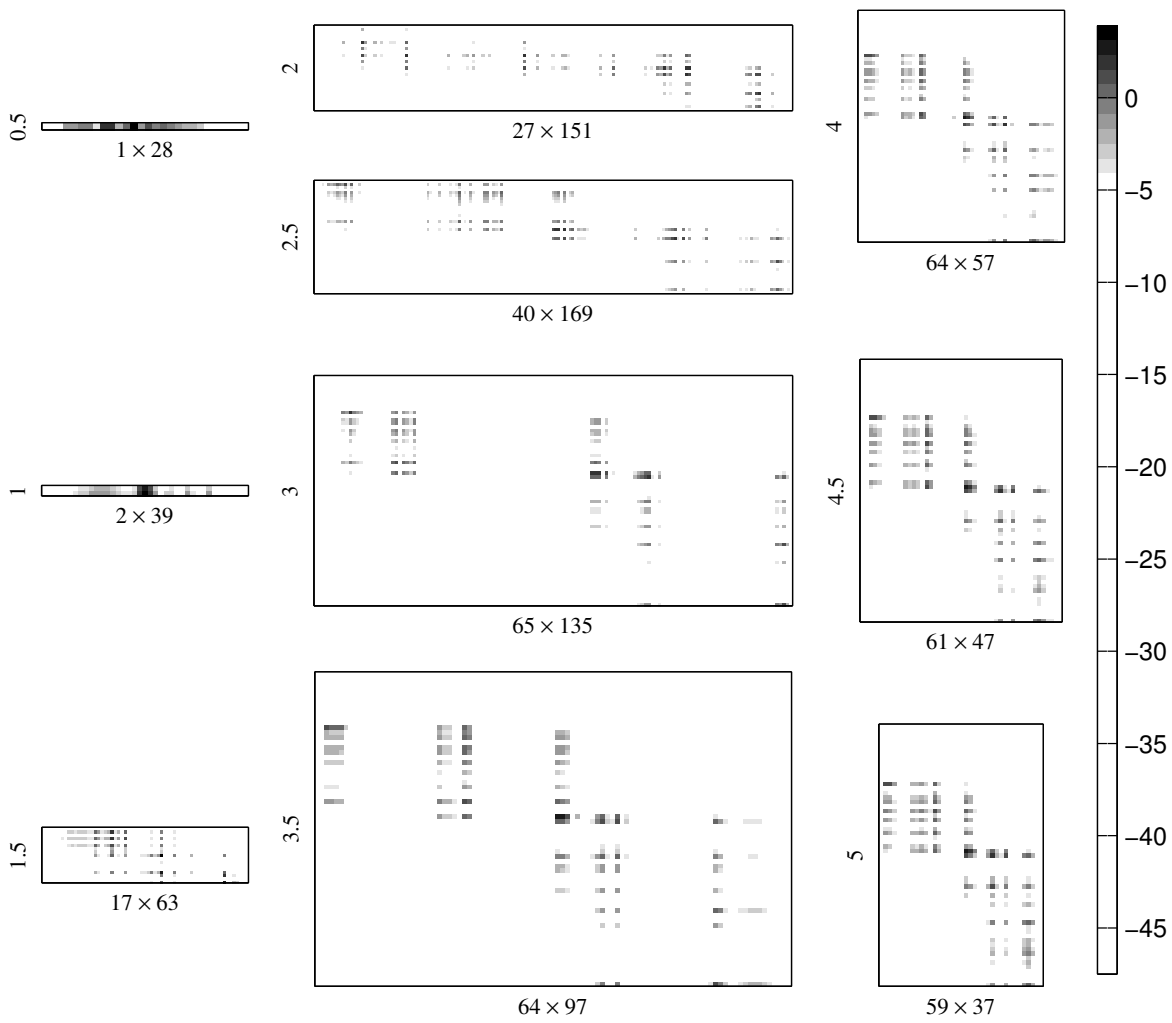

Figure 5: Coevolutionary formation of the infection rate interaction network over time. Grayscale shows pairwise interaction strength (ln $\mid \phi\theta_{ij}N_iV_j \mid$) for bacterial strains (rows) and phage strains (columns). Network size (number of bacterial strains x number of phage strains) is shown as the x-axis label for each plot, timepoint ($\times 10^7 min$) is shown as the y-axis label. Plots show timepoints corresponding to the dotted gridlines shown in Figure 2, with time increasing down each column, from top-left to bottom-right.

Figures 4 and 5 show how the adsorption rate and infection rate interaction networks respectively change over time as bacteria and phage coevolve. Initially there is low diversity and the matrices are small, but over time matrix size increases as hosts and phage diversify. At later timepoints, matrix size reduces somewhat, reflecting an overall drop in diversity as competition excludes weaker strains and the system converges to a quasi-stable state. Matrix sizes show trends in strain diversity, with host diversity rising to a stable level around 60-65 strains and phage diversity initially rising, then falling, during the course of the simulation. This trend is reflected at smaller scale in the sizes of the modules in the adsorption rate matrix. Modular inter-

action structure is always apparent in the adsorption rate matrix, though the number of modules varies over time. Structure in the infection rate matrix is harder to discern visually, though some modularity seems to be apparent at later timepoints.

We converted our quantitative interaction matrices into binary matrices using a threshold filter (see Section 2.3), to give matrices suitable for comparison with reported phage-bacteria interaction networks [22, 23]. The resulting binary networks are a coarse-grained representation of the underlying data, but can be used with the LP-BRIM and NODF algorithms to quantify modularity and nestedness respectively. Figure 6 shows the effect of different thresholds. The size and fill of the resulting binary matrix depends on the threshold used. The importance of choosing an appropriate threshold is well illustrated by the adsorption rate matrices, where choosing too low or too high a threshold results in binary matrices that do not capture the modular structure apparent in the raw data; setting too low a threshold gives overlapping modules, setting too high a threshold may result in loss of some modules. For the remainder of the analysis, we use an adsorption rate threshold of 0.8ϕ and an infection rate threshold of $0.0083\,cells\,(ml\,min)^{-1}$, which typically gave good agreement with visual interpretation of the raw data. The results presented below for adsorption rate matrices are weakly sensitive (but qualitatively robust) to the choice of threshold (data not shown). Results for infection rate matrices are robust to choice of threshold.

The order of rows and columns in the binary matrices can be permuted without changing the underlying network structure. Figure 7 shows binary networks formed from the adsorption rate and infection rate matrices for the timeslice ($t = 3.5\times10^{7}min$) shown in Figure 3. Different row and column re-orderings are applied in each panel. The upper binary matrix shows a random re-ordering that removes the phylogenetic ordering that arises from model formulation (whereby hosts and phages are ordered by genotype), thus representing how unsorted results of an experimental infection assay might appear. The middle binary matrix is sorted to maximise modularity using the LP-BRIM algorithm, with identified modules highlighted in colour. The lower binary matrix is sorted to maximise nestedness using the NODF algorithm.

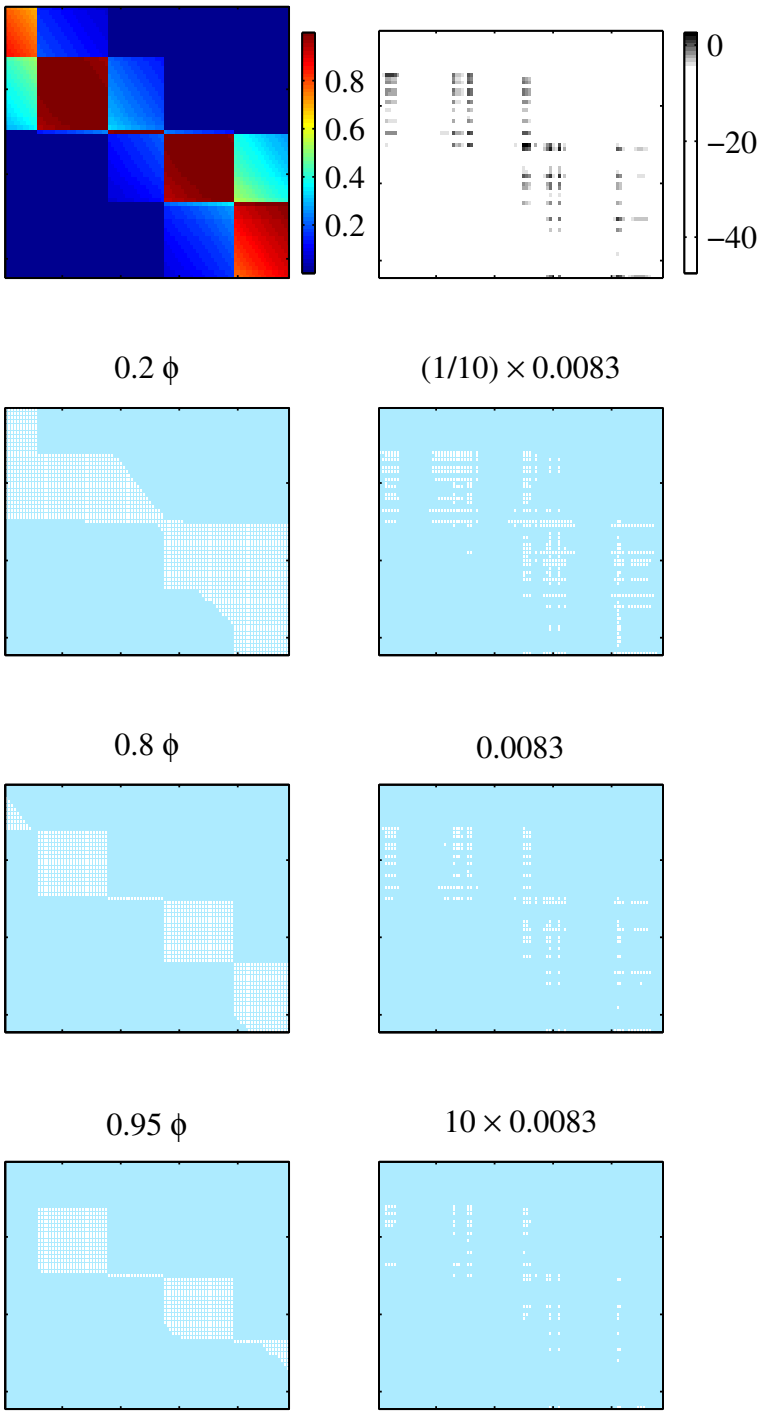

Figure 6: Effect of choice of filter threshold on binary representation of interaction matrices. Top plot in each column shows the raw (unfiltered) matrix data for pairwise adsorption rates (left column) and infection rates (right column). Lower plots show binary matrices formed by filtering raw data with different thresholds. Threshold values of 0.8ϕ for adsorption rates and $0.0083\,cells\,(ml\,min)^{-1}$ for infection rates were used for all subsequent analyses.

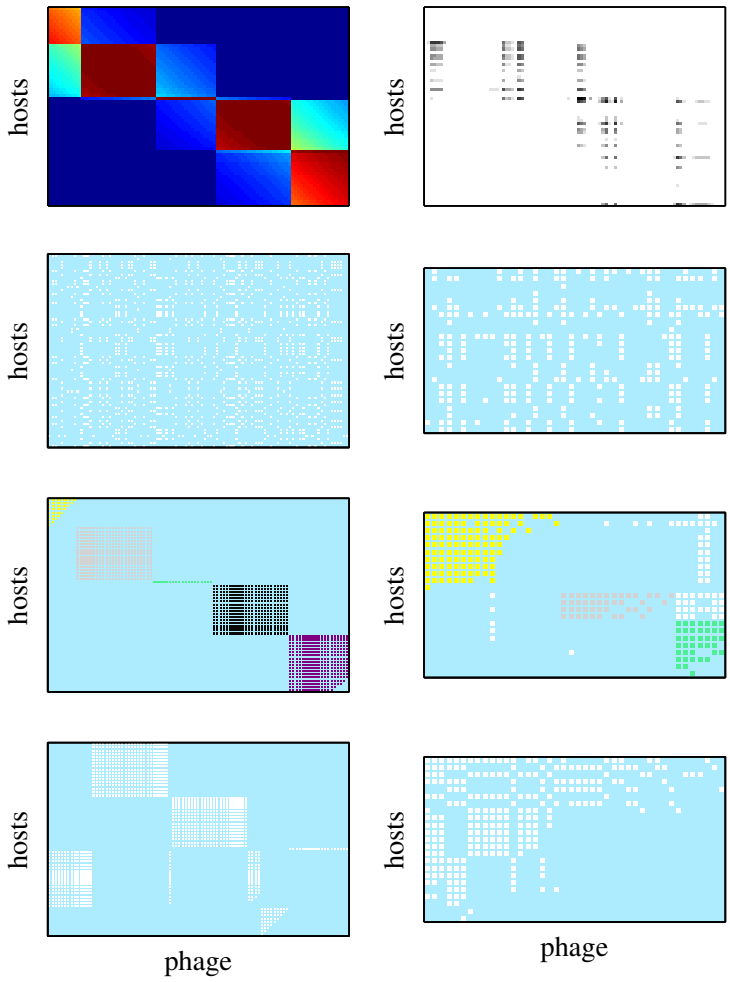

Figure 7: Comparison of binary interaction networks formed by applying a threshold filter to (left column) adsorption rate and (right column) infection rate matrices, for a snapshot of the system at $t = 3.5 \times 10^7 min$, arranged with host strains on rows and phage strains in columns. *Top panel:* quantitative interaction data. *Lower 3 panels:* binary matrices formed using a threshold filter of the quantitative data, sorted for (upper) random permutation, (middle) maximisation of modularity, (lower) maximisation of nestedness.

For the same case study, we used binary interaction matrices to study the temporal dynamics of nestedness (from the NODF algorithm) and modularity (using Barber's modularity score returned by the LP-BRIM algorithm). Figure 8 shows mean values across an ensemble of ten simulations with identical parameters. The ensemble runs differ only in the pseudo-random numbers used in the stochastic mutation process. This stochastic variation leads to different timing and order of evolutionary branching events, but the quasi-steady state reached is similar in all simulations (data not shown). Host strain diversity rises over time before levelling off, while phage strain diversity is highest during the diversification phase and decreases as quasi-steady state is approached.

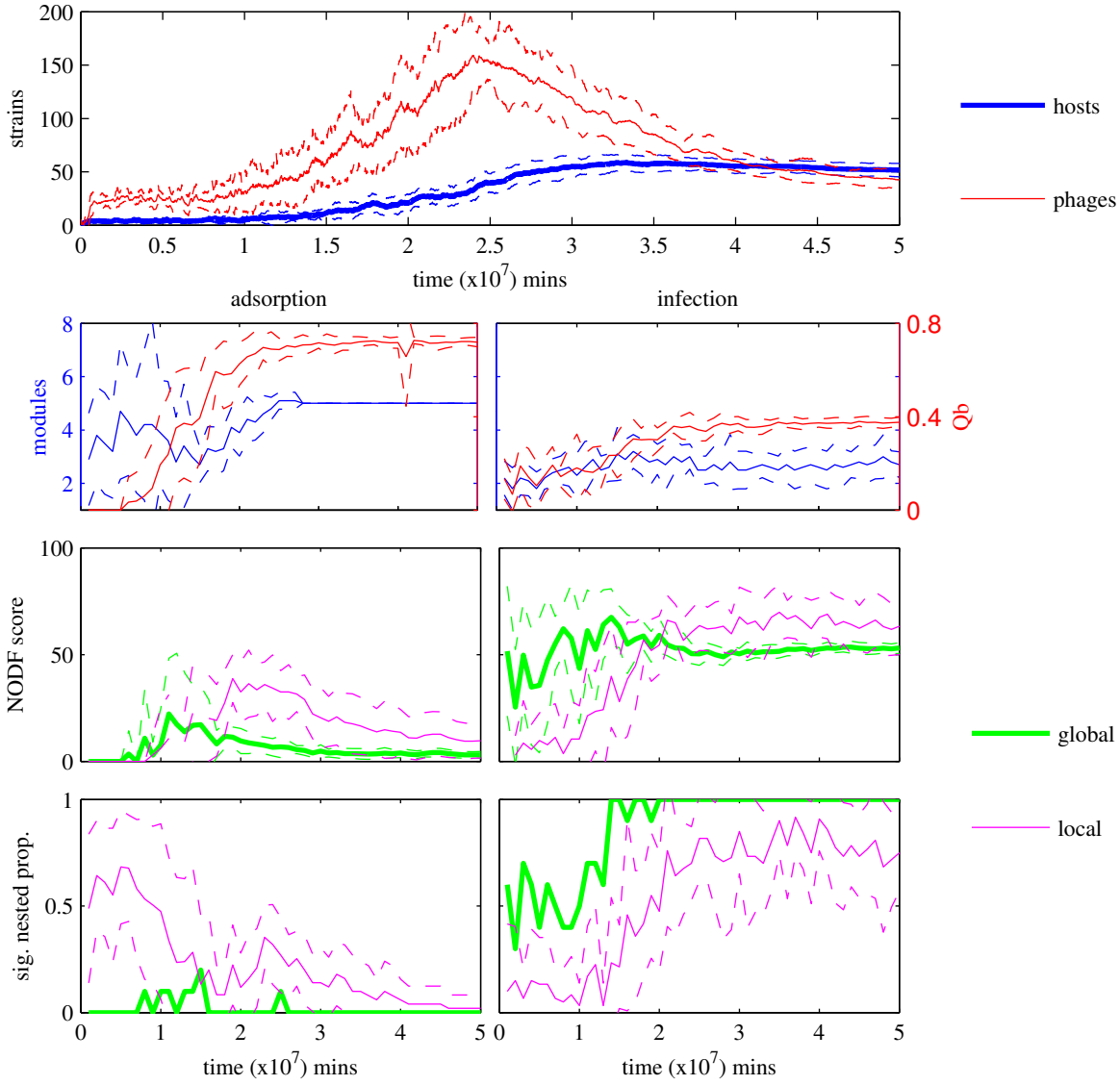

Figure 8: Timeseries of matrix metrics for nestedness and modularity of binary interaction networks during coevolutionary simulations. Data shown are mean values (solid lines) $\pm$ 1 std. dev. (dashed lines) from an ensemble of ten simulation runs with identical parameters but different stochastic mutations. Top: Number of bacterial and phage strains. The following panels show metrics for adsorption rate matrices (left column) and infection rate matrices (right column). Upper-middle: Number of modules detected (left axis) and Barber's modularity score Qb (right axis). Lower-middle: NODF nestedness score of the whole network (global) and mean within-module NODF nestedness score (local) in each simulation. Lower: Proportion of NODF nestedness scores in each simulation that were statistically significant at a level of $p < 0.05$ for the whole network (global) and within-module (local).

Figure 8 (upper-middle row) shows that the interaction structure shown by adsorption rate matrices converges on five clearly identifiable modules that are well-detected by the LP-BRIM algorithm. Module identification is less reliable early in the simulation, when the algorithm sometimes produces false positives (e.g. by identify-

ing “modules within modules”), giving large variances in module number early in the simulations. Fewer modules are detected in the infection rate matrices and there is greater variance in the number detected throughout the simulations. Barber’s modularity metric is significantly higher for the adsorption rate matrices than for the infection rate matrices throughout the simulations, confirming the visual suggestion (e.g. Figure 3) of a stronger modular structure for this form of interaction. We also performed significance tests against Barber’s modularity for the adsorption rate and infection rate matrices for all simulations at $t = 2.5 \times 10^7 min$ and $t = 5 \times 10^7 min$, finding that modularity for all matrices tested was significant at a level of $p < 0.01$.

The lower half of Figure 8 shows timeseries for nestedness (measured by NODF) and statistical significance of nestedness (here we assume statistical significance at a level of $p < 0.05$). Global nestedness of the whole adsorption rate matrix is typically low and rarely statistically significant. In contrast, global nestedness of the whole infection rate matrix is relatively higher and (after the initial diversification phase of the coevolutionary dynamics) almost always statistically significant.

For each simulation, we also calculated nestedness for the modules identified by the LP-BRIM algorithm. Figure 8 (lower four panels) shows the mean within-module NODF score across all modules detected in a particular simulation, as well as the proportion of modules which were statistically nested. Mean within-module nestedness is typically higher than global nestedness for both forms of interaction matrix. Mean within-module nestedness of adsorption rate matrices is typically low and statistically insignificant for most of the timeseries, but rises and shows a higher frequency of statistical significance approaching the mid-point of the simulation runs, when phage diversity is highest. Visual inspection of the adsorption rate matrices showed that a large proportion of identified modules were completely filled (all 1s). These cases give NODF$= 0$ at a significance level of $p = 1$. For the remaining minority of non-filled modules, mean within-module nestedness was typically strong and statisticially significant. Mean within-module nestedness of infection rate matrices is typically high and in most cases statistically significant.

Overall there is a mixed signal from our statistical comparison of binary interaction networks formed from adsorption rates and from infection rates. Both forms of interaction network show a multiscale nested-modular interaction structure to some extent. For adsorption rate matrices, global modularity is very strong while within-module nestedness is often weak and statistically insignificant; however, there is a minority of modules which are strongly and significantly nested during the middle section of the simulations when phage diversity is highest. For infection rate matrices, global modularity is weaker, but identified modules are typically significantly nested. Interestingly, the infection rate matrices are also typically significantly globally nested.

4 Discussion

Here we have shown that a parsimonious 'relaxed lock-and-key' coevolution model based on genetic matching is sufficient to reproduce several core structural features of observed natural communities of bacteria and phage. Negative frequency-dependent selection from phage drives host diversification, which is then mirrored by phage diversification to track their hosts. At steady state, a diverse community of hosts and phage is maintained by kill-the-winner ecological dynamics. Two forms of phage-bacteria interaction network representing the coevolved communities show similar multiscale structural patterns to those observed for natural communities; modularity at large phylogenetic scales and nestedness at smaller scales.

We have aimed with our theoretical study to show that a very simple coevolutionary model can produce nested-modular structures and have selected the chemostat formalism as one of the simplest models in which coevolutionary dynamics can be studied. We do not rule out the possibility that other processes, including spatial structure or multiple resources, might affect observed natural patterns. However, such processes lie beyond the scope of our current study, which aims to show what might occur due to coevolution alone, in the absence of any other (possibly confounding) additional processes. We believe that the most useful models are often the simplest and argue that the use of the chemostat formalism does not affect the generality of our results.

An important caveat is that there is currently only a single dataset [25, 23] showing the macroscale nested-modular interaction structure that the relaxed lock-and-key model produces. Thus it is possible that we are over-estimating the importance of this dataset and hence misjudging the capability of the relaxed lock-and-key model to explain natural community structures. However, there are multiple observations of stable high diversity in natural phage-bacteria communities (e.g. [3, 5, 18, 20]), as well as many examples of nested interaction networks at smaller phylogenetic scales [22]. At larger phylogenetic scales, there are good arguments to support the specificity of infection needed to produce a modular network structure; for example, phage target particular receptors which may only be present in a small number of bacterial lineages, limiting their potential host range [18, 29]. Thus we cautiously suggest that the nested-modular structure should be robust at large phylogenetic scales and propose that additional large-scale cross-infection studies would be a fruitful area for further research.

In the adsorption rate matrices, the nested-modular pattern is most clearly ob-

served when the system is still in the transient diversification phase, i.e. before the system reaches an evolutionary steady state and while the species in the system are still adapting. This section of the coevolutionary dynamics is also where the phage diversity is highest. Since the adsorption rate network includes interactions irrespective of the density of the bacteria/phage strains involved, it likely includes many interactions between low-density strains which are in process of being out-competed and excluded from the system by fitter mutants. This may have implications for observations of natural communities, where overlapping ecological and evolutionary timescales for bacteria and phage [10, 11] imply that many natural communities may not be at evolutionary steady state. We hypothesise that nested-modular structures in density-independent interaction networks will be most obvious when overall adaptation rates within the community are high, for example, in communities adapting to a changed or dynamic environment.

A note of caution must be raised about the methodological grounding for comparisons of model output to empirical observations. The two forms of interaction network that we studied (based either on adsorption rates or on infection rates) showed different statistical features. In this study, adsorption rate networks showed higher global modularity and lower within-module nestedness, while infection rate networks showed lower global modularity and higher within-module nestedness. While both forms of interaction network studied here showed broadly similar multiscale structure, in general the fit to empirical data will depend both on which form of model network was chosen and also on the method by which the empirical network was produced. With theoretical models, all information is accessible - thus our model networks accurately reflect the full diversity of the model community. However, methods for the more difficult task of constructing interaction networks for natural communities inevitably introduce different kinds of bias into the network structure that is output. A common experimental method for determining interaction networks for natural phage-bacteria communities appears to be to collect a sample from the natural environment, isolate as many strains as possible, then use a plaque assay to test for infection of each potential host by each phage. However, sampling inevitably carries a bias towards collecting only the more numerous strains, of which only a small fraction will be cultivable [37, 38]. Thus while empirical studies reporting natural interaction networks may aim to present a complete record of all strains present in the community (i.e. to produce something similar to our model adsorption rate networks), they may actually (without any failure of the experimental method) be presenting networks more akin to our infection rate networks, which only include interactions between abundant strains. Thus it is not clear which of our binary interaction networks should be compared to the networks reported for natural communities.

Another methodological issue surrounds the comparison of weighted and binary interaction networks; here we have converted quantitative networks to binary networks to aid comparison with empirical data. Infectivity assays commonly only measure presence-absence of infection, rather than the messier rate/affinity data that indicate the quantitative strength of interactions, which are harder to measure [39, 40, 24]. Thus empirical data is often given in binary form and analysed using statistical tools developed for binary matrices. Conversion of quantitative data into binary form loses information and inevitably introduces bias that will accentuate some features and mask others. A challenge for empirical researchers is to develop methods for measuring interaction strengths between phage and bacterial strains, rather than just presence-absence of interaction [40]. A challenge for theoretical researchers is to develop better statistical tools for analysing the weighted phage-bacteria interaction networks that will thereby be produced.

The original kill-the-winner model of aquatic virus ecology [16] describes one-to-one interactions between viruses and bacteria, such that no cross-infections occur. In that scenario, specialised infection leads to negative density-dependent predation from viruses, which favours rare bacteria phenotypes and acts to maintain diversity. This contradicts available data for real phage-bacteria systems [22, 23]. The relaxed lock-and-key model allows for cross-infection based on genetic similarity, while producing a stable diverse community maintained by kill-the-winner dynamics. Observations of natural kill-the-winner dynamics [19] are thus consistent with the relaxed lock-and-key model. We note that several alternative coevolution models are unlikely to capture nested-modular interaction structures and kill-the-winner ecological dynamics. Gene-for-gene genetic models are consistent with arms race dynamics [41] and transitive range expansion, but do not permit stable high diversity without the addition of explicit trade-offs [14] (and even then diversity is typically limited to dimorphism [8]). Matching-alleles genetic models can drive diversification [14, 41], but do not produce the modular structure or within-module nestedness.

There remain other mechanisms that will affect the interactions between bacteria and phage and several of these could potentially produce high diversity and nested-modular network structures. The relaxed lock-and-key model is perhaps most easily interpreted as representing coevolution of phage tail-fibres and bacterial cell-surface receptors. However, the real infection process is multi-stepped and coevolution may occur at different stages, including initial adsorption to extracellular surface receptors [42] and also subsequent intracellular defence mechanisms [43, 29, 44, 45, 46]. Additionally our model does not take into account non-mutational processes of genetic variation, such as gene loss and horizontal gene transfer [40, 18], or the possible role of environmental heterogeneity and/or spatial localisation [23]. Any of these factors

can affect natural community structures and could in principle have produced the patterns created here by coevolution. The strength of the relaxed lock-and-key model is by appeal to Occam's razor; it offers a parsimonious and sufficient explanation of multiple observed phenomena.

Acknowledgements

S.J.B and H.T.P.W acknowledge financial support from the University of Exeter. We would like to thank Richard Boyle and two anonymous reviewers for their comments that have helped improve this article.

References

[1] Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**(6736), 541–548.

[2] Wilhelm, S. W. and Suttle, C. A. 1999. Viruses and nutrient cycles in the sea. *Bioscience* **49**(10), 781–788.

[3] Weinbauer, M. G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28**(2), 127–181.

[4] Suttle, C. A. 2005. Viruses in the sea. *Nature* **437**(7057), 356–361.

[5] Suttle, C. A. 2007. Marine viruses – major players in the global ecosystem. *Nature Reviews Microbiology* **5**(10), 801–812.

[6] Brussaard, C. P. D., Wilhelm, S. W., Thingstad, F., Weinbauer, M. G., Bratbak, G., Heldal, M., Kimmance, S. A., Middelboe, M., Nagasaki, K., Paul, J. H., et al. 2008. Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME Journal* **2**(6), 575–578.

[7] Weitz, J. S. and Wilhelm, S. W. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports* **4**(17).

[8] Bohannan, B. J. M. and Lenski, R. E. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecology Letters* **3**, 362–377.

[9] Buckling, A. and Rainey, P. B. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**(1494), 931–936.

[10] Lennon, J. T. and Martiny, J. B. H. 2008. Rapid evolution buffers ecosystem impacts of viruses in a microbial food web. *Ecology Letters* **11**(11), 1178–1188.

[11] Marston, M. F., Pierciey, F. J., Shepard, A., Gearin, G., Qi, J., Yandava, C., Schuster, S. C., Henn, M. R., and Martiny, J. B. H. 2012. Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences* **109**(12), 4544–4549.

[12] Brockhurst, M. A., Morgan, A. D., Fenton, A., and Buckling, A. 2007. Experimental coevolution with bacteria and phage: The *Pseudomonas fluorescens*—$\phi 2$ model system. *Infection, Genetics and Evolution* **7**, 547–552.

[13] Forde, S. E., Thompson, J. N., Holt, R. D., and Bohannan, B. J. M. 2008. Coevolution drives temporal changes in fitness and diversity across environments in a bacteria–bacteriophage interaction. *Evolution* **62**(8), 1830–1839.

[14] Agrawal, A. and Lively, C. M. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research* **4**(1), 79–90.

[15] Gandon, S., Buckling, A., Decaestecker, E., and Day, T. 2008. Host–parasite coevolution and patterns of adaptation across time and space. *Journal of Evolutionary Biology* **21**(6), 1861–1866.

[16] Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45**(6), 1320–1328.

[17] Winter, C., Bouvier, T., Weinbauer, M. G., and Thingstad, T. F. 2010. Trade–offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. *Microbiology and Molecular Biology Reviews* **74**(1), 42–57.

[18] Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Beltran Rodriguez-Brito, L. P., Thingstad, T. F., and Forest Rohwer, A. M. 2009. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7**(11), 828–836.

[19] Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy, J., Martin-Cuadrado, A. B., Mira, A., Nulton, J., Pasic, L., Rayhawk, S., Rodriguez-Mueller, J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T. F., Tran, T., Thurber, R. V., Willner, D., Youle, M., and Rohwer, F. 2010. Viral and microbial community dynamics in four aquatic environments. *ISME J* **4**(6), 739–751.

[20] Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D. 2011. Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* **474**(7353), 604–608.

[21] Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., Ellisman, M., Deerinck, T., Sullivan, M. B., and Giovannoni, S. J. 2013. Abundant SAR11 viruses in the ocean. *Nature* **494**(7437), 357–360.

[22] Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., and Weitz, J. S. 2011. Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences* **108**(28), E288–E297.

[23] Flores, C. O., Valverde, S., and Weitz, J. S. 2013. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *ISME Journal* **7**, 520–532.

[24] Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B., and Hochberg, M. E. 2013. Phage-bacteria infection networks. *Trends in Microbiology* **21**(2), 82 – 91.

[25] Moebus, K. and Nattkemper, H. 1981. Bacteriophage sensitivity patterns among bacteria isolated from marine waters. *Helgoländer Meeresuntersuchungen* **34**(3), 375–385.

[26] Weitz, J. S., Hartman, H., and Levin, S. A. 2005. Coevolutionary arms races between bacteria and bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* **102**(27), 9535–9540.

[27] Williams, H. T. P. Coevolving parasites improve host evolutionary search on structured landscapes. In Artificial Life 13: Proceedings of the Thirteenth International Conference on the Simulation and Synthesis of Living Systems, Adami, C., Bryson, D. M., Ofria, C., and Pennock, R. T., editors, 129–136 (MIT Press, Cambridge, MA, 2012).

[28] Williams, H. T. P. 2013. Phage-induced diversification improves host evolvability. *BMC Evolutionary Biology* **13**, 17.

[29] Labrie, S. J., Samson, J. E., and Moineau, S. 2010. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**(5), 317–327.

[30] Levin, B. R., Stewart, F. M., and Chao, L. 1977. Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. *American Naturalist* **111**, 3–24.

[31] Monod, J. 1949. The growth of bacterial cultures. *Annual Reviews in Microbiology* **3**(1), 371–394.

[32] Poisot, T. and Flores, C. 2012, BiWeb: https://github.com/tpoisot/biweb. Technical report.

[33] Barber, M. J. 2007. Modularity and community detection in bipartite networks. *Physical Review E* **76**(6), 066102.

[34] Liu, X. and Murata, T. Community detection in large-scale bipartite networks. In *Web Intelligence and Intelligent Agent Technologies 2009 (WI-IAT'09)*, volume 1, 50–57. IEEE, 2009.

[35] Almeida-Neto, M., Guimaraes, P., Guimarães, P. R., Loyola, R. D., and Ulrich, W. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**(8), 1227–1239.

[36] Beckett, S. J. and Williams, H. T. P. 2013, FALCON: https://github.com/sjbeckett/falcon. Technical report.

[37] Edwards, R. A. and Rohwer, F. 2005. Viral metagenomics. *Nature Reviews Microbiology* **3**(6), 504–510.

[38] Stewart, E. J. 2012. Growing unculturable bacteria. *Journal of Bacteriology* **194**(16), 4151–4160.

[39] Jover, L. F., Cortez, M. H., and Weitz, J. S. 2013. Mechanisms of multi–strain coexistence in host–phage systems with nested infection networks. *Journal of Theoretical Biology* **332**, 65 – 77.

[40] Koskella, B. and Meaden, S. 2013. Understanding bacteriophage specificity in natural microbial communities. *Viruses* **5**(3), 806–823.

[41] Quigley, B. J. Z., López, D. G., Buckling, A., McKane, A. J., and Brown, S. P. 2012. The mode of host–parasite interaction shapes coevolutionary dynamics and the fate of host cooperation. *Proceedings of the Royal Society B: Biological Sciences* **279**(1743), 3742–3748.

[42] Scanlan, P. D., Hall, A. R., Lopez-Pascua, L. D. C., and Buckling, A. 2011. Genetic basis of infectivity evolution in a bacteriophage. *Molecular Ecology* **20**(5), 981–989.

[43] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**(5819), 1709–1712.

[44] Stern, A. and Sorek, R. 2011. The phage–host arms race: Shaping the evolution of microbes. *Bioessays* **33**(1), 43–51.

[45] Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S., and Koonin, E. V. 2012. Viral diversity threshold for adaptive immunity in prokaryotes. *mBio* **3**(6), e00456–12.

[46] Seed, K. D., Lazinski, D. W., Calderwood, S. B., and Camilli, A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**(7438), 489–491.