ARTICLES
https://doi.org/10.1038/s41559-020-01312-z

nature ecology & evolution

Check for updates

The network structure and eco-evolutionary dynamics of CRISPR-induced immune diversification

Shai Pilosof[1,2], Sergio A. Alcalá-Corona[2], Tong Wang[3,4], Ted Kim[4,5], Sergei Maslov[4,6], Rachel Whitaker[4,5] ✉ and Mercedes Pascual[2,7] ✉

As a heritable sequence-specific adaptive immune system, CRISPR–Cas is a powerful force shaping strain diversity in host–virus systems. While the diversity of CRISPR alleles has been explored, the associated structure and dynamics of host–virus interactions have not. We explore the role of CRISPR in mediating the interplay between host–virus interaction structure and eco-evolutionary dynamics in a computational model and compare the results with three empirical datasets from natural systems. We show that the structure of the networks describing who infects whom and the degree to which strains are immune, are respectively modular (containing groups of hosts and viruses that interact strongly) and weighted-nested (specialist hosts are more susceptible to subsets of viruses that in turn also infect the more generalist hosts with many spacers matching many viruses). The dynamic interplay between these networks influences transitions between dynamical regimes of virus diversification and host control. The three empirical systems exhibit weighted-nested immunity networks, a pattern our theory shows is indicative of hosts able to suppress virus diversification. Previously missing from studies of microbial host–pathogen systems, the immunity network plays a key role in the coevolutionary dynamics.

Microbial hosts have developed a range of resistance mechanisms against viruses. For example, mutations in receptors leading to surface resistance, innate immunity obtained with restriction modification systems and adaptive immunity obtained via CRISPR and its CRISPR-associated (Cas) proteins[1–3]. These and other defence mechanisms can operate alone or in tandem to affect the diversification of microbes and viruses[4,5] as well as structuring the complex infection networks into which this diversity is organized[6–8]. Infection structure ('who infects whom') is a crucial feature of microbe–virus interactions because, as for other host–parasite systems, it may affect the evolution[9,10], population stability[11] and transmission dynamics[12] of both partners. Hence, given the crucial role microbes and viruses play in virtually all of Earth's ecosystems, infection structure may have far-reaching consequences for ecosystem functions.

The microbe–virus interaction structure is non-random, with two dominant macroscopic topologies in its descriptions: modularity and nestedness[13,14]. Modularity concerns patterns of specificity, where the network is partitioned into modules of microbes and viruses that interact densely with each other but sparsely with those outside the group. Nestedness reflects instead patterns of specialization where specialist microbes are infected by subsets of viruses that in turn also infect the more generalist microbes[15]. An analysis of a large assemblage of bacteria–virus infection networks found that these are predominantly nested rather than modular[15] but a later study suggested that structure depends on phylogenetic scale, with modularity at large phylogenetic scales of species interactions and nestedness among interacting strains of the same species[16]. Several hypotheses have been put forward to explain the emergence of modularity and nestedness in bacteria–phage infection networks from immunity-related coevolutionary dynamics[6,7] (mutations that create new CRISPR alleles and variation in the relative frequencies of alleles over time). For example, Fortuna et al.[8] showed that bacteria–virus networks evolve nestedness under directional arms race dynamics, but not under fluctuating selection.

While providing crucial knowledge about the structure of microbe–virus interactions, all previous studies have focused exclusively on patterns of infection. However, infection structure should critically depend on the genetic basis of resistance of hosts to pathogens[10]. The emergence of network structure from immune selection has been shown in pathogens of humans such as *Plasmodium falciparum*[17,18]. So far, these studies have only analysed networks of genetic similarity from the parasite perspective, thereby inferring the selective impact of hosts because information on host immunity structure and diversity is typically absent. Hence, how immunity structures host–parasite interactions, for example, into modular or nested topologies, is unknown. In contrast, CRISPR is a heritable adaptive immune system conferring sequence-specific protection against viruses, plasmids and other mobile elements[1,19]. The CRISPR system functions as an adaptive immune system by incorporating DNA segments called 'protospacers' of infecting viruses into host genomes as 'spacers' that constitute sequence-specific immunity and memory[19]. Hence, it provides a direct sequence-based link that allows consideration of associated structures for both infection and immunity, and from both host and parasite perspectives.

[1]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [2]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. [3]Department of Physics, Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [4]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [5]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [6]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [7]Santa Fe Institute, Santa Fe, NM, USA. ✉e-mail: rwhitaker@life.illinois.edu; pascualmm@uchicago.edu

The diversity of CRISPR alleles in natural, experimental and simulated populations has been explored[20–24]. Simulation studies[20] followed by experimental tests showed that increased strain diversity of *Pseudomonas aeruginosa* (with strains being clones with different spacer compositions) promotes the ability of the bacterium to control the DMS3vir virus[22]. Nevertheless, how such strain diversity evolves, the emergence of the structure of host–virus interaction networks from molecular mechanisms that link genotype to defensive phenotype and how networks are assembled and disassembled have not been unexplored.

In this study, we address how the genetic basis of CRISPR immunity shapes the structure of both infection and immunity networks, and in turn, how the structure of the networks shapes the population dynamics of viruses and their hosts. Because we consider not just the presence/absence of immunity but also the weight or redundancy of immunity links, the two networks are not simply complementary and the structure of one cannot inform us about that of the other. Using a system of Lotka–Volterra-like differential equations coupled with stochastic evolution (mutation of protospacers in viruses and spacer acquisition in hosts), we investigate the emergence and consequence of the structure of immunity networks. We show that infection and immunity networks exhibit distinct structures emerging from CRISPR immunity that are involved in a constant interplay with the diversification and population dynamics of host and viral strains. These structures influence dynamics with eco-evolutionary feedbacks.

Results

Dynamics shift between virus diversification and host control. Previous work has shown the possibility of diverse populations, containing many strains, with distributed immunity of the host[20,25]. The dynamics of these populations can exhibit two alternating major regimes: escape and diversification of the viruses and dominance of the microbial host, respectively[25]. We extended the model to allow for larger host and virus richness and to track the coevolutionary history of both hosts and viruses through time (Methods). Regime switching is an emergent property of the model and the timing of the switching between the two regimes is defined and identified in this study based on the dynamics of virus abundance (Methods). In the 'virus diversification regime' (VDR), virus strains proliferate and diversify while in the 'host-controlled regime' (HCR) host strains are able to constrain virus diversification and lead to their extinction (Fig. 1). During the former, viruses and hosts coexist with fluctuating abundances, whereas during the latter, hosts reach carrying capacity and viruses exhibit declining abundances and richness followed by either escape or extinction (Extended Data Figs. 1–3).

Due to the stochastic nature of our simulations, dynamics can quantitatively change. For example, the number of alternations between regimes varies. We present in the main text results for a single simulation but these hold for multiple simulations (Supplementary Information). For our set of parameters, the long-term dynamics of the system consists of complete virus extinction during one of the HCRs. In most simulations (about 70%), viruses do not go rapidly extinct and exhibit instead one or more transient periods of diversification (Fig. 1), whereas in the rest, they show rapid extinction as hosts exert immediate control. Our analyses concentrate on the alternating transient dynamics that precede extinction.

Variation in viral or host abundance alone cannot drive the transitions between these two regimes since these do not exist in the corresponding Lotka–Volterra dynamics under neutral conditions of hosts not acquiring specific immunity (Supplementary Methods). An alternative explanation is that the structure of strain diversity, emerging from the eco-evolutionary dynamics via specific immunity, explains the transitions between these dynamical regimes. By 'eco-evolutionary dynamics' we refer to the dynamic interplay between ecological (for example, population dynamics such as changes in relative abundances) and evolutionary (for example, mutation) processes. To investigate the structure of diversity in this complex system, we consider two complementary bipartite networks, the 'infection' and 'immunity' networks, through time. In these networks, a node represents a strain of a virus (unique combination of protospacers) or a host (unique combination of spacers) and the edges represent a given type of interaction, either infection or protection from infection (Fig. 2).

Modularity of infection networks represents niches for hosts. At the start of the simulations and during VDRs, the infection network is built over time by the addition of host and virus strains (Extended Data Fig. 4). We find that at the end of each VDR, after the infection network has been built, the infection network is significantly modular ($P < 0.001$ in two of three VDRs; in the first VDR the network was too small to test; Methods). That is, infections are concentrated within groups of viruses and hosts, with more edges within than between these groups (Fig. 3b). These modules reflect different niches for virus growth whereby hosts are resources that a group of viruses is able to infect. The existence of these niches is linked to the structured genetic diversity of the hosts. Specifically, modules in host–spacer networks delineate groups of hosts that share the same spacers and are therefore susceptible to similar viruses. We found that host–spacer networks aggregated across the VDR are significantly modular ($P < 0.001$ in two of three VDRs).

The importance of immune selection in the formation of these niches can be demonstrated by asking whether clonal expansion alone could account for the modularity of the infection network. This generally is not the case: there is no consistent phylogenetic signal in the infection and host–spacer networks of the VDRs. This is shown by comparing the observed phylogenetic distance within modules to distances obtained in randomized versions of the corresponding network (Methods and Supplementary Tables 1 and 2). Moreover, in the absence of host immune memory, diversification of the viruses and coexistence of different strains is not observed. This is demonstrated by a neutral model where all the processes of the full system are retained except for the specific immunity of the host (Supplementary Methods).

The persistence of the modular structure in the infection network and the coexistence of a diverse community of viral strains is only transitory because the number of susceptible hosts available to all viruses declines rapidly and the hosts available to a given group of viruses within a module also decline in number by either lysis or acquisition of further immunity. This closing of niches, coupled with the inability of the viruses to escape makes the VDR effectively transient. To understand why virus escape is highly unlikely, we turn next to the structure of the immunity network.

Weighted nestedness of immunity networks is indicative of host control. Diversification of viruses during the VDR increasingly diversifies the host population (Extended Data Fig. 2). In other words, escape from host control via mutation of protospacers allows higher abundances of particular virus strains, which therefore also experience higher encounter rates with hosts in general. This leads to the increasing acquisition of new spacers through either the failure of their previous immunity or infection by an escape mutant. Such red queen coevolutionary dynamics (that is, the reciprocal evolution of hosts and viruses) progressively adds spacers to hosts during the VDR, building the immunity network (Fig. 3a). In this network, edges indicate at least one spacer–protospacer match and the weight of the edge encodes the number of different matches protecting a given host from a given virus (Fig. 2c).

In the immunity network, an edge value larger than 1 indicates redundancy in the protection of a given host to a given virus. In other words, the host is protected from this given virus by more

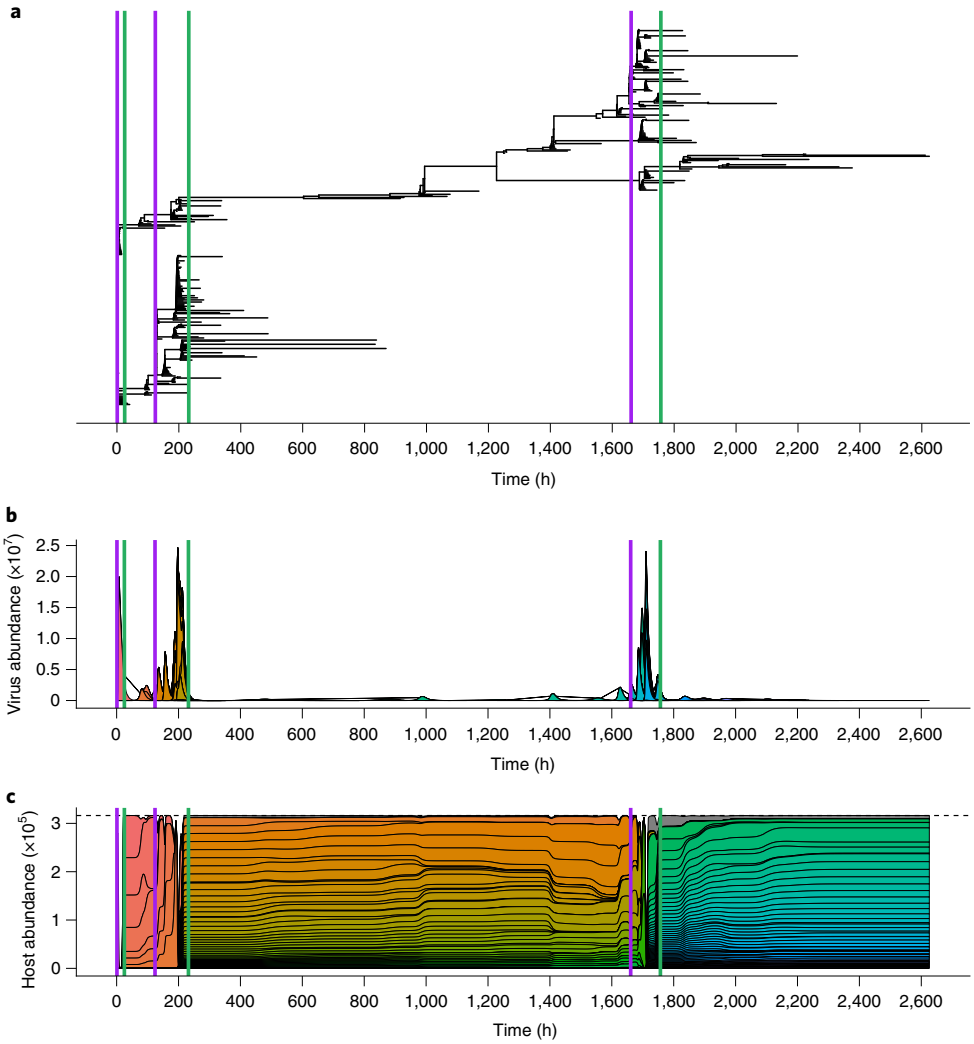


Fig. 1 | An example of typical dynamics with regime shifts. a, VDRs start at the purple vertical lines and end in the green vertical lines, when the HCRs begin. The viral phylogenetic tree shows virus diversification during VDRs and virus extinctions during the HCRs. Intermittent virus diversification in the later part of HCRs eventually leads to escape and the initiation of a new VDR. Complete virus extinction occurs at the end of this simulation during the last HCR. **b**,**c**, Abundance profiles of viruses (**b**) and hosts (**c**), respectively. The 100 most abundant strains are coloured, the rest are aggregated and shown in grey. Note that virus growth does not imply a decrease in host richness while host dominance decreases virus abundance and richness sharply. Virus diversification implies a temporary escape from host control and their rise in abundance, which in turn increases encounters with hosts. The resulting rise in per capita infection rates of hosts leads to their concomitant diversification because new host strains are generated by acquiring at least one new protospacer. Therefore, despite an initial decline in the abundance of hosts, their richness typically rebuilds (Extended Data Figs. 1 and 3).

than one match, even though a single match already confers immunity. We find that this redundancy has a characteristic quantitative nested structure[26]. That is, we can order the network such that hosts with immunity to most viruses also have more matches (that is, increasing levels of redundancy) to those viruses and subsequent hosts (from top to bottom) are immune to subsets of viruses, via fewer matches (Figs. 2g and 3a). Similarly, each virus strain (from left to right) can infect a progressively smaller subset than the virus following it, also via fewer matches. Although redundancy per se does not necessarily imply nestedness, the diversification and associated acquisition of immune memory in time leads to the particular structure of weighted nestedness.

The weighted-nested structure is assembled by a complex interplay of temporal changes in abundances, associated encounter rates and selection pressures. Hosts can increase matches to a virus strain by exposure to other strains that carry an overlapping protospacer, building redundancy. In addition, how the matrix structure is woven from bottom to top, and left to right, with the respective addition of new hosts and viruses, is reflected in the relationship between the order of the rows and columns and strain age (Extended Data Fig. 5). It is also the result of CRISPR failure, which allows infection and the subsequent acquisition of additional spacers; that is, when a virus infects a protected host whose CRISPR system failed, then the host can (if it does not die) obtain a new spacer.

A nested structure provides universal resistance to the epidemic growth of viruses when redundancy is widespread. This happens when the immunity network is close to fully connected, with insufficient zero matches to support the growth of viral strains, leading to their extinction. This inability to grow does not mean however immediate or complete extinction of all viruses because the decline in abundance of their populations takes time. During this time, virus strains can still mutate. When infection occurs in unprotected hosts, mutant offspring will be generated but will find very low and insufficient host numbers to support epidemic growth. Moreover, when infection of protected viruses occurs because immunity fails, errors in viral replication can lead to the replacement of a random

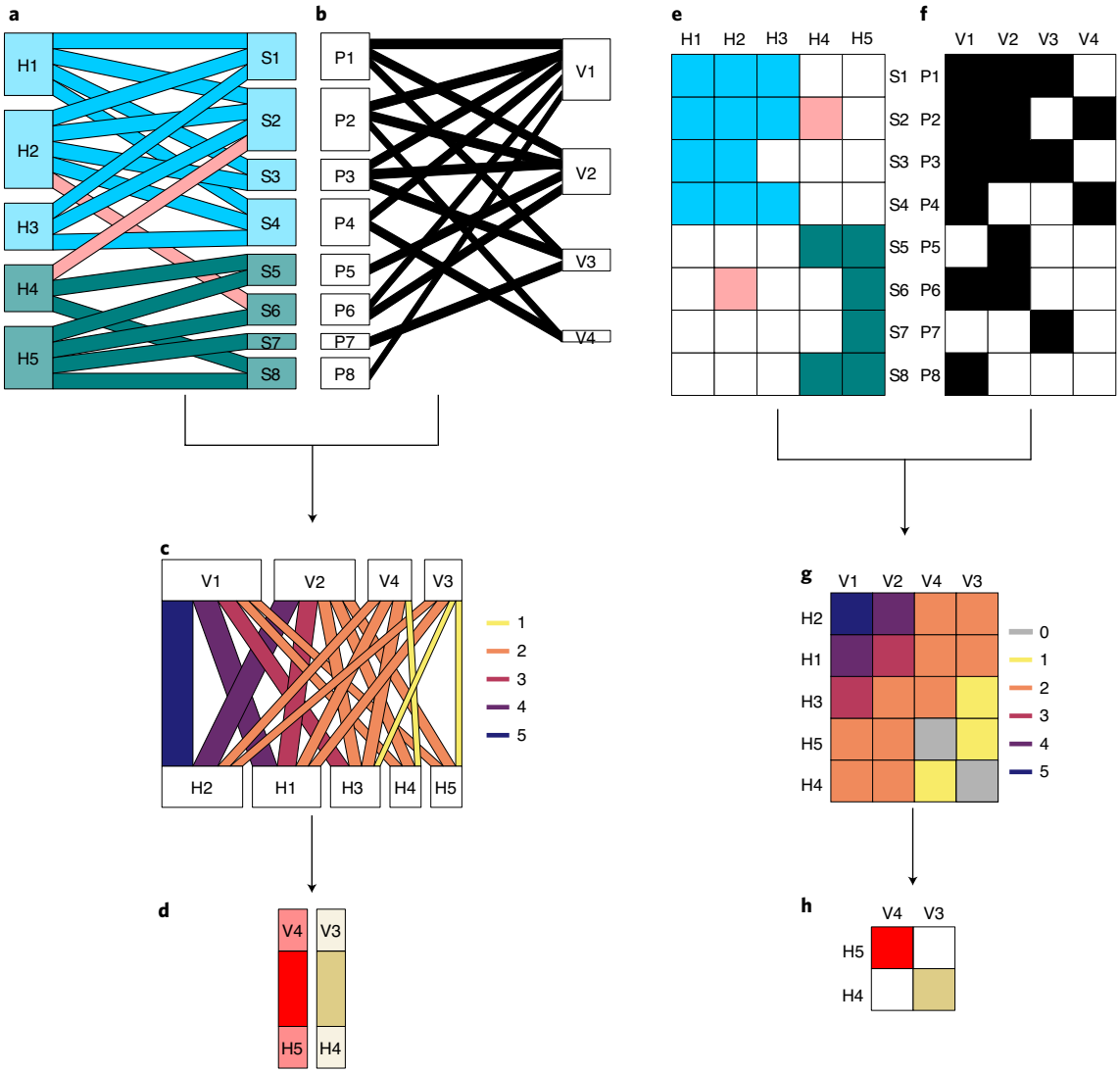


Fig. 2 | Structures of diversity. Diagrams illustrating the two different kinds of networks (left column) and associated matrices (right column) used in this study and how they are built. The toy example has five hosts (H1–H5), four viruses (V1–V4), eight spacers (S1–S8) and eight protospacers (P1–P8). **a**, A bipartite network depicting the spacer composition of hosts. Hosts are affiliated to one of two modules (depicted in light blue and dark green) and interactions can fall either within a module (coloured as the modules) or outside the module (pink). **b**, A bipartite network (not modular) depicting the protospacer composition of viruses. **c**, The immunity network is created by counting the number of shared spacers and protospacers between pairs of hosts and viruses (that is, matches). Interactions in the network are weighted by the number of matches, depicted by different colours and width. **d**, The infection network is created by considering unrealized interactions in the immunity network (equivalent to 0 matches in **g**). In this example, only two such interactions exist, between two viruses and two hosts. There are two modules, depicted in colours, with interactions occurring within the modules only. **e**, The counterpart matrix of the network in **a**. Interactions (matrix cells) depict the occurrence of a spacer in a host strain and are coloured as in **a**. **f**, The counterpart matrix of the network in **b**. The matrix cells depict the occurrence of a protospacer in a virus strain. **g**, The counterpart matrix of the network in **c**, with colours depicting the number of spacer matches. The matrix is organized by the sum of columns and rows and is quantitatively nested. **h**, The counterpart matrix for the network in **d**. The two interactions correspond to the grey empty matches in **g**.

protospacer. However, mutant offspring are unlikely to escape given that there are redundant matches to the original host. How the structure of the immunity network implies its own change during this period of virus extinction in ways that can ultimately allow and facilitate virus escape is described next.

Order in virus extinctions facilitates their eventual escape. The nested pattern in the redundancy of immunity, built during a VDR, enforces order and predictability in virus extinctions during the subsequent HCR. Viruses for which most hosts have acquired immunity via multiple matches will tend to go extinct earlier than those with fewer matches, without an opportunity to escape protection via mutant offspring (Fig. 3a and Extended Data Fig. 6). Multiple matches reflect an earlier timing of epidemic growth of the given virus strain and a concomitant depletion of the host populations this growth was based on. Somewhat paradoxically, this orderly extinction can facilitate a new viral escape and the stochastic initiation of another virus expansion cycle since it reduces redundancy and the associated nestedness structure by preferentially removing viruses with a high number of matches.

The decrease in weighted nestedness after the onset of the HCR is illustrated in Fig. 4a. Concomitantly, the extinction process progressively elevates the proportion of 0 and 1 matches (out of all interactions) between hosts and viruses in the network (Fig. 4b).

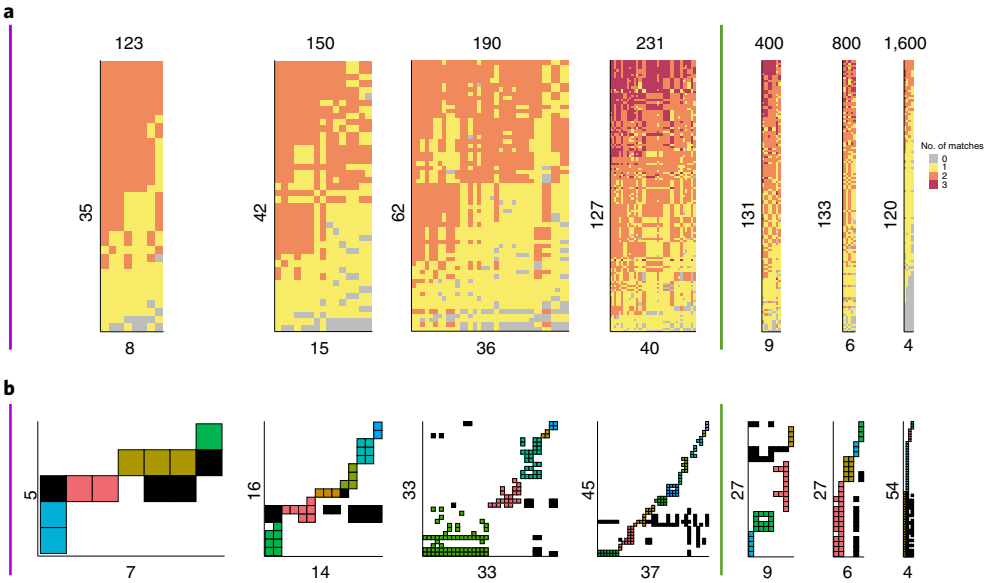


Fig. 3 | Snapshots of network structure during the two regimes. Networks are defined as in Fig. 2. Each network is a snapshot of the population, with time steps depicted above the networks corresponding to Fig. 4. The numbers in the *x* and *y* axes indicate the number of viruses and hosts, respectively, participating in each of the networks. Because the size of the matrix is limited, the width and height of the networks are not necessarily proportional to the number of strains. (A reduced number of strains results in panels with larger squares.) The purple and green lines depict the initiation of a VDR and a HCR regime, respectively. **a**, Weighted nestedness in immunity networks. The colours of the interactions depict the number of matches between viruses and hosts. The network has a weighted-nested structure that builds up during the VDR and declines during the HCR, as viruses go extinct. The extinction is orderly, with viruses to which many hosts have immunity via many spacers (those at the left) going extinct first. **b**, Modularity in the infection network. The coloured interactions fall inside modules of virus strains that infect similar hosts (each module has a different colour). Black interactions are those that fall outside all modules. The size of the network and the number of modules increase during the VDR (between the purple and green vertical lines) and decline during the HCR (right of the green vertical line).

As a result, the effective mutation rate of the overall remaining viral population also increases (Supplementary Methods). An effective mutation is one leading to the growth of the resulting new strain (even if only temporarily and with no transition to the VDR); such growth can only occur in the HCR from a virus escaping the immunity of a host protected by a single match. As the proportion of single matches increases, the probability rises that a mutation occurs in a protospacer of a virus towards which a host or multiple hosts have no additional protection. Viral escapes are indeed associated with a particular tripartite structure where hosts are immune to viruses via a single match (Extended Data Fig. 7), which provides the basis for computing the expected reproductive number of a mutant virus, as we describe next.

Epidemiologically, virus escape can be predicted with the expected reproductive number of a mutant. The ability of a mutant virus to escape can be captured via a modified measure of the basic reproductive number—a well-known quantity in epidemiology measuring the ability of a pathogen to invade a host population (Methods and Supplementary Methods). We specifically derived an expression for the expected reproductive number of a mutant virus that we called R_{mut}, which quantifies the expected number of offspring a mutant virus would produce over its typical lifetime. When R_{mut} is above 1, a mutant virus can on average grow sufficiently fast to generate a local epidemic. Fig. 4c shows that this quantity fluctuates but remains largely above 1 during the VDR, with crossings of this value accompanying the different waves of virus diversification. R_{mut} drops considerably below 1 when the system enters an HCR. Its subsequent increase and the crossing of the threshold of 1 at some later point during this phase, signals the impending escape from host control preceding a transition to a new VDR (Fig. 4c). Moreover, in the final HCR, the one leading to viral extinction, this crossing never happens, indicating that mutant viruses cannot successfully escape control. To understand the role played in a successful escape by the increasing fraction of 0 matches, and especially by that of 1 matches, one needs to consider the respective expressions of the two components of R_{mut} for each virus (Supplementary Methods and Extended Data Fig. 8). In particular, the component related to 1 matches shows an increasing trend, reflecting a growing potential for mutants that escape such non-redundant immunity to gain access to sufficient additional hosts to sustain higher and ultimately positive growth (Extended Data Fig. 8).

Empirical immunity networks are weighted-nested. Finally, we analysed the structure of immunity in three empirical systems because the weighted nestedness of the immunity network is a key structural feature in the shifts between dynamical regimes. Ideally, one would consider temporal data on host–virus CRISPR matches. Such data are currently unavailable since obtaining them requires hundreds of genotypes of virus and CRISPR alleles to be resolved. Metagenomic datasets[21] can obtain such diversity but do not link spacers to individual host strains. Hence, currently available temporal data are either resolved at the strain level but for very low numbers of strains[4,22] or contain many spacers and protospacers but without host–virus CRISPR matches between multiple strains[21]. In fact, even for non-temporal data, there are very few examples of virus and host populations sufficiently resolved at an individual strain level. We considered three such static datasets and asked whether they exhibited evidence for the weighted-nested immunity structure predicted by the theory.

In our empirical datasets, both CRISPR alleles and virus strains were carefully and manually assembled and we had previously established their profiles of diversity[24,27,28] but not their structure. The data represent lytic, chronic and temperate virus lifestyles and two dif-

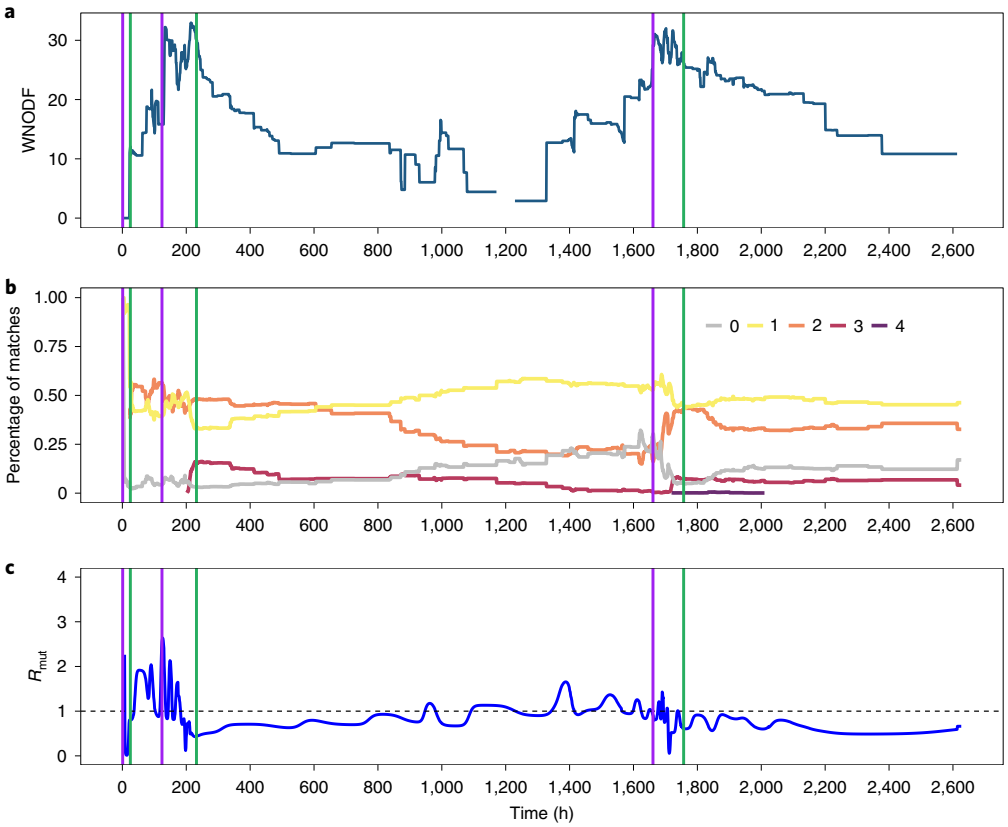


Fig. 4 | Processes leading to regime shifts. a, Weighted nestedness increases during VDRs and decreases during HCRs. **b**, The disassembly of the network as the result of virus extinctions during HCRs increases the proportion of 0 and 1 matches (calculated over all the interactions in the immunity network, regardless of virus or host), particularly towards the end of these periods. This creates the conditions for escape as a higher proportion of 1 matches makes it increasingly likely that a mutation will result in an escape. **c**, At the entrance into an HCR, the expected reproductive number of a mutant virus, R_{mut}, over all strains is below 1 indicating its inability to generate positive growth. During the central HCR and concomitant with the trend in the fraction of 1 matches, R_{mut} exhibits an increasing trend and eventually crosses the threshold value of 1. This signals that a mutant can leave enough progeny to replace itself and initiate a local epidemic. The crossing of $R_{mut}=1$ anticipates the impending virus escape from host control and the resulting transition to another VDR. Because mutations are stochastic events, the timing of such escape is not at this crossing. During the VDR, R_{mut} oscillates rapidly but remains largely above 1, with crossings of this value accompanying the successive waves of virus diversification. See Extended Data Fig. 8 and Supplementary Methods for details on the components of R_{mut} and their relationship to 1 matches.

ferent microorganisms from the two domains of life where CRISPR occurs (archaea and bacteria). The first two datasets compared genomes resolved to individual strains of the thermoacidophilic crenarchaeon *Sulfolobus islandicus* sampled from two different locations with two different populations of contemporary viruses. The third dataset is a pool of the gammaproteobacteria *P. aeruginosa*[28], isolated from the sputum of patients with cystic fibrosis from a single clinic in Copenhagen by Marvig et al.[29], with virus genomes obtained from the temperate mu-like viruses obtained from the National Center for Biotechnology Information (NCBI). We and others have shown previously that CRISPR loci do not evolve over time within a single patient. Therefore, this dataset represents a snapshot of the *P. aeruginosa* diversity present in the large panmictic population of strains that must evolve diversity in CRISPRs in unknown environmental reservoirs[30]. Viruses are not contemporary but represent the diversity of cultured strains from a single virus type.

Because we did not have a time series of matches, our analysis cannot directly address transitions between HCR and VDR regimes. The data allow us to interrogate the structure of immunity about the existence of weighted nestedness in these populations. From each dataset, the empirical immunity matrices were constructed by comparing CRISPR arrays from each individual host strain to virus genomes. We assessed the statistical significance of nestedness for each empirical network by comparing the observed value of its nestedness index (for two different indices; Methods) to a distribution obtained from 10,000 shuffled networks. We find that the probability that a shuffled network will be more nested than the observed one is effectively zero ($P<0.0001$; Methods) in all three empirical networks (Fig. 5).

In light of our theoretical results, these findings are consistent with host control of viruses via distributed immunity that is redundant in the number of matches. Moreover, we found that, as in our theoretical results, the empirical host–spacer networks in the three VDRs are also significantly modular (Extended Data Fig. 9), with only one out of five modules showing a phylogenetic signal (Supplementary Table 3). This suggests that the mechanism by which immunity is obtained is similar to the one we have described: modules delineate groups of host strains that share immunity via the same spacers and are therefore susceptible to similar viruses. While these results point to the possibility that dynamic processes similar to what we observe theoretically may play a role in nature, this hypothesis can only be confirmed via empirical studies that include the temporal dimension, providing a direction for future research.

Discussion

The role of resistance and immunity in driving the diversity, evolution and coevolution of microbes and viruses has been extensively studied[2,4,21,31–35]. However, the role it plays in structuring microbial

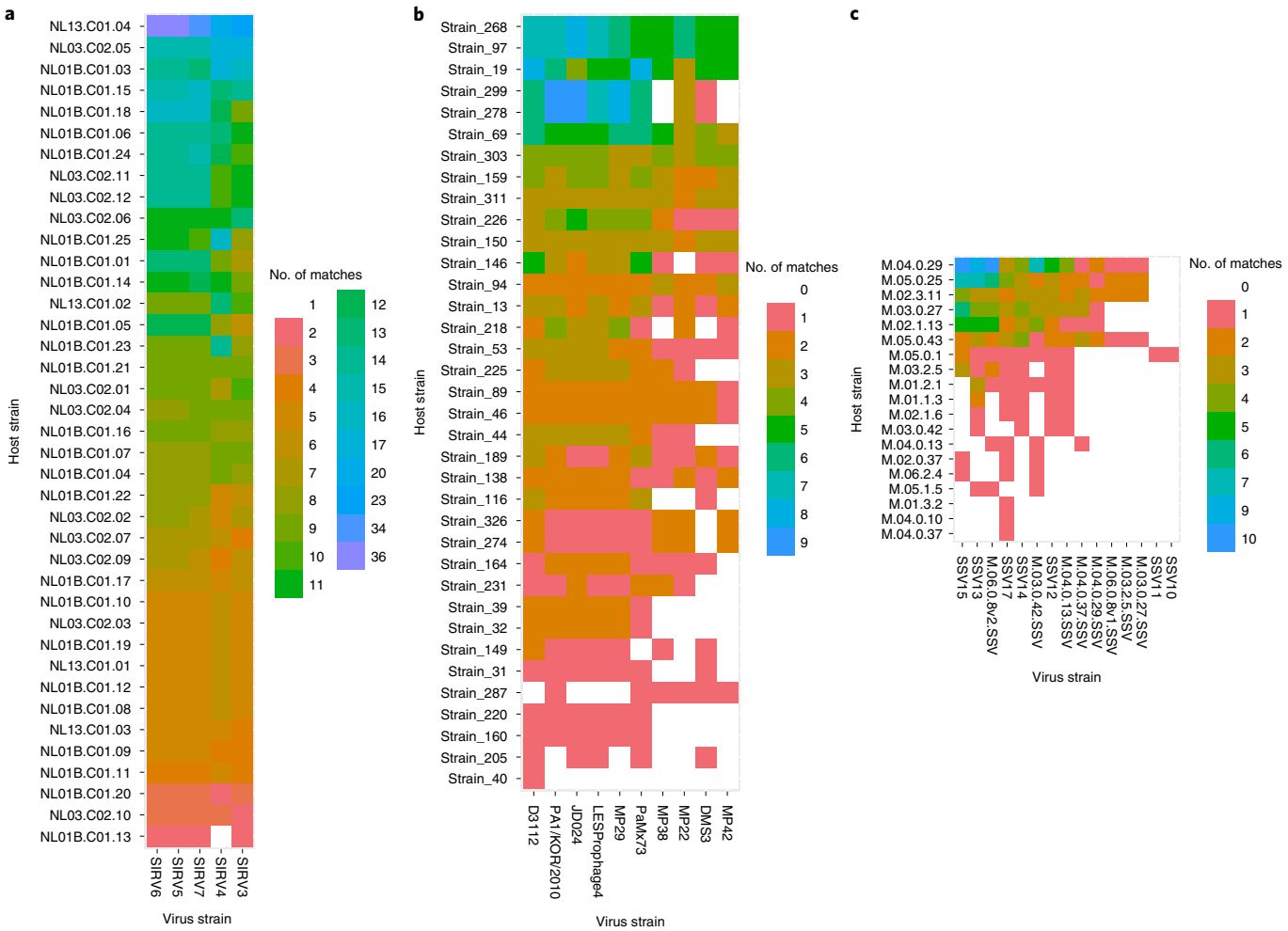


Fig. 5 | Weighted nestedness of empirical immunity networks. The matrices depict the number of shared spacers and protospacers between hosts (rows) and viruses (columns). Each network is ordered by column and row sums and is nested in the quantity of matrix cells. The networks come from three different systems. **a**, *S. islandicus* hosts from a single location in Yellowstone National Park and contemporary lytic *S. islandicus* rod-shaped viruses isolated from Yellowstone National Park. **b**, *S. islandicus* hosts and contemporary chronic *Sulfolobus* spindle-shaped viruses from the Mutnovsky Volcano in Russia (2010). **c**, *P. aeruginosa* hosts from Copenhagen and temperate mu-like viruses. To test the hypothesis that for a given distribution of matches in the population of host strains, the observed network is organized in a non-random, weighted-nested pattern, we shuffled networks by randomly distributing the interactions.

populations remains poorly investigated despite possible consequences for host–virus coevolution and population stability. Our theoretical results show that the coevolutionary dynamics of a host–pathogen system with CRISPR-induced resistance can influence, and in turn be influenced by, the network structure of strain diversity. In particular, modularity of the infection network and weighted nestedness of the immunity network interact to affect the transient nature of alternating dynamical regimes. By promoting pathogen diversification, the former builds the latter. In turn, the redundant protection of the immunity network and its nested organization put an end temporarily to pathogen diversification by enabling host control of virus growth and opposing escape of such control via effective virus mutation. Weighted nestedness contains the seeds of its own unravelling by implying a certain order in viral extinction, which eliminates preferentially viruses with the most redundant protection, and in so doing, creates the conditions for a potential new viral escape. Thus, the weighted nestedness of immunity networks is at the centre of the coevolutionary dynamics and the transient dynamical regimes we have described in this article.

Modularity has been described for host–parasite infection networks and attributed to a variety of mechanisms, such as phylogeny, specificity and coevolution[9,14,36]. It was also detected in a host–virus system, where it was suggested to arise via negative frequency-dependent selection[16]. Its emergence from immune selection was demonstrated in this study. It is consistent with existing strain theory developed mostly for human infections and pathogens with multilocus encoding of antigens when evolution is considered explicitly[17,37–40]. Strain theory posits that immune selection acting as a form of negative frequency-dependent selection, can create groups of pathogens with limiting overlap in antigenic space. Pathogens with antigens that are new to the host immune system are at a competitive advantage, while those with common ones are at a disadvantage. In the absence of data on the immunity of hosts, this pattern has been described as modules, or clusters, in networks that depict genetic similarity between the pathogens themselves[17,18]. In contrast to strain theory, we are able to consider in this study the immune genotype of pathogens and hosts simultaneously, which allows us to move beyond pathogen genetic similarity to consider actual infection patterns. In this study, niche structure arises in part from the addition of new spacers along with competition among viruses for hosts.

One major difference with previous strain theory is that modularity and associated coexistence of pathogens relying on different

groups of hosts in our model are only transient and not accompanied by stable population dynamics. Strain theory for hosts with non-heritable immunity, where fully susceptible individuals are continuously added to the population via births or immigration, allows instead for stationary coexistence of pathogens with limited similarity depending on the strength of cross-immunity and the speed of evolution[37,40]. Future work should examine whether this difference and the associated shifts in dynamical regimes arise from the heritable nature of CRISPR-based immunity, which by definition cannot produce fully susceptible offspring, needed to sustain transmission. Host–virus coexistence can also result from loss of immunity[34], a feature absent from our model and that could prolong VDRs by supplying a constant inflow of susceptible hosts.

We show that, when multiple viruses are involved, weighted nestedness with multiple spacers derived from repeated infections provides a population-level mechanism by which hosts can control viral populations. It remains an open question whether weighted nestedness can arise under non-CRISPR immunity or non-heritable immunity. For CRISPR-induced immunity, nestedness reflects the coevolutionary diversification of hosts in response to that of viruses, and more specifically, the build-up over time of the redundancy in immune memory, which allows hosts to keep viruses in check despite their occasional escape. Thus, observation of weighted nestedness in nature could be an indication of such coevolutionary history and resulting control. Variation of this network property over time contains valuable information on the relative pace of viral diversification and host acquisition of immunity. Observing the temporal course of nestedness in both dedicated experiments and in natural environments should shed light on its role for controlling virus populations in relation to host and virus diversification. Quantitative extensions should also investigate appropriate ways to normalize weighted nestedness (Song et al.[41]), so that the effects of its change can be isolated for network size and distributed redundancy of protection.

The transience of alternating dynamical regimes in the CRISPR system is not an impediment to the formation of rich structures by coevolution; in this study, it is its natural consequence. It should be recognized, however, that the alternation of dynamical regimes occurs in the model in a defined region of parameter space that allows for high diversification and is absent in some earlier simulations (for example, Childs et al.[25]). Because host control periods lead to stochastic viral extinction, identifying the conditions in terms of critical parameter combinations that give rise to these dynamics is of practical importance.

Future work in this area should consider the role of demographic stochasticity arising from small population numbers and discrete individuals. Our numerical implementation of the system treated population dynamics as deterministic and therefore relied on a given abundance threshold for extinction. It also initialized the size of a new strain at a given abundance slightly above threshold. We expect our results to hold qualitatively even though there should be some quantitative differences in the exact parameter values for which the two implementations are comparable. This should be the case for the mutation rate since the emergence probability of new strains should decrease due to demographic stochasticity. For a fully stochastic implementation, the analytical expressions of Chabas et al.[42] could be extended to compute and understand emergence probability under varying frequencies of resistance alleles. A future theory should also more broadly investigate the dynamics structure nexus over parameter space, in particular for protospacer mutation rates and spacer acquisition probabilities, which set the pace of evolution and have already been recognized as key to the outcome of viral extinction[22,25,43]. Finally, anti-CRISPR mechanisms maintained by viruses may also affect network structure.

The demonstration of significantly weighted nestedness in the empirical networks is consistent with coevolution involving CRISPR-induced immunity being at play in natural microbial populations. We present this empirical evidence in support of our theoretical results, in the same way that system-level statistical properties of ecological communities have been and are being studied as reflecting underlying assembly processes. We note that the *Sulfolobus* data are carefully positioned as local and contemporary and therefore are consistent with ongoing population dynamics. In contrast, the assemblage of *P. aeruginosa* strains (one per patient with cystic fibrosis) are interpreted to represent a snapshot of the ongoing dynamics that occur outside of the human lung, which is consistent with literature on *P. aeruginosa* population structure[28,30]. Despite the importance of *P. aeruginosa*, very little is known about where and on what timescales virus–host interactions (mediated by CRISPR) occur outside of patients with cystic fibrosis. Therefore, the alignment with our theory may have alternative explanations, although we are not aware of such alternatives at this stage. Obtaining time series data on CRISPR-mediated interactions between multiple host and virus strains should be a priority to test our theoretical results.

Ultimately, knowledge of the dynamics and stability of diverse host and viral populations can be applied to the control of microbes in food and industrial sciences, infectious disease emergence and treatment with phage therapy, microbiome dynamics, agriculture and environmental engineering.

Methods

The model. The model implements the formulation by Childs et al.[20], which combines three main components: ecological population dynamics; stochastic coevolution generating diversity; and molecular identity of hosts and viruses defining CRISPR immunity. Diversification events (spacer acquisition by the host and protospacer mutation of viruses) are modelled stochastically, whereas population dynamics of different subpopulations (that is, strains) of hosts and viruses are represented deterministically with Lotka–Volterra differential equations (equations (1) and (2)).

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = rN_i\left(1-\frac{\sum_i N_i}{K}\right) - \left[(1-q)\sum_j (1-M_{ij})V_j + p\sum_j M_{ij}V_j\right]\phi N_i \quad (1)$$

$$\frac{\mathrm{d}V_j}{\mathrm{d}t} = \beta\phi\left[(1-q)\sum_i (1-M_{ij})N_i + p\sum_i M_{ij}N_i\right]V_j - \left(\phi\sum_i N_i + m\right)V_j \quad (2)$$

Each strain of host i and virus j is defined by a unique genomic state of their spacer and protospacer sets S_i and G_j, respectively. In the ecological population dynamics, each host strain i has abundance N_i (the carrying capacity of all strains is K) and reproduces at a per capita rate of r. Each viral strain j has abundance V_j, which increases due to infection and lysis of hosts and decays at a density-independent rate m. Extinction of any host or viral strain occurs when these fall below a critical threshold ρ_c. Viruses infect at a constant 'adsorption rate' ϕ either hosts that do not have protection or those whose protection fails (see below).

Host immunity to a virus is defined in the molecular component of the model and is based on genomic sequence matches between the spacer and protospacer sets. Specifically, CRISPR immunity is defined using the function $M_{ij}=M(S_i, G_j)$, which equals 1 if there is at least 1 match between the sets ($|S_i \cap G_j| \geq 1$) or 0 otherwise. The CRISPR immune mechanism is not perfect and can fail. When $M_{ij}=1$, there is a probability p that the host strain is lysed and correspondingly, $1-p$ that it survives and the virus is eliminated. On the other hand, when $M_{ij}=0$, there is a probability q that the virus strain is eliminated, resulting in the acquisition of a protospacer by the host, and $1-q$ that it is lysed by the virus. Both p and q are small ($p, q \ll 1$).

The above Lotka–Volterra dynamics are implemented in between any two stochastic events concerning evolution. Specifically, errors in viral replication can result after successful infection of a host, leading to the replacement of a random protospacer with mutation rate μ (per protospacer per viral replication). This incorporates a new viral strain into the system with an initial low abundance of 1.1 times the extinction threshold ρ_c. Viruses cannot mutate multiple protospacers at the same time. During an unsuccessful infection attempt by a virus (regardless of host immunity), there is also a probability q of acquiring a new spacer by incorporating a protospacer and integrating it into the host's CRISPR system at its leading end. Hosts cannot acquire multiple spacers at the same time. The maximum number of spacers per host's strain is constant. If the maximum number of acquired spacers is reached, the addition of a spacer to the leading end

is accompanied by the deletion of a spacer at the trailing end. In our simulations, the length of the spacer cassette was set to a sufficiently large value to avoid loss of acquired immune memory.

We numerically implemented the model in C++ to increase computational efficiency. This enables consideration of a larger number of spacers/protospacers and hence host and virus richness, for longer simulation times, than in the original MATLAB code[20]. Our implementation combines: (1) a Gillespie algorithm to determine the time between two stochastic events and randomly select a virus/host strain to mutate a protospacer/acquire a new spacer; and (2) a numerical ordinary differential equation solver using Euler's method. The code includes the following features to facilitate subsequent network (and other) analyses: (1) all data related to virus or host strains (for example, identity of protospacers/spacers, abundance values) at each time are written to files during simulation; (2) the parent ID of each newly generated virus or host strain is tracked to generate phylogenetic trees; (3) checkpoint implementation for running longer simulations. Details on the implementation are in the Supplementary Information.

We used the parameters summarized in Supplementary Table 4 based on Childs et al.[20,25]. The host growth rate corresponds to a doubling time of 1 h within the range of observed values for *P. aeruginosa* (20 min) and *Sulfolobus* (8 h). The adsorption rate also falls within typical range values of 10^{-8} to 10^{-9} ml min^{-1}. The order of magnitude of our mutation rate is consistent with standard mutation rates for DNA-based organisms, if we assume that mutation at a single site enables protospacer escape from a spacer match. The burst size of 50 is on the low side with values of 200 observed in *P. aeruginosa*. This is motivated by a lower protospacer number than currently documented in nature. Specifically, the speed of virus escape ultimately depends on the effective rate of mutation per protospacer, which increases linearly with the product of burst size and the mutation rate parameter μ and decreases inversely with the number of protospacers. Therefore, our parameter set should generate similar speeds of protospacer evolution than for higher burst sizes and also higher numbers of protospacers (for example, $\beta = 200$ and $g_p \times 4$). Our numerical implementation allows consideration of a higher number of protospacers than before[20]; for simplicity, in our study we used a number of protospacers that is still below those typically observed but have compensated with a lower value of offspring produced (burst size; Supplementary Information).

Definition of regimes. Our general approach to define regimes was to classify each point in the virus abundance time series to either an HCR or a VDR. We did so by detecting changes in relative virus abundance, defined as the total virus abundance at any given time t, $V_T(t)$, divided by the maximum abundance in the whole time series, V_T: $A(t) = V_T(t) / \max(V_T)$. Using relative abundance allows for comparisons and analysis across multiple simulations; considering absolute abundance does not change the regime definition (Extended Data Fig. 10). An HCR is a sequence of points where $V_T(t)$ changes very little, which can be captured by calculating the changes between consecutive time points. We consider the first $A' = A(t_x) - A(t_{x-1})$ and second difference $A'' = A'(t_x) - A'(t_{x-1})$. Values close to 0 in A'' would occur when (1) there is no or low change in A between two consecutive time intervals so that A' itself is small in each of these periods or (2) A' varies and changes sign in these consecutive intervals but the size of these changes 'up' and 'down' are small, as well as the peak in A this reflects. (In this study, we ignore the case of an almost linear increase, which does not apply.) We classified each point as belonging to the HCR if the absolute value of A'' is smaller than a given small threshold, so that $|A''_t| < 0.001$, and to a VDR otherwise. This procedure creates a sequence of points $C(t)$ with each point classified into an HCR or VDR (Extended Data Fig. 10). We added two further conditions. The first includes time points in the HCR when isolated increases in virus abundance are not recognized as sufficiently small peaks by our threshold, but these increases are localized outbreaks in the sense of not resulting in a complete escape and a regime switch. To include such events within the HCR, we calculated a threshold f defined as the 75% quantile value of the distribution of VDR lengths in C. Any sequence of VDR points shorter than f was converted to the HCR. The last condition for classification to an HCR was that a sequence of points had to be longer than the longest-lasting virus outbreak.

While the values for the thresholds we have used may seem, to some degree, arbitrary, this would be the case for any other algorithm for classification because a 'regime' is not naturally defined by the dynamical system, but rather detected in the emerging time series. As an independent corroboration for our method, we imposed the regime boundaries calculated using the virus abundance time series to the time series of host abundance and virus diversification. The regimes are evident in those time series also (Extended Data Figs. 1–3). For example, in the virus diversification time series, the rate of diversification is much steeper in the VDR than in the HCR (Extended Data Fig. 3). This additional verification increases the confidence we have in our ability to classify these regimes.

Network construction. The networks we use are bipartite networks, which contain two sets of nodes (for example, hosts and spacers or hosts and viruses). Bipartite networks are mathematically represented using incidence matrices. A graphical overview of the networks (and associated matrices) and how we constructed them can be found in Fig. 2. In this section, we provide details on network construction.

Networks of genetic composition. At each time t, we defined a host–spacer network, $\mathcal{S}(t)$ (and associated matrix $S_{xi}(t)$), as a bipartite network where each edge is drawn between a host strain i and a spacer x (Fig. 2a,e). We analogously defined a virus–protospacer network, $\mathcal{P}(t)$ (and associated matrix $P_{yj}(t)$) as a bipartite network where each edge is drawn between a virus strain j and a protospacer y (Fig. 2b,f). Hence, in each network a host or virus strain's genome is the set of its neighbouring spacer or protospacer nodes, respectively.

Immunity network. We used $\mathcal{S}(t)$ and $\mathcal{P}(t)$ to define an immunity network at a time t, $\mathcal{I}(t)$ (and associated matrix $I_{ij}(t)$), where edges are drawn between a virus strain j and a host strain i. Edge weights were defined as the number of matching spacers and protospacers between $\mathcal{S}(t)$ and $\mathcal{P}(t)$ for any given host–virus pair (Fig. 2c,g). That is, $I_{ij}(t) = \sum_x^{S_i} \sum_y^{G_j} M(S_i^x, G_j^y)$, where M has a value of 1 if spacer x in the spacer set of host i, S_i, is the same as protospacer y in the protospacer set of virus j, G_j and 0 otherwise.

Infection network. We defined an infection network at a time t, $\mathcal{G}(t)$ (and associated matrix $G_{ij}(t)$), as a subset of the immunity network where $I_{ij}(t) = 0$ (Fig. 2d,h). That is, the edges in $G_{ij}(t)$ were drawn between hosts and viruses that did not have an edge in $\mathcal{I}$. We weighted the edges of $G_{ij}(t)$ by a normalized measure of encounter between virus strain j and a host i, based on their abundance values: $G_{ij}(t) = \frac{V_j(t) N_i(t)}{N_T(t)}$, where $V_j(t)$ and $N_i(t)$ are the abundances of a virus strain j and a host strain i at time t and N_T is the total abundance of hosts.

Network analysis. *Weighted nestedness.* We evaluated the nestedness of $I_{ij}(t)$ using weighted nestedness based on overlap and decreasing fill (WNODF)[26]. Briefly, the index ranges from 0 to 100, with the maximum 100 representing perfect nestedness. Perfect nestedness occurs when all 2 × 2 sub-matrices of the form:

$$\begin{array}{c} \\ b_1 \\ b_2 \end{array} \begin{array}{c} \begin{array}{cc} v_1 & v_2 \end{array} \\ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \end{array}$$

satisfy the conditions for host b_1 to be immune to the two viruses via more matches than b_2 ($a > c$, $b > d$, $a > d$) and for v_2 to have fewer matches to the two hosts than v_1 ($a > b$, $c > d$). We calculated the WNODF with the networklevel function in the bipartite package v.2.11 in R. We calculated the WNODF when at least two strains of both hosts and viruses were present.

In the empirical data, we evaluated nestedness in two ways. First, we used the WNODF. Second, we calculated the largest eigenvalue of $I_{ij}(t) I_{ij}^{T}(t)$, ρ, as suggested by Staniczenko et al.[44]. We did not use ρ in the simulations because preliminary analyses indicated that this measure is highly dependent on network size and therefore cannot be used to compare between networks. However, it is suitable for comparing an observed network to its shuffled counterparts because for a given distribution of weights and network size, high values of ρ indicate a more quantitatively nested structure[44]. WNODF and ρ are the only two measures for quantitative nestedness we are aware of. WNODF cannot handle networks that are fully connected (that is, density of 1); therefore, we did not use it for our Yellowstone National Park dataset.

Community detection. To find 'communities', or as commonly referred to, 'modules', we used the map equation objective function to calculate the optimal partition of the network[45,46] with the R package infomapecology v.0.1.2 (ref. [47]). Briefly, the map equation is a flow-based and information-theoretic method (implemented with Infomap), which calculates network partitioning based on the movement of a random walker on the network. In any given partition of the network, the random walker moves across nodes in proportion to the direction and weight of the edges. Hence, it will tend to stay longer in dense areas representing groups of, for example, viruses and hosts with high interaction density. These areas can be defined as 'modules'. The time spent in each module can be converted to an information-theoretic currency using an objective function called the map equation and the 'best' network partition is the one that minimizes the Map Equation[45,46]. For convenience, we use the term 'modules' because it is commonly used to refer to partitions of networks across different disciplines, but Infomap does not calculate a modularity function (Newman and Girvan[48]). We chose the map equation over the more commonly used modularity objective function because of the computational efficiency of its implementation (Infomap)[49], given that we needed to analyse hundreds of thousands of networks.

Significance of modularity and weighted nestedness. To test if a given network structure is non-random, we evaluated the statistical significance of the structural index (map equation value for modularity and WNODF or ρ for weighted nestedness) by comparing it to a distribution of the index for shuffled versions of the matrix. We shuffled binary matrices (host–spacer and infection networks) with function r00 in the package vegan v.2.5-4 in R, an algorithm that maintains the density of the network. This was done using infomapecology[47]. We shuffled weighted matrices (immunity networks in simulated and empirical data) by randomly distributing the interactions (function r00_samp in the package vegan in R). This algorithm maintains the density of the network and the distribution of

weights while shuffling the structure. Therefore, it tests the hypothesis that for a given distribution of matches, the observed network is non-randomly structured in a quantitatively nested way.

Phylogenetic signal in modules. We first computed the pairwise distances between pairs of strains within each module from branch lengths, with the function cophenetic.phylo in the package ape in R. Then, we permuted the identity of strains in modules, maintaining module size, and recalculated the mean pairwise phylogenetic distance. The null hypothesis is that the permuted distance is smaller than the observed one (per module) and therefore there is no phylogenetic signal. Rejecting this hypothesis indicates a phylogenetic signal because the observed phylogenetic distance between hosts within each module would be smaller than expected by chance (closely related hosts share a module). Because each network has several modules, the threshold for significance was adjusted by dividing 0.05 by the number of modules (Bonferroni correction).

Basic reproductive numbers. To derive the expected reproductive number of a mutant R_{mut}, we need to first define a set of different reproductive numbers that provide the building blocks for this quantity (Supplementary Methods). The first measure, $R_0^j(t)$, is the number of offspring produced at time t by a virus strain j from infecting all hosts with no protection to it (0 matches) and is defined as:

$$R_0^j(t) = \frac{\beta\phi(1-q)}{\phi N_T(t) + m} \sum_i M_{ij}^0 N_i(t) \quad (3)$$

where growth from failure of immunity has been neglected. The function M_{ij}^0 equals 1 (and 0 otherwise) when there are no matches between the set of spacers of host i and the set of protospacers of virus j. The sum over i is over all host strains at time t.

The second measure, R_1^j, quantifies the expected growth allowed by the additional infections made possible by mutation of one of the protospacers of virus j whose match to the spacer of a given host represents its only protection to the virus:

$$R_1^j(t) = \frac{\beta\phi(1-q)}{\phi N_T(t) + m} \times \frac{1}{k_j} \sum_i M_{ij}^1 N_i(t) \quad (4)$$

with $M_{ij}^1 = 1$ (and 0 otherwise) if there is exactly 1 matching spacer–protospacer pair between the set of spacers of host i and the set of protospacers of virus j. The number k_j counts the protospacers of virus j involved in these pairs. (This can be visualized as the degree of the node representing the virus in Extended Data Fig. 7.)

Together, these two measures constitute the expected number of offspring that an escape mutant of virus strain j would produce over its typical lifespan. We call this number the 'potential reproductive number' of virus j and define it as:

$$R_{pot}^j(t) = R_0^j + R_1^j \quad (5)$$

We now have all the pieces to write the expectation that defines R_{mut} as:

$$R_{mut} = \sum_j P_{mut}^j \left[(1 - \frac{k_j}{g}) R_0^j + \frac{k_j}{g} R_{pot}^j \right] \quad (6)$$

where the probability P_{mut}^j that virus j receives the mutation is given by:

$$P_{mut}^j = \frac{g\mu_j}{\sum_l g\mu_l} \approx \frac{\sum_i N_i V_j (1 - M_{ij})}{\sum_i \sum_l N_i V_l (1 - M_{i,l})} \quad (7)$$

We note that in addition to this probability, we need to consider in the above expectation that defines R_{mut} the probability that a mutation results in an escape, k_j/g, or not $(1 - k_j/g)$. Only mutation of one of the k_j key protospacers will open up additional hosts for infection and population growth. If a mutation falls on one of these protospacers, the number of offspring is given by R_{pot}^j; if it does not, the number of offspring is simply that of the original parent, namely R_0^j, which explains the expression for R_{mut}.

Empirical data. The first dataset represents a single time point from three adjacent hot springs (NL01, NL10 and NL13) in the Nymph Lake region of the Yellowstone National Park. These three springs have been shown to share a single well-mixed population of *S. islandicus* hosts[50]. Viruses were collected from contemporary populations and CRISPR spacers and virus isolates suggest that these are dominated by the lytic virus *S. islandicus* rod-shaped virus[51]. The second dataset is from *S. islandicus* strains isolated from several springs in the Mutnovsky Volcano in Kamchatka (Russia) that were also shown to share populations of hosts. The predominant viral population in these hosts is the chronic *Sulfolobus* spindle-shaped virus[24]. The third dataset consists of longitudinal sampling of human-adapted *P. aeruginosa* isolates from sputum samples of patients with cystic fibrosis collected at a hospital in Copenhagen and a global set of temperate mu-like viruses from *P. aeruginosa* viruses extracted from the NCBI to substitute for a lack of sequenced contemporary viruses[28,29]. Detailed methods are shown in the Supplementary Methods.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Simulated and empirical network data are available in the dedicated GitHub repository associated with this paper at: https://github.com/Ecological-Complexity-Lab/CRISPR_networks.

Code availability

The code for the simulations and their analysis is available in the dedicated GitHub repository associated with this paper at: https://github.com/Ecological-Complexity-Lab/CRISPR_networks.

Received: 2 February 2020; Accepted: 14 August 2020;
Published online: 19 October 2020

References

1. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
2. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
3. Mayer, A., Mora, T., Rivoire, O. & Walczak, A. M. Diversity of immune strategies explained by adaptation to pathogen statistics. *Proc. Natl Acad. Sci. USA* **113**, 8630–8635 (2016).
4. Chevallereau, A., Meaden, S., van Houte, S., Westra, E. R. & Rollie, C. The effect of bacterial mutation rate on the evolution of CRISPR–Cas adaptive immunity. *Philos. Trans. R. Soc. Lond. B* **374**, 20180094 (2019).
5. Gurney, J., Pleška, M. & Levin, B. R. Why put up with immunity when there is resistance: an excursion into the population and evolutionary dynamics of restriction–modification and CRISPR–Cas. *Philos. Trans. R. Soc. Lond. B* **374**, 20180096 (2019).
6. Weitz, J. S. et al. Phage–bacteria infection networks. *Trends Microbiol.* **21**, 82–91 (2013).
7. Gurney, J. et al. Network structure and local adaptation in co-evolving bacteria-phage interactions. *Mol. Ecol.* **26**, 1764–1777 (2017).
8. Fortuna, M. A. et al. Coevolutionary dynamics shape the structure of bacteria-phage infection networks. *Evolution* **73**, 1001–1011 (2019).
9. Krasnov, B. R. et al. Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. *Am. Nat.* **179**, 501–511 (2012).
10. Pilosof, S. et al. Host–parasite network structure is associated with community-level immunogenetic diversity. *Nat. Commun.* **5**, 5172 (2014).
11. Dallas, T. & Cornelius, E. Co-extinction in a host-parasite network: identifying key hosts for network stability. *Sci. Rep.* **5**, 13185 (2015).
12. Pilosof, S., Morand, S., Krasnov, B. R. & Nunn, C. L. Potential parasite transmission in multi-host networks based on parasite sharing. *PLoS ONE* **10**, e0117909 (2015).
13. Vázquez, D. P., Poulin, R., Krasnov, B. R. & Shenbrot, G. I. Species abundance and the distribution of specialization in host–parasite interaction networks. *J. Anim. Ecol.* **74**, 946–955 (2005).
14. Fortuna, M. A. et al. Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* **79**, 811–817 (2010).
15. Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host–phage interactions. *Proc. Natl Acad. Sci. USA* **108**, E288–E297 (2011).
16. Beckett, S. J. & Williams, H. T. P. Coevolutionary diversification creates nested-modular structure in phage–bacteria interaction networks. *Interface Focus* **3**, 20130033 (2013).
17. He, Q. et al. Networks of genetic similarity reveal non-neutral processes shape strain structure in *Plasmodium falciparum*. *Nat. Commun.* **9**, 1817 (2018).
18. Pilosof, S. et al. Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in plasmodium falciparum. *PLoS Biol.* **17**, e3000336 (2019).
19. van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **12**, 479–492 (2014).
20. Childs, L. M., Held, N. L., Young, M. J., Whitaker, R. J. & Weitz, J. S. Multiscale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. *Evolution* **66**, 2015–2029 (2012).
21. Paez-Espino, D. et al. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *mBio* **6**, e00262-15 (2015).
22. van Houte, S. et al. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–388 (2016).
23. Daly, R. A. et al. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat. Microbiol.* **4**, 352–361 (2019).

24. Pauly, M. D., Bautista, M. A., Black, J. A. & Whitaker, R. J. Diversified local CRISPR–Cas immunity to viruses of *Sulfolobus islandicus*. *Philos. Trans. R. Soc. Lond. B* **374**, 20180093 (2019).
25. Childs, L. M., England, W. E., Young, M. J., Weitz, J. S. & Whitaker, R. J. CRISPR-induced distributed immunity in microbial populations. *PLoS ONE* **9**, e101710 (2014).
26. Almeida-Neto, M. & Ulrich, W. A straightforward computational approach for measuring nestedness using quantitative matrices. *Environ. Model. Softw.* **26**, 173–178 (2011).
27. Held, N. L., Herrera, A., Cadillo-Quiroz, H. & Whitaker, R. J. CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS ONE* **5**, e12988 (2010).
28. England, W. E., Kim, T. & Whitaker, R. J. Metapopulation structure of CRISPR–Cas immunity in *Pseudomonas aeruginosa* and its viruses. *mSystems* **3**, e00075-18 (2018).
29. Marvig, R. L., Sommer, L. M., Molin, S. & Johansen, H. K. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat. Genet.* **47**, 57–64 (2015).
30. Pirnay, J.-P. et al. *Pseudomonas aeruginosa* population structure revisited. *PLoS ONE* **4**, e7740 (2009).
31. Hall, A. R., Scanlan, P. D., Morgan, A. D. & Buckling, A. Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecol. Lett.* **14**, 635–642 (2011).
32. Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genet.* **9**, e1003312 (2013).
33. Koskella, B. & Brockhurst, M. A. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–931 (2014).
34. Weissman, J. L. et al. Immune loss as a driver of coexistence during host-phage coevolution. *ISME J.* **12**, 585–597 (2018).
35. Braga, L. P. P., Soucy, S. M., Amgarten, D. E., da Silva, A. M. & Setubal, J. C. Bacterial diversification in the light of the interactions with phages: the genetic symbionts and their role in ecological speciation. *Front. Ecol. Evol.* **6**, 6 (2018).
36. Fontaine, C. et al. The ecological and evolutionary implications of merging different types of networks. *Ecol. Lett.* **14**, 1170–1181 (2011).
37. Gupta, S. & Day, K. P. A strain theory of malaria transmission. *Parasitol. Today* **10**, 476–481 (1994).
38. Buckee, C. O., Recker, M., Watkins, E. R. & Gupta, S. Role of stochastic processes in maintaining discrete strain structure in antigenically diverse pathogen populations. *Proc. Natl Acad. Sci. USA* **108**, 15504–15509 (2011).
39. Artzy-Randrup, Y. et al. Population structuring of multi-copy, antigen-encoding genes in *Plasmodium falciparum*. *eLife* **1**, e00093 (2012).
40. Zinder, D., Bedford, T., Gupta, S. & Pascual, M. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLoS Pathog.* **9**, e1003104 (2013).
41. Song, C., Rohr, R. P. & Saavedra, S. Why are some plant–pollinator networks more nested than others? *J. Anim. Ecol.* **86**, 1417–1424 (2017).
42. Chabas, H. et al. Evolutionary emergence of infectious diseases in heterogeneous host populations. *PLoS Biol.* **16**, e2006738 (2018).
43. Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR–Cas in an explicit ecological context. *J. Bacteriol.* **195**, 3834–3844 (2013).
44. Staniczenko, P. P. A., Kopp, J. C. & Allesina, S. The ghost of nestedness in ecological networks. *Nat. Commun.* **4**, 1391 (2013).
45. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123 (2008).
46. Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *Eur. Phys. J. Spec. Top.* **178**, 13–23 (2010).
47. Farage, C., Edler, D., Eklöf, A., Rosvall, M. & Pilosof, S. A dynamical perspective to community detection in ecological networks. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.04.14.040519v1 (2020).
48. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
49. Lancichinetti, A. & Fortunato, S. Community detection algorithms: a comparative analysis. *Phys. Rev. E* **80**, 056117 (2009).
50. Campbell, K. M. et al. *Sulfolobus islandicus* meta-populations in Yellowstone National Park hot springs. *Environ. Microbiol.* **19**, 2334–2347 (2017).
51. Bautista, M. A., Black, J. A., Youngblut, N. D. & Whitaker, R. J. Differentiation and structure in *Sulfolobus islandicus* rod-shaped virus populations. *Viruses* **9**, 120 (2017).

Acknowledgements

We thank W. England and M. Pauly for data processing and collection and Q. He for guidance on the construction of the phylogenetic trees from the model outputs. R.W. acknowledges the support of the Cystic Fibrosis Foundation (no. CFF C2480) and an Allen Distinguished Investigator Award from the Allen Frontiers Institute. M.P. acknowledges the support of the University of Chicago. We are grateful for the access to the computer cluster of the Research Computing Center of the University of Chicago.

Author contributions

S.P. conceptualized the study, carried out the formal analysis (networks and data analysis, model dynamics, visualizations) and wrote the original manuscript draft. S.A.A.-C. carried out the formal analysis (networks analysis, model simulation) and reviewed and edited the manuscript. T.W. was involved with the software, methodology and formal analysis (model dynamics). T.K. curated and analysed the data and reviewed and edited the manuscript. S.M. was involved with the software, methodology and funding acquisition, and reviewed and edited the manuscript. R.W. conceptualized the study, curated and analysed the data, acquired the funding, supervised the study and reviewed and edited the manuscript. M.P. conceptualized the study, carried out the formal analysis (model dynamics), acquired the funding and prepared the original manuscript draft.

Competing interests

The authors declare no competing interests.

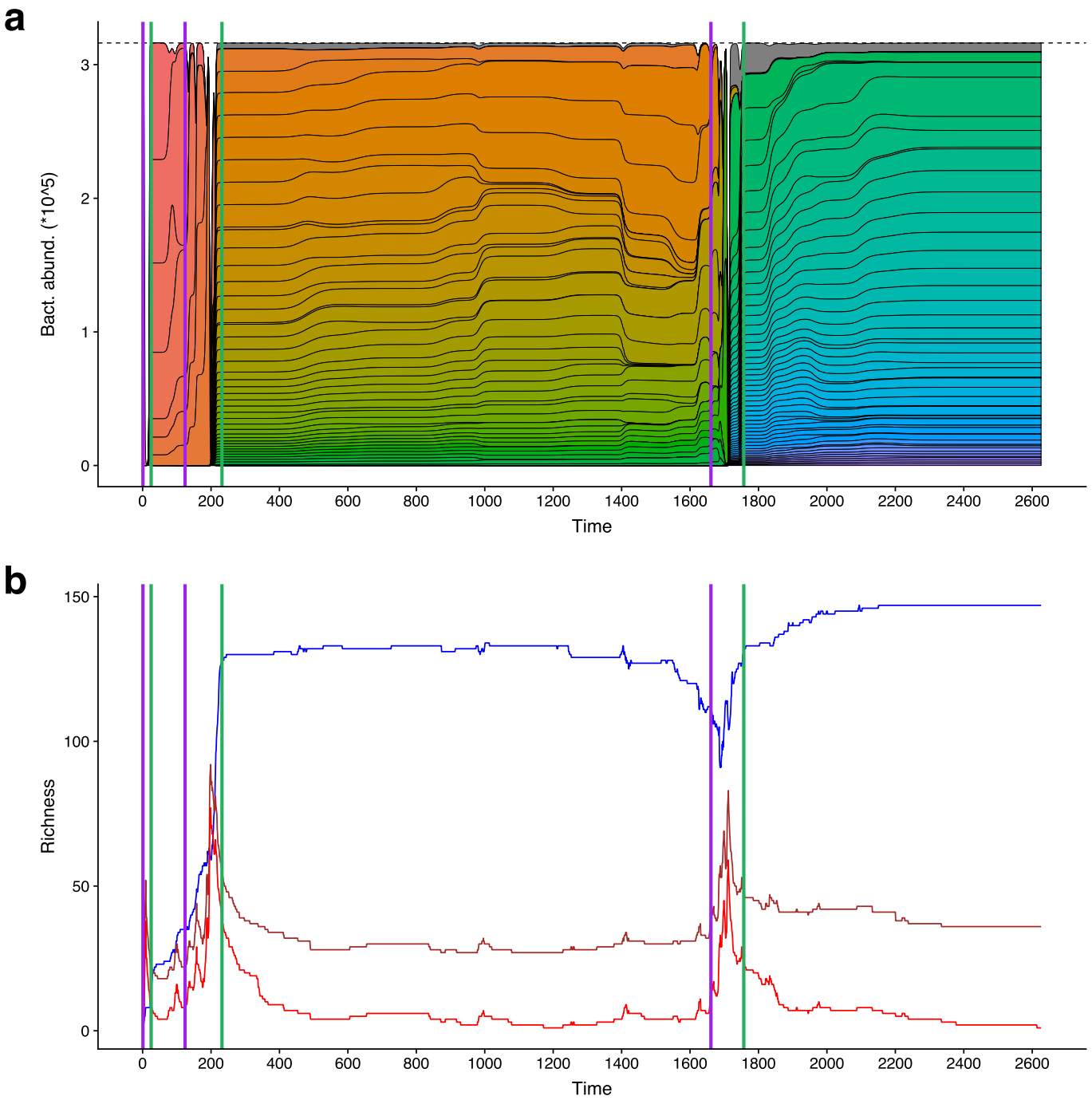
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-020-01312-z.

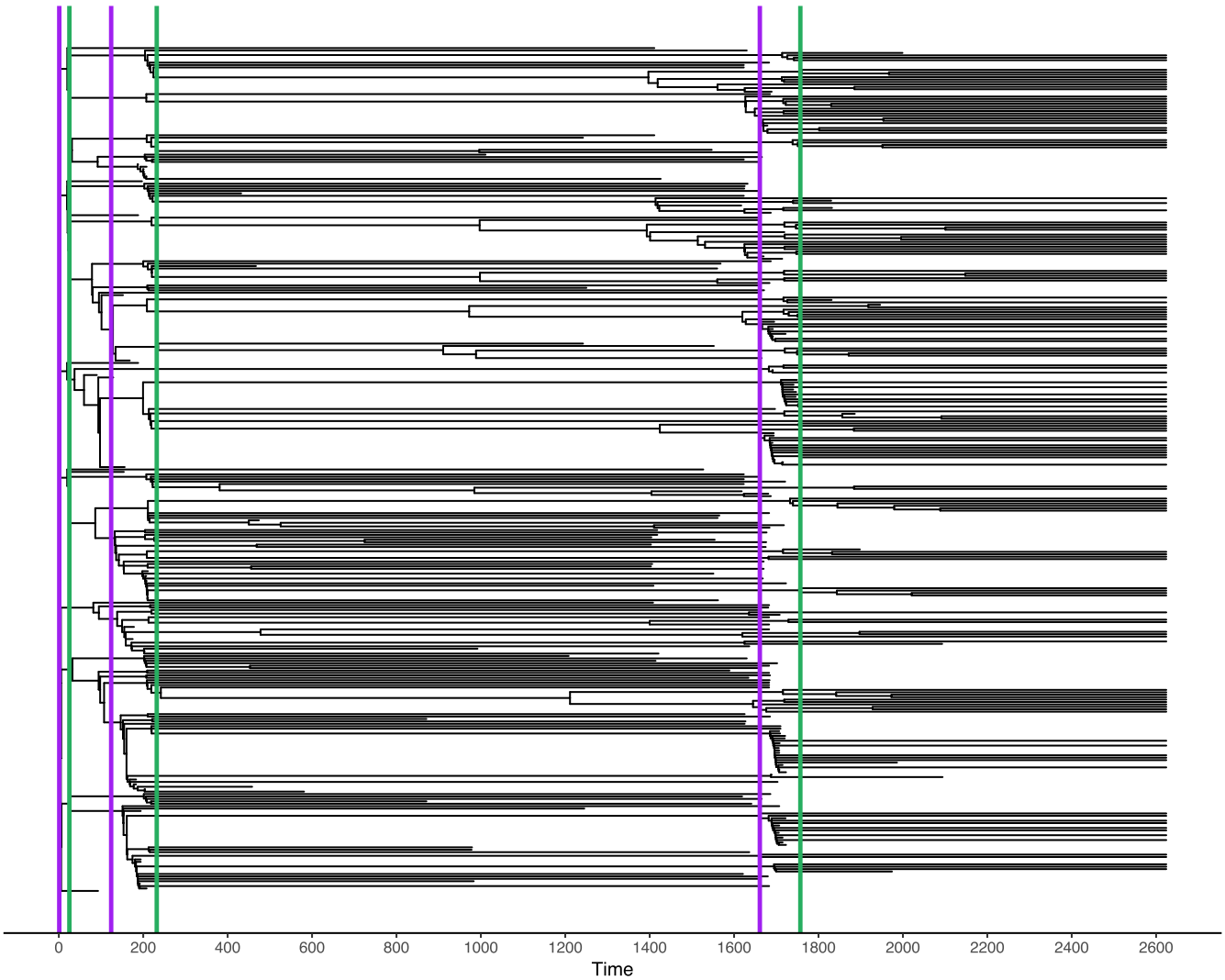
Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-01312-z.

Correspondence and requests for materials should be addressed to R.W. or M.P.

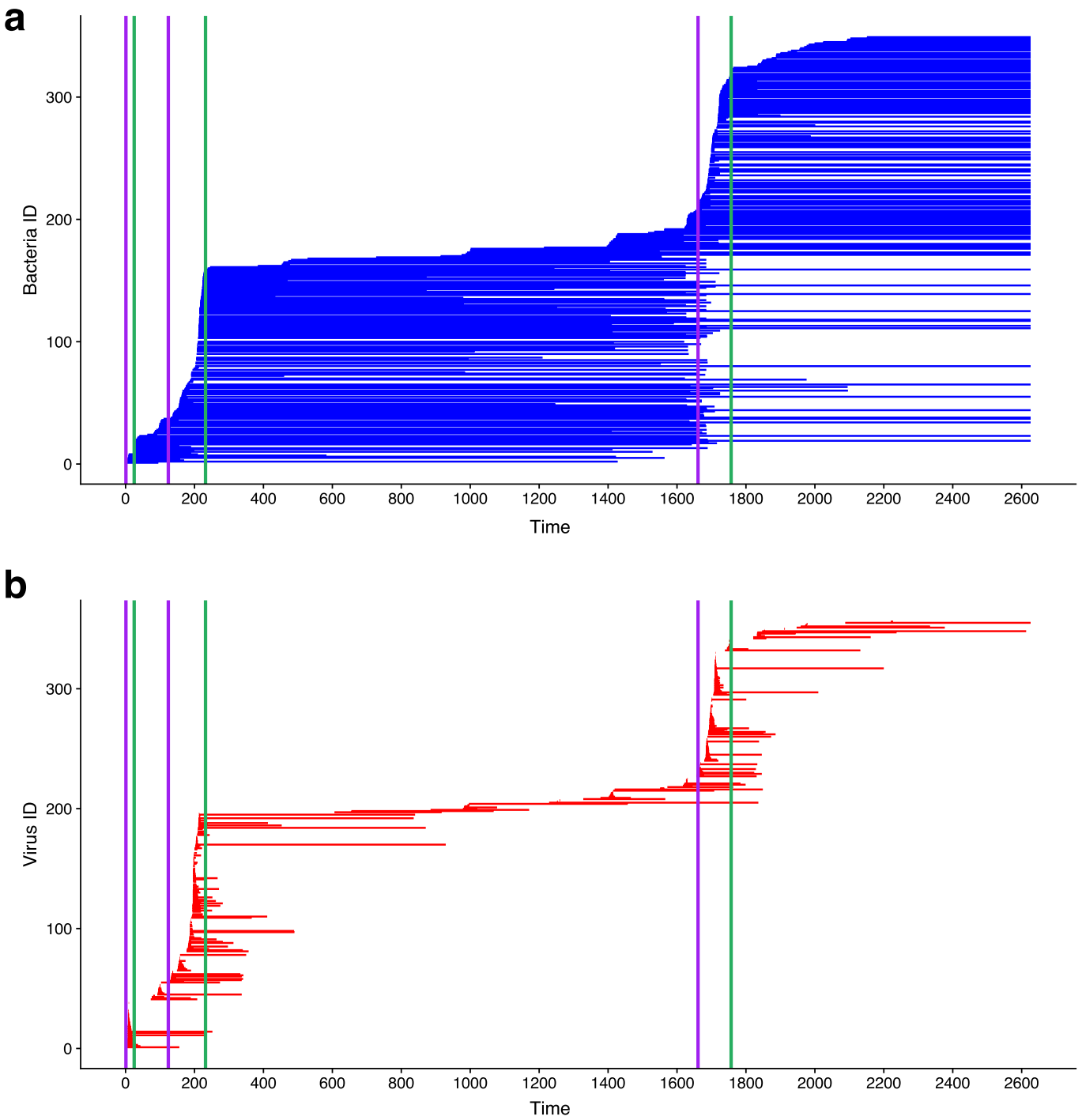
Peer review information Peer reviewer reports are available.



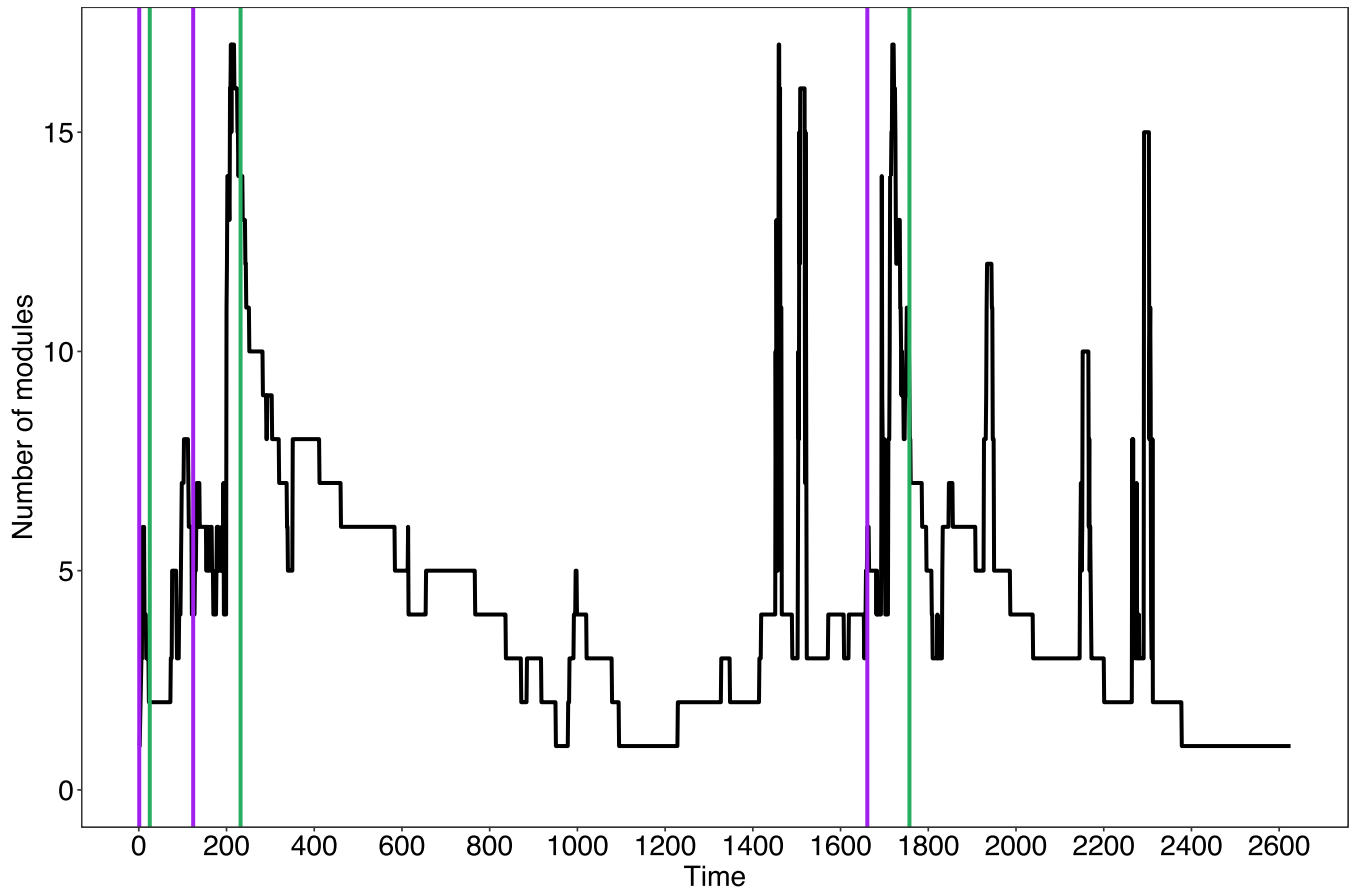
Extended Data Fig. 1 | Viral and host abundance and richness. a, Host abundance. The 100 most abundant strains are colored, the rest are aggregated and shown in gray. **b**, Richness (i.e., number of unique strains) of hosts (blue) and viruses (red). The number of unique spacers (spacer richness) is depicted in brown. During VDRs the abundance of both hosts and viruses fluctuates. As a response to virus diversification, host richness eventually increases despite possible declines at the beginning of the VDR resulting from the initial viral attack.



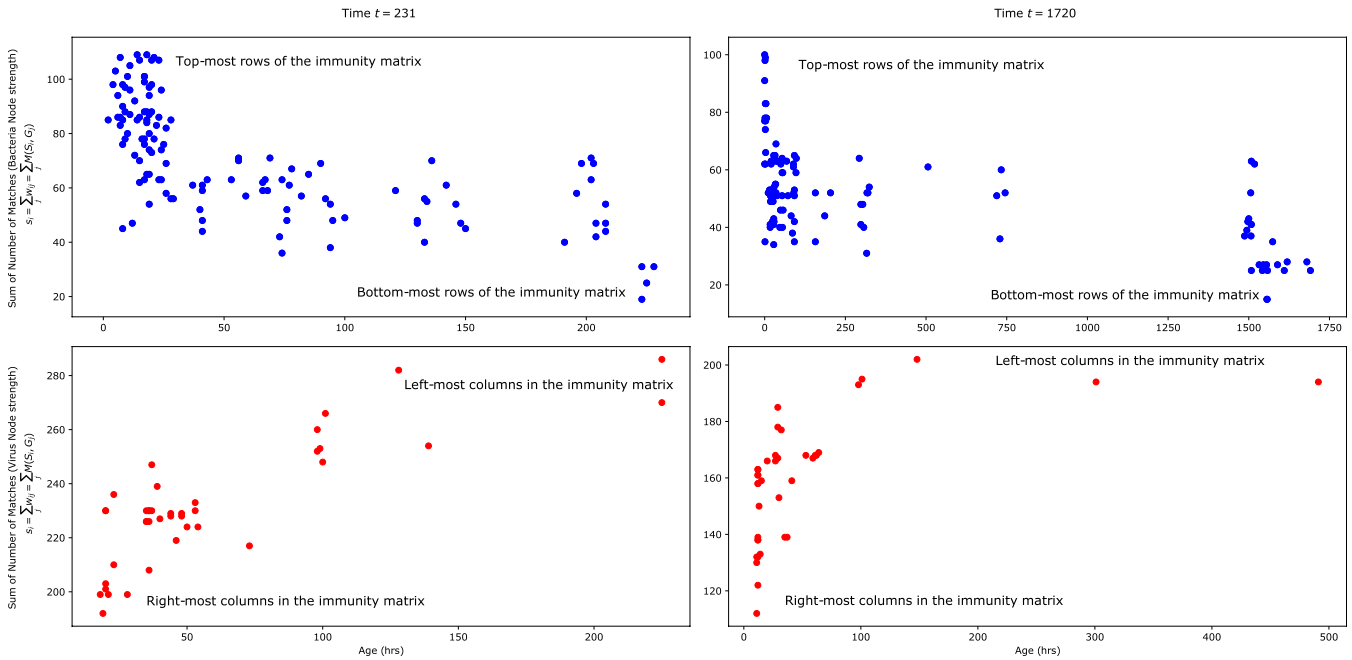
Extended Data Fig. 2 | Host phylogenetic tree. The tree is not inferred, but rather drawn based on exact genealogical data (which strain descends from which) collected during the simulation. Branch length indicates the lifetime of any given host strain. Hosts diversify and go extinct primarily during VDRs, resulting in strain replacement, but strains can also persist from one VDR to the next.



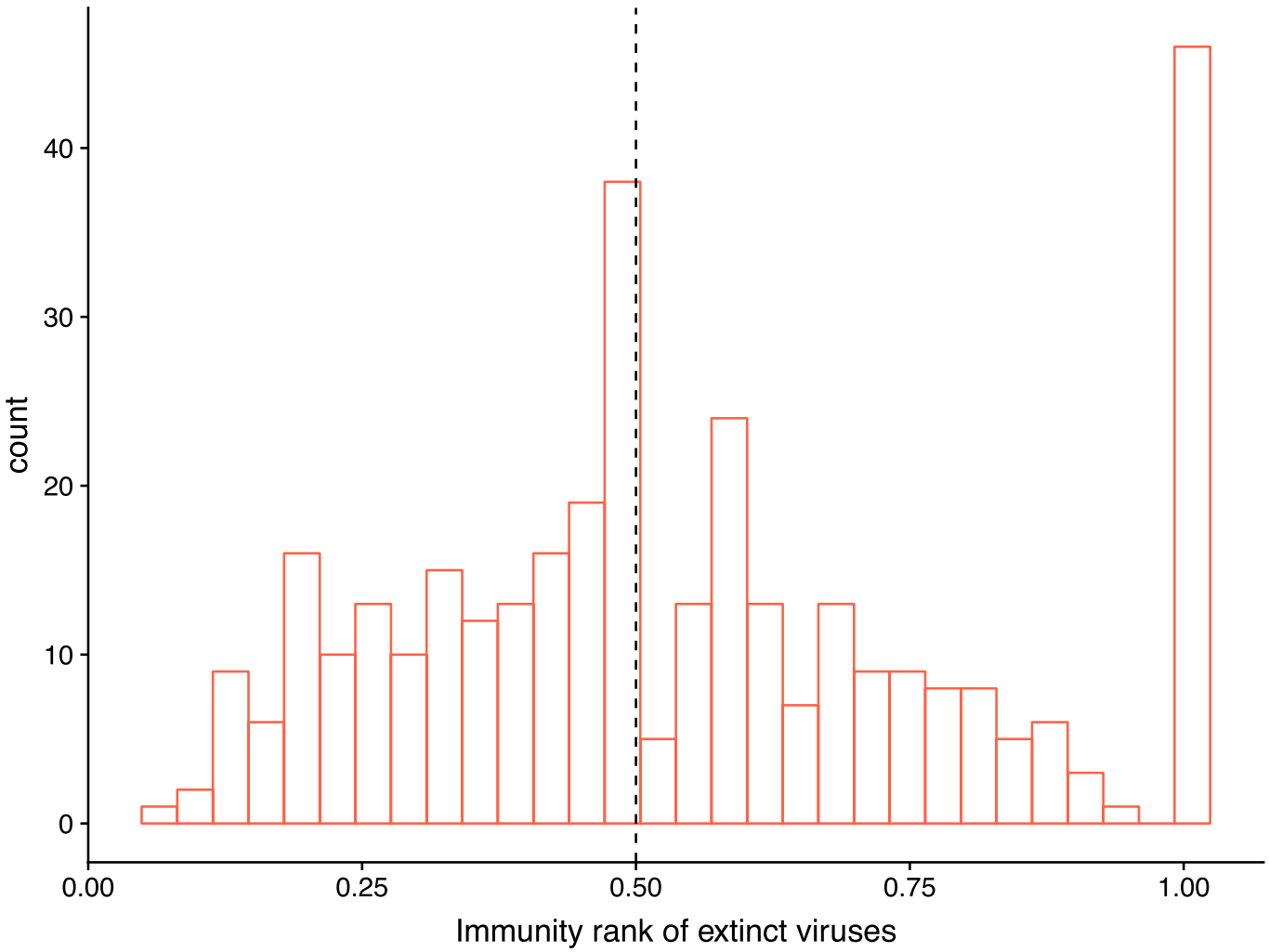
Extended Data Fig. 3 | Host and virus diversification and extinction. Each host (panel **a**) or viral (panel **b**) strain is plotted with a line, starting at the time when the strain was generated and ending when the strain went extinct. During VDRs the rate of diversification of both viruses and hosts is higher than during HCRs. While viruses have relatively short persistence (shorter line lengths), hosts have long persistence and can persist during an entire VDR.



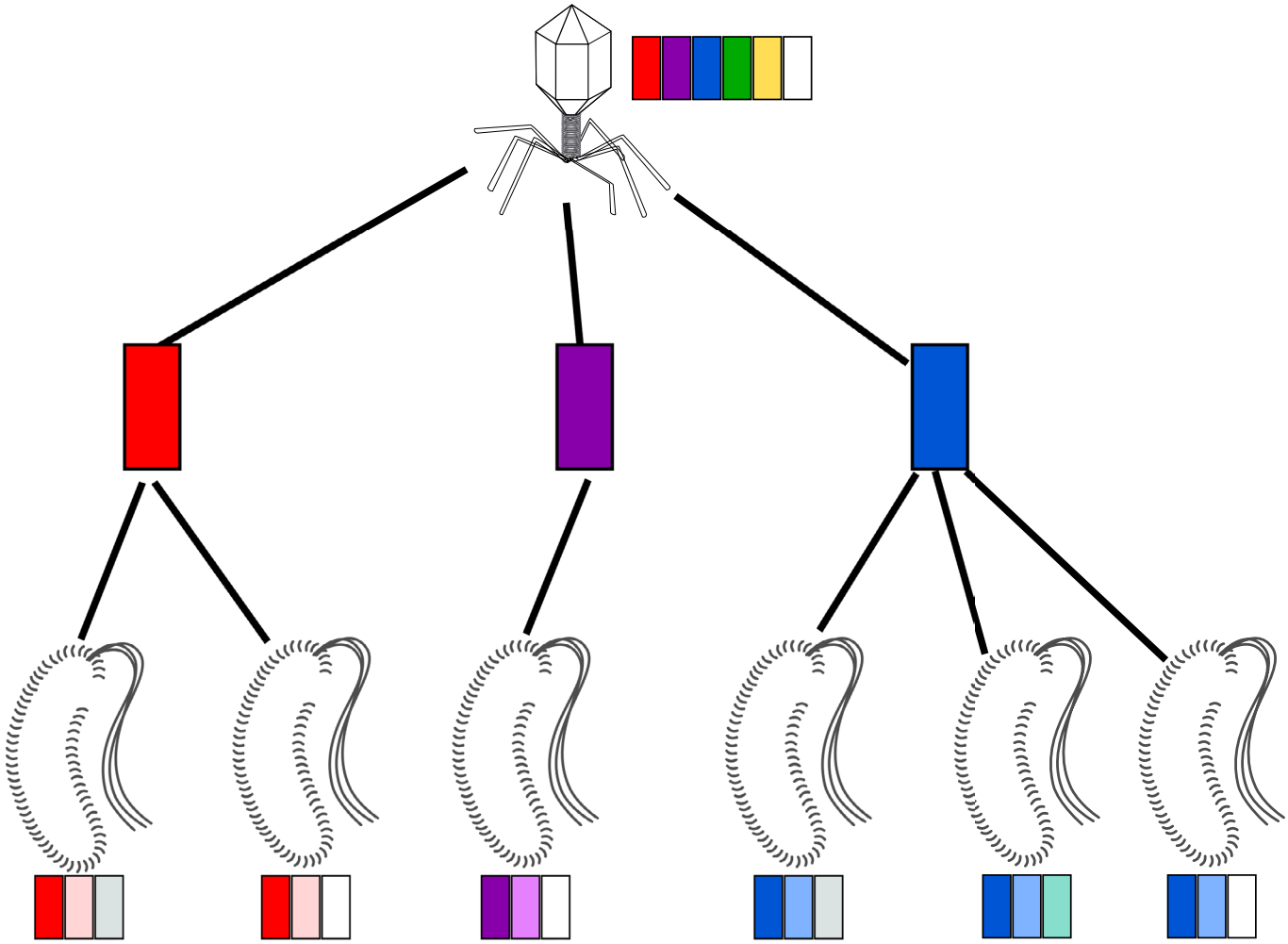
Extended Data Fig. 4 | Modules in the infection network. A time series of the number of modules in the infection network for a single simulation. Modularity enables diversification since it allows the temporary coexistence of different groups of viruses.



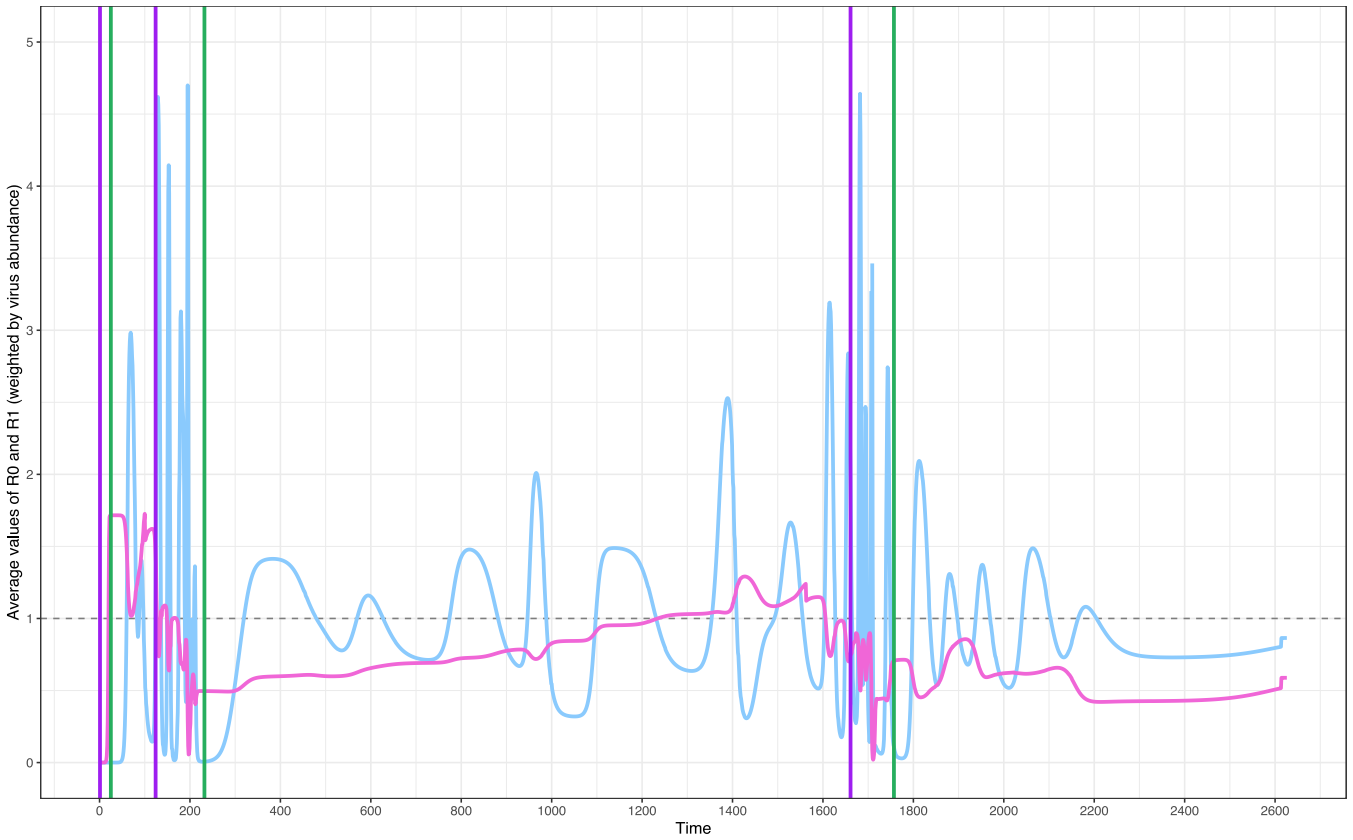
Extended Data Fig. 5 | Host and virus rankings in the weighted nested immunity matrix as a function of age. The plots show node strength (sum of the corresponding number of matches in the immunity network) of hosts (top) and viruses (bottom) against their age measured from the time of their birth, for two selected times before the start of an HCR (leY, t=231 and right, t=1720). Each data point represents a host (top) or virus (bottom) strain. On the different plots, selected groups of youngest and oldest strains are indicated. The oldest host strains occupy the lower rows (low node strength), and their rankings tend to decrease with age. This is because descendants of a given host inherit all of its spacers and add a new one, which always results in an increase in total matches (host node strength). Because they have acquired additional protection, they can grow in abundance and through resulting enhanced encounters and infections, the failure of existing spacers can add redundancy (more than one spacer to the same virus), further contributing to their ranking. The oldest viruses occupy the leftmost columns, with the highest column sums of matches to hosts, since longer lifetimes provide the opportunity for many encounters, and therefore for both the failure of existing spacers (which adds redundancy to a given entry) and the addition of new spacers (which distributes immunity throughout entries). A successful offspring will have mutated a protospacer that confers escape from a given match of the parent; thus, successful descendants exhibit one less match and are placed to the right. It is worth noting that there is considerable variation around the general trends with age, reflecting the complex interplay of the stochastic acquisition of spacers and protospacers with the abundance dynamics which affect both encounter and mutation rates.



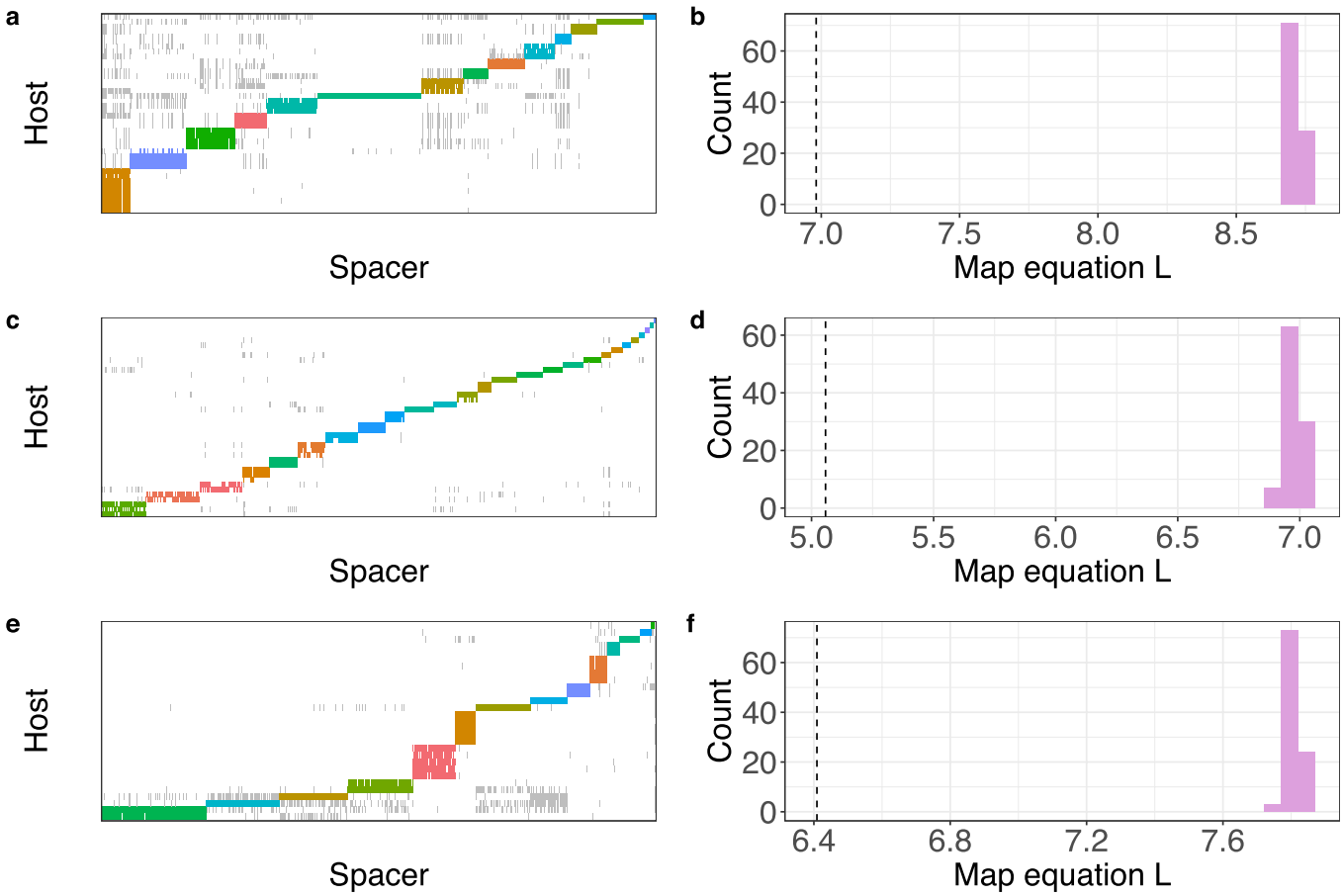
Extended Data Fig. 6 | Order of extinctions. We tested for 'orderly' extinctions in which extinction preferentially happens from the viruses to which hosts have most immunity to those that can infect many hosts. For any virus that went extinct we calculated an 'immunity rank'. Specifically, for a given time step, we calculated the strength of all n virus nodes in the immunity network, $s=(s_1,s_2,\ldots,s_j,\ldots,s_n)$, where s_j is the node strength of virus j (i.e., the sum of the columns in Fig. 2g in the main text). Viruses with higher values of s_j are those that are more to the left in Fig. 2g, and to which hosts have high immunity. We removed duplicate values in s (to avoid ties) and ordered it in ascending order to obtain s'. We then calculated the relative position of s_j in s'. A rank of 1 means that the virus that went extinct was highly ranked (e.g., position 5 out of 5 values will render a rank of 1). 50% of viruses (median indicated by a vertical dashed line) had an extinction rank of 0.5.



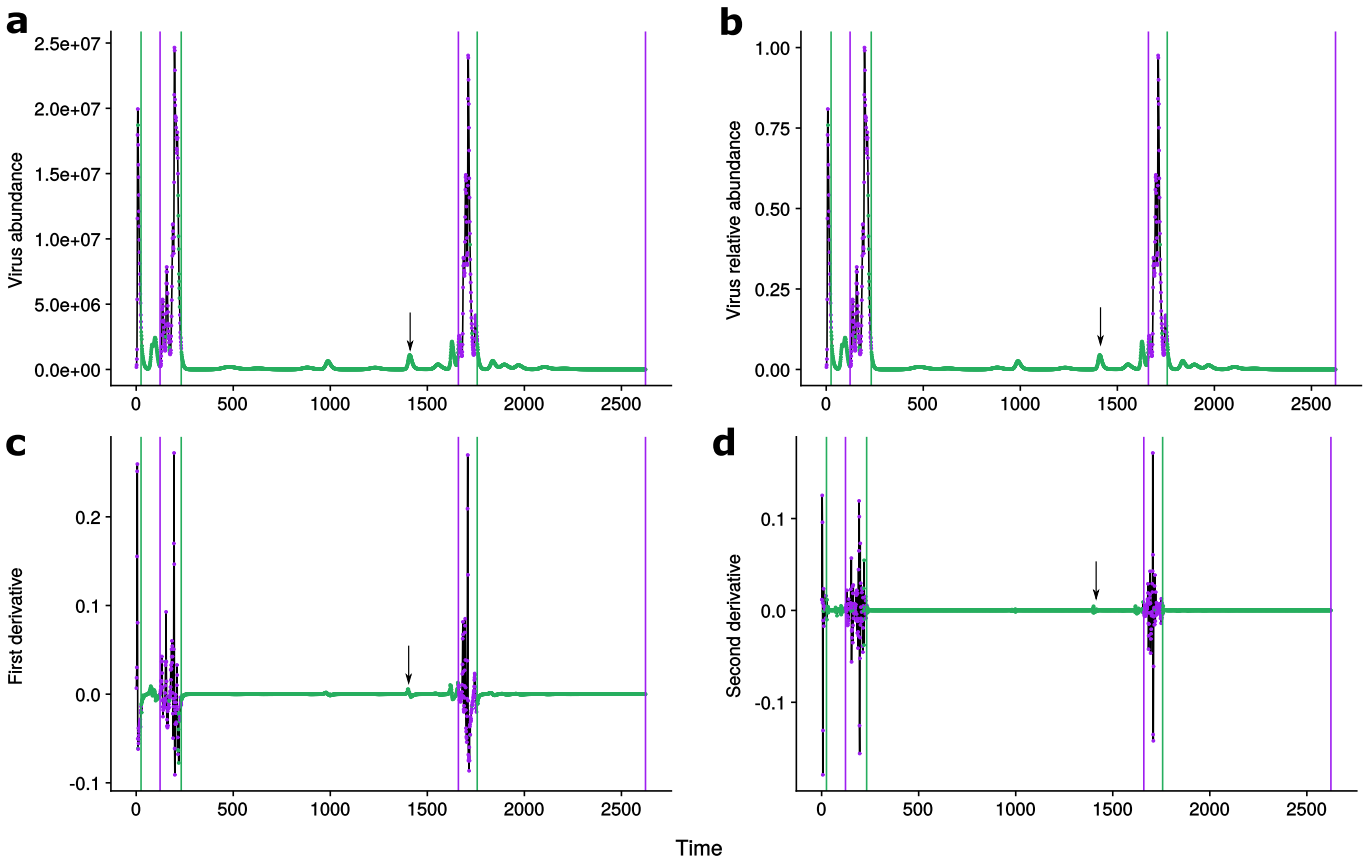
Extended Data Fig. 7 | Viral escape via 1-matches. A tripartite virus-protospacer-host network depicting escape routes for a single virus. Each host is connected to a single protospacer (colored boxes). The spacer composition of strains is shown. Escape occurs through matching colors.



Extended Data Fig. 8 | Reproductive numbers R_0 and R_1. At the entrance into an HCR, the average of reproductive numbers R_j^0 over all virus strains j (weighted by their respective abundances) (light blue line) is considerably below 1. After this initial stage, this mean measure hovers around 1 (horizontal dashed line) and exhibits considerably larger values towards the end of this period before the transition to a VDR. Growth based purely on hosts made available by a single escape mutation is shown here as the weighted-mean of R_j^1 across viruses j (in pink). This quantity exhibits an increasing trend during the HCR, which also raises the potential R_{pot}^j, defined as the sum of R_0^j and R_1^j (not shown, for clarity).



Extended Data Fig. 9 | Modularity of empirical host-spacer networks. Each row represents a different data set: *Sulfolobus islandicus* hosts from Yellowstone (Top). *Pseudomonas aeruginosa* hosts from Copenhagen (middle). *S. islandicus* hosts from the Mutnovsky Volcano in Russia, 2010 (bottom). Panels **a**, **c** and **e**, are host-spacer networks in which interactions within host-spacer modules are colored. Panels **b**, **d** and **f**, are distributions of the map equation (*L*) obtained from networks shuffled by randomly distributing interactions. Value of the observed *L* is depicted with a vertical dashed line.



Extended Data Fig. 10 | Regime definition. Each point in the virus abundance time series in panel **a** is first converted to relative abundance (panel **b**) and then classified into a HCR (green) or VDR (purple). This classification is based on the second difference (panel **d**) (for comparison we also show the first difference in panel **c**). Momentary virus growth periods (marked with an arrow) are not classified as VDR. The final classification is shown using vertical lines. HCRs start with a purple line and VDRs with a green line.

nature research

Corresponding author(s): Mercedes Pascual, Rachel Whitaker

Last updated by author(s): Jul 1, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Simulated and empirical network data are available in the dedicated GitHub repository associated with this paper at: https://github.com/Ecological-Complexity-Lab/CRISPR_networks. All the analysis was done in R version 4.0
Data analysis	Code for simulations and data analysis is available in the dedicated GitHub repository associated with this paper at: https://github.com/Ecological-Complexity-Lab/CRISPR_networks. All the analysis was done in R version 4.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Simulated and empirical network data are available in the dedicated GitHub repository associated with this paper at: https://github.com/Ecological-Complexity-Lab/CRISPR_networks. Data are concentrated in file CRISPR_database_NEE.sqlite. Complete description of the data and all references from which data were obtained are in the paper. There is no restriction on data availability.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We use a combination of simulated and empirical data. Description of empirical data is in the next section. Our simulated data: The model implements the formulation by Childs et al. 2012 (journal Evolution), which combines three main components: ecological population dynamics, stochastic coevolution generating diversity, and molecular identity of hosts and viruses defining CRISPR immunity. Diversification events (spacer acquisition by the host and protospacer mutation of viruses) are modeled stochastically, whereas population dynamics of different sub-populations (i.e., strains) of hosts and viruses are represented deterministically with Lotka-Volterra differential equations. We ran 100 simulations.

Research sample

Our data consists of three data sets, and all relevent references are in the paper, too:
1. Yellowstone: This dataset was collected in hot springs in Nymph Lake at Yellowstone National Park and consists of a population of Sulfolobus islandicus and its contemporary lytic (Sulfolobus islandicus rod-shaped viruses: SIRVs).

2. Russia: This dataset consists of a set of S. islandicus strains isolated from Kamchatka, Russia (in 2000) and sympatric chronic viruses (Sulfulobus spindle-shaped viruses: SSVs).

3. Pseudomonas: This dataset consists of longitudinal sampling of human-adapted Pseudomonas aeruginosa isolates from sputum samples of Cystic Fibrosis patients collected at a hospital in Copenhagen, Denmark and a global set of temperate and lytic P. aeruginosa viruses extracted from NCBI to substitute for a lack of sequenced contemporary viruses. Viruses were grouped based on nucleotide similarity into families known as clusters, and these clusters were assigned a number identifiers which have been described in a previous study (4). In this study we only used viruses from cluster 3 to avoid a false positive result in which we obtain a nested structure due to immune patterns that depend on the cluster (phylogeny).

Illumina sequenced reads from samples were quality filtered using prinseq with the following arguments: -derep 1245 -lc_method entropy -lc_threshold 50 -trim_qual_right 30 \newline -trim_qual_left 30 -trim_qual_type min -trim_qual_rule lt -trim_qual_window 5 \newline -trim_qual_step 1 -trim_tail_left 5 -trim_tail_right 5 -min_len 66 -min_qual_mean 30 \newline -ns_max_p 1 -verbose. Spacers were extracted from quality filtered sequencing reads using an in-house bioinformatic pipeline (in preparation, code available upon request). These scripts utilize known repeats from S. islandicus and P. aeruginosa and BLASTn to extract spacers from reads located between any repeats. BLASTn cutoffs for repeats against reads were based on an e-value of 0.001 with the -task blastn-short argument. After spacer extraction, spacers are grouped based on a hamming distance cutoff by comparing spacers as strings of nucleotides and using a sliding window across each string of basepairs (in preparation, code available upon request). We define unique spacers that have 100\% nucleotide identity to one another, using a hamming distance of 0 between nucleotide sequences. Unmatched overhanging base pairs between spacers were considered mismatches since these are likely independently acquired spacers from sequential protospacers with different PAM sequences. Unique spacers were mapped to strains in order to determine the spacer set per strain. Sulfolobus islandicus isolates in Nymph Lake contained on average 256 spacers per strain ranging from 62 to 520 spacers. Sulfolobus islandicus isolates from Kamchatka contained on average 181 spacers per strain ranging from 20 to 795 spacers. P. aeruginosa isolates contained 34 spacers on average with a range of 4 to 64 spacers. P. aeruginosa, Nymph Lake S. islandicus, and Mutnovsky S. islandicus isolates contained 40, 40 and 50 total alleles respectively.

Spacer matches to protospacers were initially found using BLASTn with a -task blastn-short argument, with an e-value minimum of 0.01. The P. aeruginosa database contained 6,231,702 total bp, with 98 viral genomes ranging from 3,588 bp to 309,208 bp. The SSV BLASTn database contained 34 genomes containing 514,147 total bases with the longest sequence being 18,548 bp and the shortest being 11,323. The SIRV BLASTn database was composed of 10 genomes containing 347,896 total bases with the longest sequence being 32,308 bp and the shortest being 32,308. Protospacer BLAST matches were extended to 3 base pairs longer than the length of the spacer and retrieved with blastdbcmd tool from the blast+ package to retrieve the PAM sequences. Gaps found between alignments were added to these extended protospacer matches. Gaps, or insertion/deletion events, were considered as mismatch when comparing along the entire length of the aligned protospacer and spacer.

Sampling strategy

N/A

Data collection

Described above and in references:
Bautista MA, Black JA, Youngblut ND, Whitaker RJ. Differentiation and Structure in Sulfolobus islandicus Rod-Shaped Virus Populations. Viruses. 2017;9. doi:10.3390/v9050120

Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ. CRISPR associated diversity within a population of Sulfolobus islandicus. PLoS One. 2010;5. doi:10.1371/journal.pone.0012988

Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat Genet. 2015;47: 57–64.

England WE, Kim T, Whitaker RJ. Metapopulation Structure of CRISPR-Cas Immunity in Pseudomonas aeruginosa and Its Viruses. mSystems. 2018. doi:10.1128/mSystems.00075-18

Timing and spatial scale: N/A

Data exclusions: N/A

Reproducibility: On the GitHub repository we include a list with all the simulation seeds in addition to the exact code we used. Running the code with these seeds will perfectly reproduce our simulations.

Randomization: N/A

Blinding: N/A

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging