Coexistence of nestedness and modularity in host–pathogen infection networks

Sergi Valverde[1,2] ✉, Blai Vidiella[3], Raúl Montañez[3], Aurora Fraile[4], Soledad Sacristán[4] and Fernando García-Arenal [4] ✉

[1]Evolution of Technology Laboratory, Institute of Evolutionary Biology, CSIC–Universitat Pompeu Fabra, Barcelona, Spain. [2]European Centre for Living Technology, Venice, Italy. [3]ICREA–Complex Systems Laboratory, Health and Experimental Science Department, Institute of Evolutionary Biology, CSIC–Universitat Pompeu Fabra, Barcelona, Spain. [4]Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) and Instituto nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) and E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, UPM, Madrid, Spain. ✉e-mail: sergi.valverde@ibe.upf-csic.es; fernando.garciaarenal@upm.es

The long-term coevolution of hosts and pathogens in their environment forms a complex web of multi-scale interactions. Understanding how environmental heterogeneity affects the structure of host–pathogen networks is a prerequisite for predicting disease dynamics and emergence. Although nestedness is common in ecological networks, and theory suggests that nested ecosystems are less prone to dynamic instability, why nestedness varies in time and space is not fully understood. Many studies have been limited by a focus on single habitats and the absence of a link between spatial variation and structural heterogeneity such as nestedness and modularity. Here we propose a neutral model for the evolution of host–pathogen networks in multiple habitats. In contrast to previous studies, our study proposes that local modularity can coexist with global nestedness, and shows that real ecosystems are found in a continuum between nested-modular and nested networks driven by intraspecific competition. Nestedness depends on neutral mechanisms of community assembly, whereas modularity is contingent on local adaptation and competition. The structural pattern may change spatially and temporally but remains stable over evolutionary timescales. We validate our theoretical predictions with a longitudinal study of plant–virus interactions in a heterogeneous agricultural landscape.

The coevolution of hosts and pathogens in their environment forms a complex web of multi-scale interactions. Network characterization is central to models of disease spread and emergence, as the frequency and strength of host–pathogen interactions provides information regarding pathogen host ranges and differential host usage, respectively, two traits that determine transmission dynamics. Network characterization is a difficult task because infection structure depends on multiple and interdependent factors. Environmental heterogeneity in space (habitat) and time in host–pathogen interactions is particularly relevant, and spatial scales have been shown to be critical determinants of infection[1–7]. There is abundant evidence of seasonality of infection[8]. Spatial and temporal fluctuations in the distribution and abundance of species determine the probability of encounters with pathogens[9], thereby shaping the structure of host–pathogen interactions[9–13].

Network theory is a powerful approach for identifying non-random associations in ecological interactions[14–19]. Nested and modular structures have been seen as opposite extremes of the ecological spectrum. Specialists in nested networks tend to interact with subsets of species interacting with more generalist species[20]; whereas modular networks consist of groups (or modules) of species interacting less frequently with dissimilar species[14,20–23]. It is assumed that coevolutionary mechanisms could lead to nestedness or modularity, but not both. For example, strong associations lead to species-poor networks characterized by modularity, whereas weak associations promote species-rich, nested networks[24,25]. This is at odds with the ad hoc analysis of specific systems, which suggests a complex interplay between structural patterns.

Both mutualistic networks[26] and host–parasite networks[27] are frequently nested. Extensive analyses of interactions among animal hosts and macroparasites, such as between fish and helminths, mammals and fleas or insects and parasitoids[28–32] revealed that asymmetry of interactions is the rule. Nevertheless, modules defining highly specific sets of interactions can be detected within nested networks[28]. Nestedness has been associated, at least in part, with phages evolving to expand their host range, and with hosts evolving towards increasing resistance to larger sets of phages[23]. Fortuna et al.[33] suggested that the pattern of interactions between bacteria and phages depends on the way the coevolutionary process unfolds, the degree of nestedness being explained by trade-offs between fitness benefits of evolving resistance and infectivity traits and their maintenance costs.

The above interpretation emphasizes deterministic, evolutionary factors of host–pathogen interactions to the detriment of non-deterministic, ecological factors. The process of ecological fitting provides a complementary mechanism[34,35]. Ecological fitting is an inevitable and frequent process in nature that occurs with interactions between plastic organisms and heterogeneous biotic and abiotic environments[34]. Phenotypic plasticity enables many species to persist, colonize a wide range of habitats and form new species interactions even when they have not evolved and continue to exploit the same resources[35,36]. It is relevant that a pathogen's realized host range, that is, the set of host species that are infected by a parasite in a given spatial or temporal context[37], depends on genetic determinants and is also contingent on host community structure[6]. In general, network structure is predicted to be the outcome of both evolutionary and ecological factors.

In this study, we develop a theoretical framework to reconcile inconsistent views about nestedness and modularity in host–pathogen interaction networks. An important limitation is that network

representations cannot easily disentangle genetic and environmental components. The association of host–pathogen networks with single habitats and/or seasons results in a loss of information; that is, inter-habitat or inter-season connections are missing. We show that ecotype–pathogen networks (EPNs) properly capture three-way interactions between host species, pathogen species and habitats. These ecological structures are instances of hypergraphs, which have been seldom used in ecology[38]. We apply this framework to a representative dataset of plant–virus interactions in an agricultural landscape in central Spain with a mosaic of successional habitats. Comparison with a neutral model suggests that nestedness in empirical systems is a consequence of ecological assembly—that is, how species are introduced in the system. This confirms the view that nestedness may have no inherent origin in dynamics or stability[20,39–42]. Furthermore, we do not observe modularity at the ecosystem level because habitats are interdependent. However, the specific interplay between nestedness and modularity depends on temporal and spatial scales. First, finite scaling—for example, seasonal variations in temperature—affects statistical significance of structural patterns. Second, deviations from neutrality suggest a role of local adaptation in modularity. We found that compartmentalization and modularity relate to fitness trade-offs, which, in pathogens, often occur across hosts[9,43–45]. Competition–dispersal trade-offs predict increased modularity in colder seasons and temporal coexistence of nestedness and modularity in warmer seasons.

Results

Defining host ecotype–pathogen infection networks. We sample host infections in a heterogeneous environment following a spatially hierarchical scheme—that is, at a small (habitat) and a large (ecosystem) scale. Within this scheme, our network analysis differs from previous approaches in two ways. First, we aggregate all infections over sites to minimize the effect of fluctuations in measured host ranges, recovering the genetic determinants of host specialization. Second, we consider that a pathogen species infects host ecotypes and not just host species. Infections by a pathogen of a given host species in multiple habitats are really different interactions. Although the identities of the pathogen and host species do not change, community and spatial context are different. Moreover, hosts of the same species differentiate into host ecotypes in distinct habitats. From the point of view of network analysis, a host ecotype is the natural association between host species and habitats. This concurs with the evolutionary biology concept: an ecotype defines a variant of a population or breed—within a species—that has adapted to specific environmental conditions.

Pairwise modelling is insufficient to implement ecotypes. Network analysis omits the spatial context from host–pathogen interactions, which is crucial when understanding the origin of structural patterns—that is, the interplay between nestedness and modularity. Relevance of non-pairwise interactions in ecology has been recognized by studies of trait-mediated indirect interactions, that is, interactions between two species mediated by some other trait of a third species[38]. A trait-mediated indirect interaction can occur whenever a species reacts to the presence of a second species by altering its phenotype. Hypergraphs are mathematical representations of such tripartite, rather than pairwise, ecological interactions[38]. We capture environmentally mediated host–pathogen infections with EPNs (see Fig. 1), which formally corresponds to a hypergraph (see 'Hypergraph modelling of EPNs' in Methods) As we will see, hypergraphs offer an unified framework for the analysis of ecological networks in natural environments.

The neutral evolution of nested infection networks. The fingerprint of assembly rules can be found in the architecture of EPNs. Even in the absence of selective forces, neutrality predicts a non-random community structure[42], such as nestedness in mutualistic networks[41]. Our theoretical model suggests that neutrality is a sufficient condition for nestedness in EPNs, without further assuming other coevolutionary processes[46–48]. This is in agreement with the heterogeneous structure predicted by previous studies[49,50]. Figure 2 illustrates the rules of a neutral model for the growth of EPNs (see Supplementary Discussion, section 2 for details). Starting from a set of almost empty habitats, we simulate a stochastic process of community assembly under equal rates of speciation[42]. At each time step, a novel species is produced from some ancestor species by means of speciation and coevolutionary rules until a target level of species richness is achieved. Unlike previous models, speciation takes into account the internal structure of ecotype species, which depends on both evolutionary and ecological constraints[51]. Pathogen speciation rate α is the probability of generating a new pathogen; reciprocally, we generate a new ecotype with probability $1-\alpha$. Descendant nodes (ecotype or pathogen species) duplicate all interactions from its ancestor (see Fig. 2a). Dynamics of host ecotypes is driven by trade-offs between competition and dispersal[52]. Sympatric speciation[53] occurs with probability θ (host speciation rate), which increases intra-habitat host competition. Conversely, an existing host colonizes a new habitat with probability $1-\theta$, which mitigates intra-habitat competition (see Fig. 2b). External fluctuations can also adjust the strength of inherited interactions. Species divergence is caused by the random redistribution of link weights between parent and daughter species. Finally, link removal takes place depending on predefined thresholds, along with the possible extinction of whole species (see Fig. 2c).

A mean-field approximation to the dynamics of species richness. In parallel with the neutral model, we develop a mean-field approach to species diversity (see 'Mean-field model of EPNs' in Methods). Although this is a deterministic set of equations that reduces the many degrees of freedom in the ecosystem to a single, average effect, it provides insights about the growth of species richness and the origin of structural patterns. A major ecological principle is that competitive exclusion limits species richness. For example, Lotka–Volterra equations predict saturation of species richness due to competition in limited spaces. Competition–colonization trade-offs allow the coexistence of similar hosts within the same habitat (see above). An important difference from previous population-based models is that species coexistence depends on the equilibrium value of relative species richness (which is potentially unbounded) and not on species richness. Figure 3 shows the numerical simulation of the mean-field model after 10^5 speciation events (S). The white line highlights the boundary between existence (non-zero species richness) and extinction of species (that is, $P=T=H=0$; where P is pathogen species richness, T is ecotype richness and H is host species richness). Long-term dynamics of host and pathogen species richness are both functions of ecotype richness, but their dependence can change markedly with the diversity of habitats (C). Figure 3a shows there is little difference between host-species richness and ecotype richness when host speciation and dispersal occur in the same habitat ($C=1$). This situation is noticeably different in ecosystems composed of a mosaic of multiple habitats (see Fig. 3b). More habitats ($C=4$) result in higher ecotype richness than host richness ($T>H$) in the same region of the (α,θ)-parameter space (compare Fig. 3a with Fig. 3b). For low host-speciation rates, growth of ecotype richness $\mathrm{d}T/\mathrm{d}S$ depends on high dispersal; that is, the differ-

ence between C and T/H is large until the landscape is biotically saturated. High host-speciation rates (and very limited dispersal rates) yield a system similar to the single-habitat scenario. A second white line in Fig. 3b visualizes the boundary between ecotype-rich and host-rich systems.

Relative species richness and the coexistence of modularity and nestedness. Disentangling nestedness and modularity is a difficult task[22]. Spatial independence was assumed in the 'nested-modular' organization of a natural assemblage in the Atlantic Ocean, which displays large-scale modularity and local nestedness within each module[54,55]. As shown here, we can still observe spatial coexistence of local modularity and large-scale nestedness when structural patterns are linked to the evolutionary and ecological drivers of species richness. Given N_t ecotypes and N_p pathogen species, the quotient of maximal number of links $N_t \times N_p$ by the total number of nodes is an upper bound to the average degree $k < N_t N_p/(N_t + N_p)$ and also a measure of connectance—that is, the proportion of realized interactions among the potential interactions[56]. When considering ecotypes, the structure of host–habitat association also depends on relative ecotype richness. Specialization in resource use is often considered as an invariant species-property that is consistent across locations, but species display different levels of specialization in varying locations[57,58]. High habitat specificity corresponds to a low average degree $N_t/(N_h + C)$ of the host–habitat network (see Fig. 4a), for example, when there is only one habitat. $N_t/(N_h + C) > 1$ indicates that host species are multi-habitat, and thus we can no longer assume independence of structural patterns from the environment. In this case, there is a dependence of nestedness and modularity with species-richness ratios (see Fig. 4a).

Nestedness and modularity reach a local optimum at the transition from ecotype-rich ($\mathrm{d}T > \mathrm{d}H$) to host-rich ecosystems ($\mathrm{d}T < \mathrm{d}H$). Figure 4c shows that the number of links can be maximized here because of the large combinatorial space of pathogen and ecotype interactions[59]. Figure 4b shows the average number of habitats occupied per host, $N_t/(N_h + C) \approx L_{\mathrm{inter}}$, decreases with host competition and thus increasing modularity (where L_{inter} is the number of intermodular links). More specifically, coexistence between nestedness and modularity corresponds to a species-rich ecosystem; that is, ecotype richness concurs with pathogen richness and there is a high degree of host habitat specificity. Using the mean-field predictions, we can summarize the preconditions for structural coexistence as follows: (1) high pathogen and ecotype speciation rates, (2) availability of multiple habitats $C > 1$, and (3) high intraspecific competition.

Seasonal variation of nestedness and modularity in plant–virus infection networks. Our theoretical predictions for nestedness and modularity are supported by the empirical study of plant–virus infection networks surveyed at different habitats in central Spain from 2000 to 2002 (see Fig. 5 and 'Field work for plant–virus interactions' in Methods). For this, we manually adjusted model parameters (ecotype richness and pathogen-species richness) until comparable structures to the full EPN were obtained (see Supplementary Discussion, section 3 for a validation of parameter choices). High statistical significance of spectral radius $\rho(A)$ matches the neutral prediction for the full EPN (see 'Nestedness in HPINs' in Methods). However, modularity is not always observed in aggregated interactions (see Extended Data Fig. 1). Full EPN is non-modular—that is, it consists of dependent habitats with a large number L_{inter} of inter-habitat connections (see 'Fuzzy bridgeness and habitat modularity in EPNs' in Methods). The implicit assumption of habitat independence in analyses of individual host–pathogen networks (see Fig. 5a) is not universal. To understand habitat interdependence, we decompose network modularity as the sum of local habitat modularities (as described in Methods). Reported values reveal that habitats do not contribute equally to global modularity (see Extended Data Fig. 2). Figure 5c compares local modularity in the four habitats with an ensemble of networks generated with the neutral model. This shows that distance to the centre of the random ensemble decreases with statistical significance of local modularity, suggesting that neutrality is a valid hypothesis for the community structure of the fallow field (that is, it is located at the centre of the ensemble) but not for the other habitats. Habitat-specific variations in local modularity are not always neutral.

Temporal variation of species richness across seasons also influences nestedness and modularity. This is most apparent when studying seasonal subnetworks (see below), which display intermediate modularity values (Q_c) varying between approximately 0.04 and 0.16 (to be compared with the lower modularity of the full EPN; see Extended Data Fig. 1). Figure 6 shows how modularity varies seasonally. Empirical values of seasonal modularity and the average bridgeness B (see 'Fuzzy bridgeness and habitat modularity in EPNs' in Methods) are very close to the model predictions (see Fig. 6a,c) even when the neutral model was fit to the full EPN (see Fig. 6b). In all panels of Fig. 6, the line $B = Q_c$ separates ecotype-rich ecosystems from host-rich ecosystems (see Supplementary Discussion, section 4). Half space $B < Q_c$ corresponds to strong dispersal limitations. In this case, modularity and nestedness are seen as antagonistic patterns that act at different spatial scales[54,55]. The other half space $B > Q_c$ suggests that habitats are not independent from each other. Overall, the degree of habitat interdependence (and modularity) is a continuum driven by the competition–dispersal trade-off.

Seasonal subnetworks shift towards higher modularity in cold seasons (see Fig. 6c). This may be explained by at least two factors related to plant and virus species distributions across space and time. High ecotype richness in the summer compared with the winter network (66 versus 23; Extended Data Fig. 1) offer generalist viruses more opportunities for plasticity in host range and in specialized host use. This results in a more complex network in summer, with more inter-habitat connections (see Supplementary Figs. 1 and 3). In addition, the lowest N_h/N_t ratio in summer yields higher bridgeness and lower modularity (see Fig. 6c). Intermediate modularity of the spring and autumn networks (see Fig. 6c), with intermediate and similar values of N_h/N_t ratios (Fig. 6c, inset) agrees with this hypothesis but does not explain the higher modularity of the autumn network (Extended Data Fig. 1 and Supplementary Figs. 2 and 4). This might be related to reduced inter-habitat virus transmission in the autumn compared with the spring. Indeed, 10 of the 11 viruses analysed are transmitted by aphid vectors, which in central Spain are active along the whole year, with morphs and aphid flights having a maximum in late spring and a secondary maximum in early autumn[60,61]. This factor, combined with the higher host-community differentiation among habitats in the autumn than in spring (see bridgeness values in Extended Data Fig. 1), will result in aphid flights leading to higher inter-habitat transmission in the spring than in the autumn.

Discussion

A main goal of evolutionary biology is to understand the origin of biodiversity. A well-studied hypothesis is that species distribution is constrained, as some species necessarily perform better in some environments that others. Fitness trade-offs establish niche differences that allow coexistence of similar species within the same

habitat[51,52]. Conversely, neutral theory assumes that all species are ecologically identical and that niche differences are not required to explain species coexistence[62]. Real ecosystems represent an intermediate situation involving both selection and contingency mechanisms. To what extent do differences between species determine the evolution and structure of ecological communities?

Hypotheses of neutrality and niche differentiation have deeply influenced our understanding of host–pathogen networks. It has been assumed that modularity and nestedness are antagonistic[23] and that coexistence of nestedness and modularity implies a separation of spatial scales, as in the 'nested-modular' hypothesis[54,55]. This does not fully recognize phenotypic plasticity to environmental variation, owing in part to the limitations of pairwise interactions in host–pathogen networks. By explicitly representing three-way interactions, hypergraphs can show the coexistence of local modularity with large-scale nestedness. The modularity of EPNs distinguishes between groups of hosts competing in the same environment (increasing modularity) or dispersing to different habitats (decreasing modularity), even when host species overlap habitats. Widespread host adaptation to ecotypes at small spatial scales[63–65] will be associated to high modularity. Further, bridgeness will reveal that generalist pathogens act in a range of spatial scales. The picture that emerges suggests that ecosystems exist in a continuum between nested-modular and nested networks driven by intraspecific competition (or relative ecotype richness per host-species richness). Depending on the spatiotemporal scales involved, these apparently opposing structures may coexist.

Agricultural landscapes provide a good opportunity to analyse highly interdependent habitats. Empirical data on plant–virus interactions in an agricultural landscape in central Spain provides a good example of EPN. As in other studies, we divide the landscape into different habitats, but each habitat is characterized by overlapping species assemblages. Our detailed surveys of this rich ecosystem, conducted monthly during three years in four different habitats[6], capture spatial and temporal variations at the landscape scale. We report significant nestedness but low modularity, which is consistent with the assumption of highly interdependent habitats. Deviations from neutrality in habitat-specific networks expose their local adaptations. In winter, lower richness and abundance of host plant species promotes virus specialization within their host range[66] and thus increasing modularity. Low density of insect vector populations[60,61] also reduces across-habitat transmissions. Accordingly, frequency of broad-host-range viruses bridging multiple habitats drops in cold seasons.

Our framework should be expanded to accommodate predictions of species abundances in each habitat. This has been addressed by island biogeography, where small isolated islands have fewer species with low abundances[67]. This cannot be represented by our model unless we relax the assumption of homogeneous habitats with unlimited capacity. Host and pathogen species richness is unbounded and grows over time, that is, the fixed point for species richness is a repulsor. However, the ubiquity of global patterns (in particular, nestedness) appears to be independent of specific extinction events. Statistical significance of these patterns can fluctuate (for example, seasonally) but will nevertheless remain stable over evolutionary timescales.

In summary, our work represents an experimentally based modelling study of EPNs. More data than we presently have is needed to fully characterize the coexistence of modularity and nestedness in natural systems. Nevertheless, we offer a theoretical explanation for structural coexistence at the landscape level. This suggests that the morphospace of host–pathogen interaction networks is far larger than was previously reported in the literature. These results will bring us closer to understanding how environmental heterogeneity mediates structural patterns in many natural systems—from host–pathogen infections to microbial interactions.

Methods

Hypergraph modelling of EPNs. As described in the main text, we implement network analysis of ecotypes using hypergraphs. A hypergraph connects three kinds of entities: host species, pathogen species and habitats (see Supplementary Discussion, section 7), whereas the host–pathogen network does not capture the influence of the environment on host–pathogen interactions (see Fig. 1a). To ease interpretation, we use EPN instead of the underlying (and equivalent) hypergraph. We define the EPN as a decorated bipartite graph $G=(P,T,E)$ with links in the set $E \quad P\times T$ connecting pathogens in the set P with host ecotypes in the set $T \quad H\times C$. Here, an ecotype node $t=(h,c)\in T$ is the association between host species $h\in H$ and habitat $c\in C$ in which at least one infection was recorded. The above resembles a common bipartite network and yet usage of ecotypes marks an important difference. Figure 1b compares the EPN and HPIN representations of a small ecosystem. Both networks involve pairwise interactions, but links in the EPN connect nodes of different complexity. It can be shown that the ecotype set T defines the links of a host–habitat bipartite network $G_T=(H,C,T)$, which is an important case for the analysis of modularity (see 'Fuzzy bridgeness and habitat modularity in EPNs' section).

Mean-field model of EPNs. This section develops a mean-field approach to species diversity in ecotype–pathogen networks. We define the state of the system as the pathogen species richness P, host species richness H and ecotype richness T. The following system of equations couples the dynamics of ecotypes, pathogen and hosts in a habitat-explicit setting (see Supplementary Discussion, section 6 for a full analysis):

$$\frac{\mathrm{d}}{\mathrm{d}S}H=(1-\alpha)\theta\frac{H}{T}-\frac{\delta}{C}$$

$$\frac{\mathrm{d}}{\mathrm{d}S}P=\alpha-\delta\frac{P}{T}$$

$$\frac{\mathrm{d}}{\mathrm{d}S}T=\underbrace{(1-\alpha)(1-\theta)}_{\rho_\mathrm{d}}T\left(C-\frac{T}{H}\right)+\underbrace{(1-\alpha)\theta}_{\rho_\mathrm{s}}\frac{H}{T}-\delta T$$

where $0\le\alpha\le1$ is the pathogen speciation rate, $0\le\theta\le1$ is the ecotype speciation rate, S is the number of speciation events, δ is the intrinsic ecotype extinction rate, $\rho_\mathrm{d}=(1-\alpha)(1-\theta)$ is the dispersal rate, $\rho_\mathrm{s}=(1-\alpha)\theta$ is the speciation rate, and there are $C>1$ different habitats. This is a crude representation of antagonistic pathogens and host species in space. Competition forbids more than one speciation event happening at each time step. Growth of host-species richness $\mathrm{d}H/\mathrm{d}S$ is proportional to habitat specificity or the ratio $0\le H/T\le1$ between host species richness and ecotype. Habitat specificity is minimal $H/T\approx0$ when a few host species dominate many habitats, and maximal $H/T\approx1$ when local host adaptations are common. These are opposite extremes represent spatially segregated (or broad) and aggregated (or restricted) hosts, respectively. Real ecosystems lie somewhere in between due to competition–dispersal trade-offs. In this case, the condition $\mathrm{d}H/\mathrm{d}S=0$ defines an equilibrium value H^*:

$$H^*=\frac{\delta}{\rho_\mathrm{s}C}T^*$$

Similarly, we let $\mathrm{d}P/\mathrm{d}S=0$ to find pathogen species richness at equilibrium:

$$P^*=\frac{\alpha}{\delta}T^*$$

which reflects intrinsic ecotype–pathogen competition (see above). The mean-field model predicts ecosystem existence when (see Supplementary Discussion, section 6):

$$\delta^2+C\rho_\mathrm{d}(\rho_\mathrm{s}-\delta)>0$$

Nestedness in HPINs. A standard approach[23] defines a HPIN as a bipartite graph $G=(P,H,E)$ having two disjoint set of nodes, that is, pathogens P and hosts H, where a link $(p,h)\in E$ indicates that the pathogen $p\in P$ can infect the host $h\in H$ (see Fig. 1a). Useful network measures include node degree (the count of links attached to any node):

$$k_i=\sum\nolimits_j A_{ij}$$

where the adjacency matrix $A_{ij}=1$ if there is a link between nodes i and j, or $A_{ij}=0$ otherwise[68]. Degree is a local measure of potential infections; that is, pathogen degree is a surrogate of host range, whereas host degree corresponds to propensity to infection. Infection heterogeneity, where a small proportion of hosts support a large proportion of pathogens, is a common finding in many empirical systems[6,9,69]. In this context, nestedness is a global measure of the tendency of low-degree species to interact with a subset of highly connected species. This pattern has a clear fingerprint in the network structure; that is, nestedness is a systematic arrangement of non-zero entries in the adjacency matrix[70]. Although there is empirical evidence that many ecological networks tend to be binarily nested, this picture can change when using information about abundances and

interaction frequencies[71]. An accurate characterization of nestedness considers the frequency of infections $\omega = \omega_{ij}$ between the nodes i and j. Then, we can compute quantitative nestedness over the following square $\mathbb{S} \times \mathbb{S}$ matrix:

$$\omega = \begin{matrix} 0 & \omega_{|P|\times|H|} \\ \omega_{|H|\times|P|} & 0 \end{matrix}$$

where $\mathbb{S} = |P| + |H|$ is the total species diversity (that is, the sum of host and pathogen richness). Notice this quantitative matrix has a block off-diagonal form due to the bipartite nature of host–pathogen interactions. The eigenvalues of this matrix, $\{\lambda_k\}(k = 1, \ldots, \mathbb{S})$,can be systematically calculated. It has been shown that the largest eigenvalue of this matrix, that is, the spectral radius $\rho(\omega) = \max\{|\lambda_1|, \ldots, |\lambda_{\mathbb{S}}|\}$, is a measure of nestedness[71]. Spectral radius is a better nestedness estimator than binary nestedness because: (1) it does not depend on the spatial arrangement of presence–absence interactions in the matrix and (2) it makes a formal link between structure and the dynamic stability analysis. Measures of ecosystem richness, such as network size $N = |P| + |H|$, are proxies for the statistical significance of nestedness. Theoretical and empirical studies have found that nestedness is sensitive to network size N and also to average number of connections $\langle k \rangle$. Network size is the most fundamental among all parameters defining a network, and determines many traits, including large-scale structural patterns[59]. For example, nestedness scales as $\rho(\omega) \sim \sqrt{N}$ in large random graphs[72]. In addition, empirical nestedness values should always be interpreted in the context of null distributions of random networks with similar size and density of connections (see Supplementary Discussion, section 4 for a full statistical analysis).

Fuzzy bridgeness and habitat modularity in EPNs. A pathogen can infect the same host species in different habitats. One way to classify species interactions is according to the location where they take place. For example, we can provide a fuzzy habitat membership vector for each species i as follows:

$$\mathbf{u}_i = u_{i1}, u_{i2}, \ldots, u_{ic}, \ldots, u_{i|C|}$$

where $u_{ic} = w_{ic} / \sum_j w_{ij}$ is the relative frequency of species i in the habitat $c \in C$. Each species distributes a total membership degree of $\sum_i u_{ic} = 1$ among all habitats $1 \leq c \leq |C|$. This vector explains how the species is spatially distributed. Bridgeness $0 \leq b_i \leq 1$ measures the degree to which a given species i is shared among different habitats[73]. This is the (normalized) Euclidean distance of membership vector $\mathbf{u}_i$ from reference vector $(1/|C|, \ldots, 1/|C|)$:

$$b_i = 1 - \sqrt{\frac{K}{K-1}\sum_{c=1}^{K}\left(u_{ic} - \frac{1}{K}\right)^2}$$

where $b_i = 0$ if species i can be found in only one habitat and $b_i = 1$ when it participates in all habitats with exactly the same membership degree. The latter will be the case for some species that function as 'bridges', that is, nodes that connect together several habitats in the network. Average bridgeness measures the overall spatial distribution of interactions:

$$\langle B \rangle = \frac{1}{N}\sum_{i=1}^{N} b_i$$

Modularity is the quality that subsets of nodes (modules) display higher density of connections among them than with the rest of nodes[74]. We compute the modularity of EPNs using the fuzzy membership vectors defined above. Following Nepusz et al.[73], we define habitat modularity Q_c as:

$$Q_c = \frac{1}{2m}\sum_{k\in C} Q_c^k = \frac{1}{2m}\sum_{k\in C}\sum_{p_i,h_j}\left(A_{ij} - \frac{k_i k_j}{2m}\right)s_{ijk}$$

where the inner sum Q_c^k (or local habitat modularity) defines the contribution of habitat $k \in C$ to the full ecosystem, and $s_{ijk} = u_{ik}u_{jk}$ is the habitat similarity between pathogen $p_i \in P$ and host $h_j \in H$, and m is the number of edges between nodes in the network. Modularity relates habitat diversity with the locality of infections (see Supplementary Discussion, section 5). Positive modularity values indicate that habitats integrations are more aggregated than random expectation. A neutral value of $Q_c \approx 0$ corresponds to a non-modular (or random) EPN. High modularity corresponds to a set of independent habitats, whereas low modularity values indicates a high number of inter-habitat connections. Local habitat modularity evaluates the difference between intra-modular interactions and their expected frequency based on a random null model[74]. By summing all local modularities we obtain an overall modularity value for the whole ecosystem.

Trajanovski et al[75]. derive an alternative expression for the modularity, which is equivalent to the previous equation:

$$Q_c = 1 - \frac{1}{C} - \frac{L_{\text{inter}}}{L} - \frac{1}{2C}\sum_{i=1}^{C}\sum_{j=1}^{C}\left(\frac{D_i - D_j}{2L}\right)^2$$

where C is the number of habitats (modules), L is the total number of links in the network, L_{inter} is the number of intermodular links, and D_i is the sum of node degrees within module c_i. This equation suggests that maximally modular networks with $Q_c = 1 - 1/C$ have disconnected ($L_{\text{inter}} = 0$) and well-balanced ($D_i \approx D_j$) modules. Modularity will be high when host species infected by the same pathogen species share few habitats (low L_{inter}), whereas low modularity corresponds to an uneven distribution of host species infected by pathogens in many habitats (high L_{inter}).

Field work for plant–virus interactions. Plant–virus infection networks were empirically analysed in a heterogeneous agricultural landscape in central Spain in four habitats characterized by different degrees of human intervention: crop (melon) fields, agricultural fields between crop seasons (fallow fields), edges between agricultural fields (edges) and abandoned fields that have not been tilled for at least 2 yr (wastelands). Plants were sampled monthly during 3 yr by systematic collections along fixed itineraries, with no consideration of symptom expression, as described[6,66]. Infection by 11 generalist viruses, that is, alfalfa mosaic virus, beet western yellow virus, bean yellow mosaic virus, cucumber mosaic virus, lettuce mosaic virus, papaya ringspot virus, potato virus Y, tomato spotted wilt virus, turnip mosaic virus, watermelon mosaic virus and zucchini yellow mosaic virus, in the sampled plants was analysed by double-antibody sandwich enzyme-linked immunosorbent assay, using commercial antisera (Bio-Rad), according to the manufacturer's instructions. A total of 80 different plant species were sampled over habitat and season, 51 species were found in edges, 45 in wastelands, 40 in fallow fields and 12 in melon fields (excluding melon). Of the 80 identified plant species, 53 were infected by at least one virus at least once, 36 species were infected in edges, 26 in wastelands, 18 in fallow fields and 10 in melon fields. Temporal sampling generates a large number of distinct subnetworks, capturing seasonal and habitat variability of species assemblages and interactions. By aggregating all interactions from this dataset, we reconstruct a full EPN of $N_h = 53$ different host plant species, $N_p = 11$ pathogen (virus) species, and $N_i = 90$ ecotypes spanning $C = 4$ different habitats (edge, wasteland, fallow field and melon field; see Fig. 5a). Aggregation aims to improve the estimate of the fundamental host range of pathogens, and thus recover the genetic determinants of host specialization.

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

Code availability

For the analysis of nestedness and modularity we used the open-source software packages FALCON (https://github.com/sjbeckett/FALCON) and BiMat (http://bimat.github.io). Network simulation codes will be provided by the authors upon reasonable request.

References

1. Handel, A., Lebarbenchon, C., Stallknecht, D. & Rohani, P. Trade-offs between and within scales: environmental persistence and within-host fitness of avian influenza viruses. *Proc. R. Soc. B* **281**, 20133051 (2014).
2. Laine, A.-L. Evolution of host resistance: looking for coevolutionary hotspots at small spatial scales. *Proc. R. Soc. B* **273**, 267–273 (2006).
3. Hamer, G. L. et al. Fine-scale variation in vector host use and force of infection drive localized patterns of West Nile virus transmission. *PLoS ONE* **6**, e23767 (2011).
4. Viana, M. et al. Dynamics of a morbillivirus at the domestic-wildlife interface: canine distemper virus in domestic dogs and lions. *Proc. Natl Acad. Sci. USA* **112**
5. Blumenthal, D., Mitchell, C. E., Pysek, P. & Jarosík, V. Synergy between pathogen release and resource availability in plant invasion. *Proc. Natl Acad. Sci. USA* **106**, 7899–7904 (2009).
6. McLeish, M., Sacristán, S., Fraile, A. & García-Arenal, F. Scale dependencies and generalism in host use shape virus prevalence. *Proc. R. Soc. B* **284**, 20172066 (2017).
7. Carlsson-Granér, U. & Thrall, P. H. Host resistance and pathogen infectivity in host populations with varying connectivity. *Ecol. Lett.* **69**, 926–938 (2015).
8. Martinez, M. E. The calendar of epidemics: seasonal cycles of infectious diseases. *PLoS Pathog.* **14**, e1007327 (2018).
9. McLeish, M., Fraile, A. & García-Arenal, F. Ecological complexity in plant virus host range evolution. *Adv. Virus Res.* **101**, 293–339 (2018).
10. Ostfeld, R. S. & Keesing, F. The function of biodiversity in the ecology of vector-borne zoonotic diseases. *Can. J. Zool.* **78**, 2061–2078 (2000).
11. Gurarie, D. & Seto, E. Y. W. Connectivity sustains disease transmission in environments with low potential for endemicity: modelling schistosomiasis with hydrologic and social connectivities. *J. R. Soc. Interface* **6** 495–508 (2009).

12. Yang, H., Tang, M. & Gross, T. Large epidemic thresholds emerge in heterogeneous networks of heterogeneous nodes. *Sci. Rep.* **5**, 13122 (2015).
13. Webster, J. P., Borlase, A. & Rudge, J. W. Who acquires infection from whom and how? Disentangling multi-host and multi-mode transmission dynamics in the 'elimination' era. *Phil. Trans. R. Soc. B.* **372**, 20160091 (2017).
14. Bascompte, J. & Jordano, P. Plant–animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **38**, 567–593 (2007).
15. Lewinsohn, T. M., Prado, P. I., Jordano, P., Bascompte, J. & Olesen, J. M. Structure in plant–animal interaction assemblages. *Oikos* **113**, 174–184 (2006).
16. Ings, T. C. et al. Ecological networks—beyond food webs. *J. Animal Ecol.* **78**, 253–269 (2009).
17. Blüthgen, N., Fründ, J., Vázquez, D. P. & Menzel, F. What do interaction network metrics tell us about specialization and biological traits? *Ecology* **89**, 3387–3399 (2008).
18. Blüthgen, N. Why network analysis is often disconnected from community ecology: a critique and an ecologist's guide. *Basic Appl. Ecol.* **11**, 185–195 (2010).
19. Dormann, C. F., Fründ, J. & Schaefer, H. M. Identifying causes of patterns in ecological networks: opportunities and limitations. *Annu. Rev. Ecol. Evol. Syst.* **48**, 559–584 (2017).
20. Pawar, S. Why are plant–pollinator networks nested? *Science* **345**, 383 (2014).
21. Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E. & Taylor, W. W. Compartments revealed in food-web structure. *Nature* **426**, 282–285 (2003).
22. Fortuna, M. A. et al. Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* **79**, 811–817 (2010).
23. Weitz, J. S. et al. Phage–bacteria infection networks. *Trends Microbiol.* **21**, 82–91 (2013).
24. Guimarães, P. R. et al. Interaction intimacy affects structure and coevolutionary dynamics in mutualistic networks. *Curr. Biol.* **17**, 1797–1803 (2007).
25. Pires, M. M. & Guimarães, P. R. Interaction intimacy organizes networks of antagonistic interactions in different ways. *J. R. Soc. Interface* **10**, 20120649 (2013).
26. Thebault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**, 853–856 (2010).
27. Poulin, R. Network analysis shining light on parasite ecology and diversity. *Trends Parasitol.* **26**, 492–498 (2010).
28. Bellay, S. et al. The patterns of organisation and structure of interactions in a fish–parasite network of a neoropical river. *Int. J. Parasitol.* **45**, 549–557 (2015).
29. Vazquez, 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).
30. Krasnov, B., Mouillot, D., Khokhlova, I., Shenbrot, G. I. & Poulin, R. Compositional and phylogenetic dissimilarity of host communities drives dissimilarity of ectoparasite assemblages: geographical variation and scale-dependence. *Parasitology* **139**, 338–347 (2012).
31. Morris, R. J., Gripenberg, S., Lweis, O. T. & Roslin, T. Antagonistic interaction networks are structured independently of latitude and host guild. *Ecol. Lett.* **17**, 340–349 (2014).
32. Maunsell, S. C., Kitching, R. L., Burwell, C. J. & Morris, R. J. Changes in host–parasitoid food web structure with elevation. *J. Animal Ecol.* **84**, 353–363 (2015).
33. Fortuna, M. A. et al. Coevolutionary dynamics shape the structure of bacteria–phage infection networks. *Evolution* **73**, 1001–1011 (2019).
34. Janzen, D. H. On ecological fitting. *Oikos* **45**, 308–310 (1985).
35. Agosta, S. J. & Klemens, J. A. Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecol. Lett.* **11**, 1123–1134 (2008).
36. Agosta, S. J. & Klemens, J. A. Resource specialization in a phytophagous insect: no evidence for genetically based performance trade-offs across hosts in the field or laboratory. *J. Evol. Biol.* **22**, 907–912 (2009).
37. Wells, K. & Clark, N. J. Host specificity in variable environments. *Trends Parsitol.* **35**, 452–465 (2019).
38. Golubski, A. J., Westlund, E. E., Vandermeer, J. & Pascual, M. Ecological networks over the edge: hypergraph trait-mediated indirect interaction (TMII) structure. *Trends Ecol. Evol.* **31**, 344–354 (2016).
39. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptionist programme. *Proc. R. Soc. B* **205**, 581–598 (1979).
40. Barrett, R. D. & Hoekstra, H. E. Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* **12** 767–780 (2011).
41. Valverde, S. et al. The architecture of mutualistic networks as an evolutionary spandrel. *Nat. Ecol. Evol.* **2**, 94–99 (2018).
42. Maynard, D. S., Serván, C. A. & Allesina, S. Network spandrels reflect ecological assembly. *Ecol. Lett.* **21**, 324–334 (2018).
43. Acevedo, M. A., Dillemuth, F. P., Flick, A., Faldyn, M. J. & Elderd, B. D. Virulence-driven trade-offs in disease transmission: a meta-analysis. *Evolution* **73**, 636–647 (2019).
44. Ashby, B., Gupta, S. & Buckling, A. Spatial structure mitigates fitness costs in host–parasite coevolution. *Am. Nat.* **183**, E64–E74 (2014).
45. Woolhouse, M. E. J. & Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842–1847 (2005).
46. Weiher, E. & Keddy, P. A. Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* **74**, 159–164 (1995).
47. Ulrich, W. & Gotelli, N. J. Null model analysis of species nestedness patterns. *Ecology* **88**, 1824–1831 (2007).
48. Bello, Fde The quest for trait convergence and divergence in community assembly: are null-models the magic wand? *Glob. Ecol. Biogeogr.* **21**, 312–317 (2012).
49. Feng, W. & Takemoto, K. Heterogeneity in ecological mutualistic networks dominantly determines community stability. *Sci. Rep.* **4**, 5912 (2014).
50. Jonhson, S., Domínguez-García, V. & Muñoz, M. A. Factors determining nestedness in complex networks. *PLoS ONE* **8**, e74025 (2013).
51. Pellissier, L. Stability and the competition-dispersal trade-off as drivers of speciation and biodiversity gradients. *Front. Ecol.* **3**, 00052 (2015).
52. Levins, R. & Culver, D. Regional coexistence of species and competition between rare species. *Proc. Natl Acad. Sci. USA* **68**, 1246–1248 (1971).
53. Fitzpatrick, B. M., Fordyce, J. A. & Gavrilets, S. What, if anything, is sympatric speciation? *J. Evol. Biol.* **21**, 1452–1459 (2008).
54. Flores, C. O., Valverde, S. & Weitz, J. S. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *ISME J.* **7**, 520–532 (2013).
55. Beckett, S. J. & Williams, H. T. P. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. *Interface Focus* **3**, 20130033 (2013).
56. May, R. M. Will a large complex system be stable? *Nature* **238**, 413–414 (1972).
57. Keyser, C. A., De Fine Licht, H. H., Steinwender, B. M. & Meyling, N. V. Diversity within the entomopathogenic fungal species *Metarhizium flavoviride* associated with agricultural crops in Denmark. *BMC Microbiol.* **15**, 249 (2015).
58. Büchi, L. & Vuilleumier, S. Ecological strategies in stable and disturbed environments depend on species specialization. *Oikos* **125**, 1408–1420 (2016).
59. Poisot, T. & Gravel, D. When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ* **2**, e251 (2014).
60. Nebreda, M. et al. Activity of aphids associated with lettuce and broccoli in Spain and their efficiency as vectors of lettuce mosaic virus. *Virus Res.* **100**, 83–88 (2004).
61. Nebreda, M., Michelena, J. M. & Fereres, A. Seasonal abundance of aphid species on lettuce crops in central Spain and identification of their main parasitoids. *J. Plant Dis. Prot.* **112**
62. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
63. Fournier-Level, A. et al. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86–89 (2011).
64. Hendrick, M. F. et al. The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone *Mimulus guttatus*. *Mol. Ecol.* **25**, 5647–5662 (2016).
65. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
66. Sacristán, S., Fraile, A. & García-Arenal, F. Population dynamics of cucumber mosaic virus in melon crops and in weeds in central Spain. *Phytopathology* **94**, 992–998 (2004).
67. MacArthur, R. H., & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
68. Newman, M. E. J. *Networks: An Introduction* (Oxford Univ. Press, 2010).
69. Susi, H., Vale, P. F. & Laine, A.-L. Host genotype and coinfection modify the relationship of within and between host transmission. *Am. Nat.* **186**, 252–263 (2015).
70. Atmar, W. & Patterson, B. D. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* **96**, 373–382 (1993).
71. Staniczenko, P. P. A., Kopp, J. C. & Allesina, S. The ghost of nestedness in ecological networks. *Nat. Commun.* **4**, 1391 (2013).
72. Wigner, E. P. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **62**, 548–564 (1955).
73. Nepusz, T., Petróczi, A., Négyessy, L. & Bazsó, F. Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E* **77**, 016107 (2008).
74. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
75. Trajanovski, H., Wang, H. & Van Mieghem, P. Maximum modular graphs. *Eur. Phys. J. B* **85**, 244 (2012).

Acknowledgements

The authors thank L. Alsedà, J. Sardanyés, T. Lázaro, S. Duran-Nebreda, N. Conde and R. Solé for useful comments and discussions. This work was supported by the Spanish Ministry of Economy and Competitiveness, grant FIS2016-77447-R MINEICO/AEI/FEDER and the European Union (to S.V.), and by grant RTI2018-094302-B-I00, Plan Estatal de I+D+i, Spanish Ministry of Economy and Competitiveness (to F.G.-A.). B.V. and R.M. were funded by the PR01018-EC-H2020-FET-Open MADONNA project.

Author contributions

S.V. and F.G.-A. designed and coordinated the study, S.V. conceived the theoretical framework and led its development. B.V. and S.V. contributed to the mean-field model. S.V., R.M. and B.V. developed the network analysis, S.S., A.F. and F.G.-A. collected field samples and the plant–virus network data. S.V. generated all the final illustrations and plots. S.V. and F.G.-A. wrote the manuscript. All authors made revisions and approved the final draft.

Season	N_p	N_h	N_t	C	N_t/N_h	$\langle k \rangle$	Q_c	$\langle B \rangle$	$\rho(A)$	z-score
Full	11	53	90	4	1.69	6.21	0.036	0.6775	1.6816	4.75
Summer	11	38	66	4	1.73	5,35	0.044	0.6049	2.7974	7.04
Spring	11	23	33	3	1.43	3,5	0.0775	0.4878	1.4047	2.76
Fall	10	26	36	4	1.38	3	0.0996	0.471	1.4844	1.48
Winter	9	20	24	3	1.2	2	0.1577	0.2908	1.486	1.37

Extended Data Fig. 1 | Summary of statistics for full and seasonal subnetworks. From left to right: pathogen richness N_p, host richness N_h, ecotype richness N_t, number of habitats C, host habitat specificity N_t/N_h, average species degree k, habitat modularity Q_c, average bridgeness B, spectral radius $\rho(A)$ (a quantitative measure of nestedness), and statistical significance of nestedness.

Habitat	Index (i)	N_i/N	N_p^i	N_t^i	Q_c^i
Full	0	1	11	90	0.036
Edge	1	0.46	11	36	0.0250442
Fallow Field	2	0.287129	11	18	0.00643848
Melon Field	3	0.207921	11	10	0.000642529
Wasteland	4	0.366337	11	26	0.003921

Extended Data Fig. 2 | Summary of statistics for habitat subnetworks. From left to right, relative fraction of species richness N_k/N, local ecotype richness N_p^k, local pathogen species richness N_t^k, and local habitat modularity Q_c^k. For reference, the first row gives the corresponding values in the full network ($N_0 = N$, $N_p^0 = N_p$, $N_t^0 = N_t$ and $Q_c^0 = Q_c = Q_c^1 + Q_c^2 + Q_c^3 + Q_c^4$).

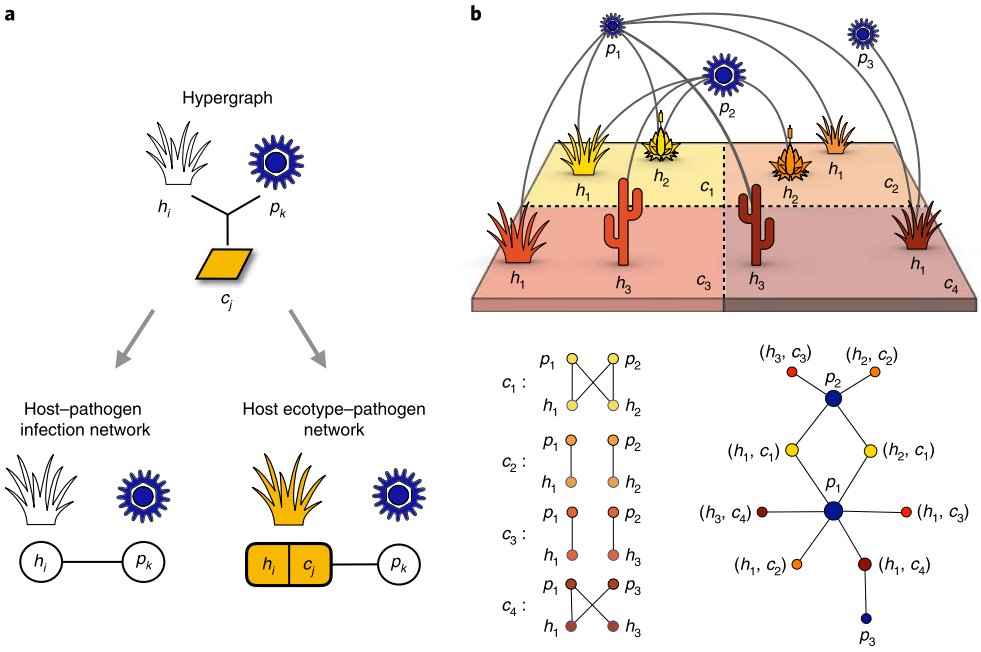


Fig. 1 | Hypergraph representation of environmentally mediated host–pathogen infections. a, Associations between host species (h_i), pathogen species (p_k) and habitats (c_j) are defined by hyperlinks. Networks depict interactions between pair of entities. From the hypergraph we can derive two networks. The HPIN represents interactions between pathogens and hosts within a single habitat, whereas host EPN is a decorated bipartite network replacing hosts with host ecotypes as the target of infection. A host ecotype is the natural association between host species and habitats. **b**, A landscape is divided into different habitats characterized by overlapping host assemblages. Below the schematic we compare the corresponding HPIN and EPN. A HPIN misrepresents pathogen ranges when host richness is different from ecotype richness, for example, pathogen $p_1 \in P$ is a generalist in habitat $c_4 \in C$ and a specialist in habitat $c_2 \in C$. EPN accurately represents how pathogen $p_1 \in P$ species can infect the same host species $h_1 \in H$ in multiple habitats. Colours indicate pathogens (blue) and the different habitats and node size is proportional to degree.

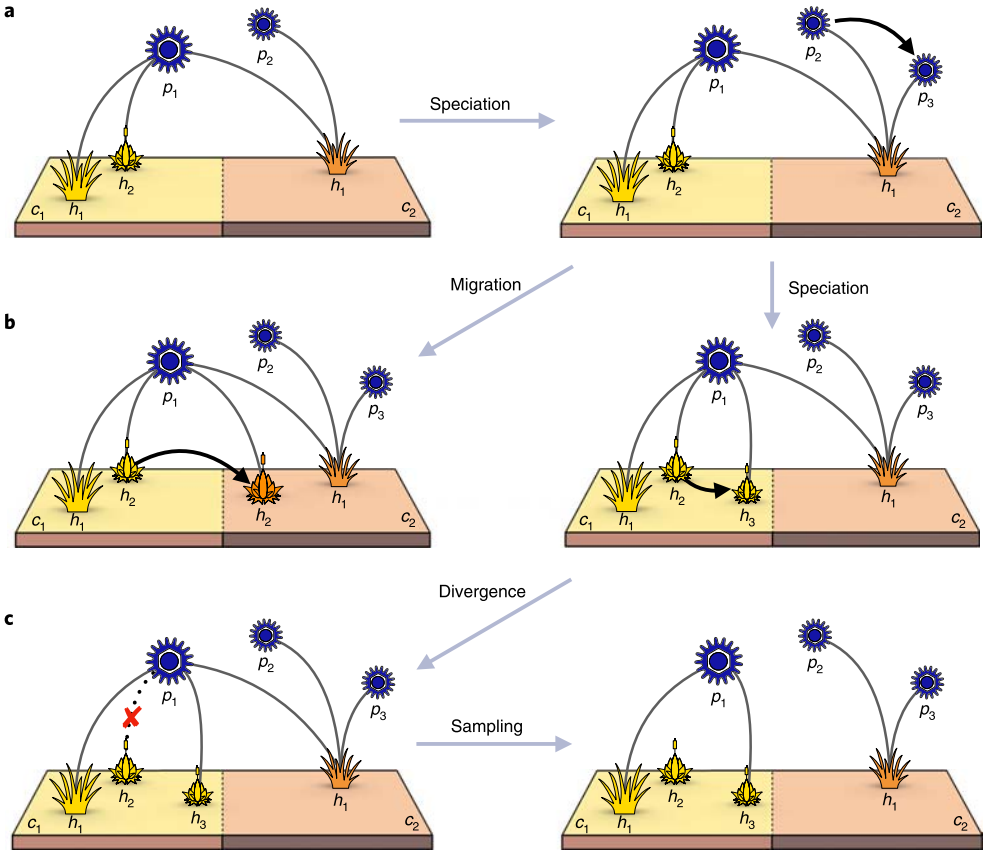


Fig. 2 | Neutral evolution of infection networks. a–c, A simple neutral model grows EPNs on the basis of duplication–divergence rules. **a**, We start from a small ecotype–pathogen network $G=(P,T,E)$ with $N_p = |P| = 2$ pathogens and $N_t = |T| = 3$ ecotypes (left). New descendant nodes inherit a subset of interactions from ancestors, that is, the new pathogen species $p_3 \in P$ duplicates the interactions of the mother species $p_2 \in P$ (right). **b**, We can generate the new ecotype $t_4 \in T$ from a previous ecotype $t_2=(h_2,c_1) \in T$ by 'migrating' the existing host to a different habitat $c_2 \in C$ (and thus $t_4=(h_2,c_2)$, left) or by 'speciating' a new host in the same habitat (and thus $t_4=(h_3,c_1)$, right). **c**, Subsequently, the network experiences a divergence caused by the random redistribution of link weights between the parent and the daughter species (left). At the ecological scale, we can only detect a fraction of the genetic interactions due to seasonality, spatial constraints and sampling effects (right). Overall, the neutrality hypothesis predicts a significant nestedness pattern but habitat subnetworks might not preserve it (for example, they can be modular).

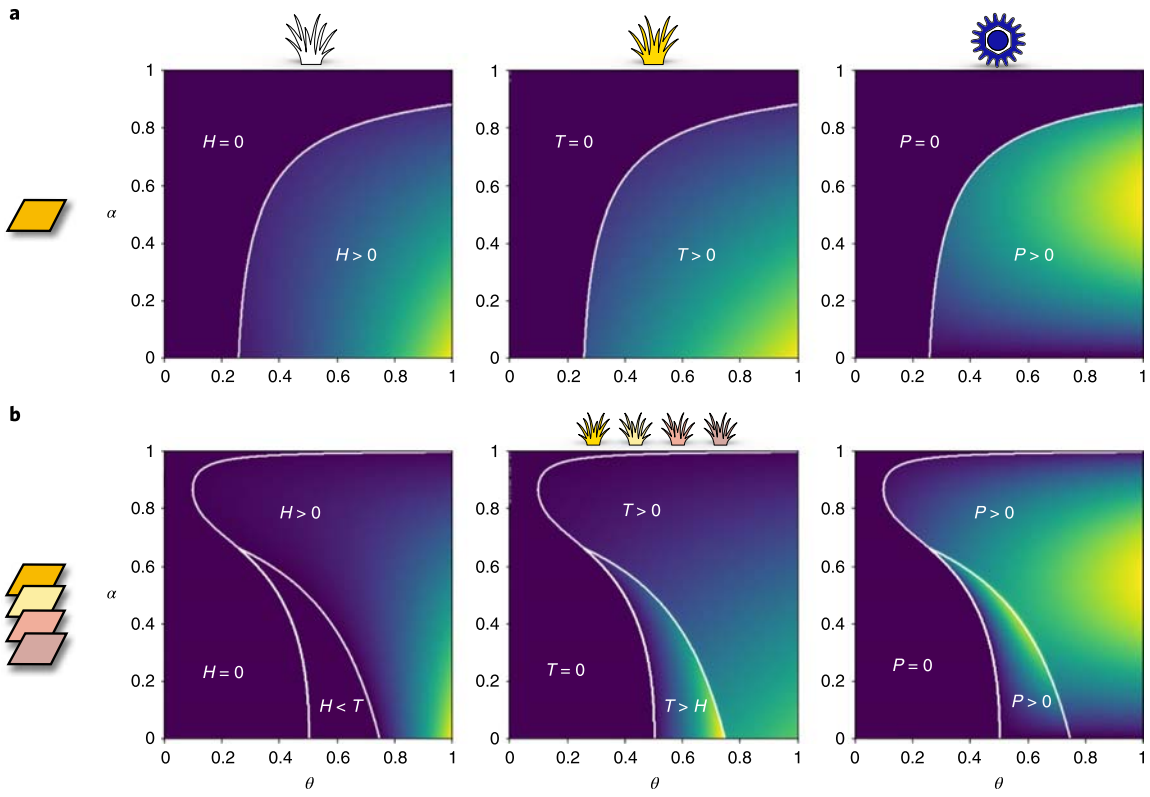


Fig. 3 | Emergence of host ecotypes depends on habitat diversity. a,**b**, Mean-field predictions for host species richness H (left), ecotype richness T (middle) and pathogen species richness P (right). **a**, There is little difference between host species richness and ecotype richness when there is only one habitat ($C=1$). White lines define the boundary between species coexistence (colours) and extinction (dark blue). Increasing pathogen species richness reduces host species richness (left). **b**, Emergence of ecotypes (middle) is a consequence of increasing the number of available habitats ($C=4$). Availability of space reduces the effective extinction rate. In this case, the system displays three different behaviours: extinction ($H=T=P\approx 0$), coexistence of ecotypes ($H<T$) and pathogens (middle), and coexistence of hosts ($H>T$) and pathogens (right). Bright colours indicate higher richness. Parameters: $\delta=1$.

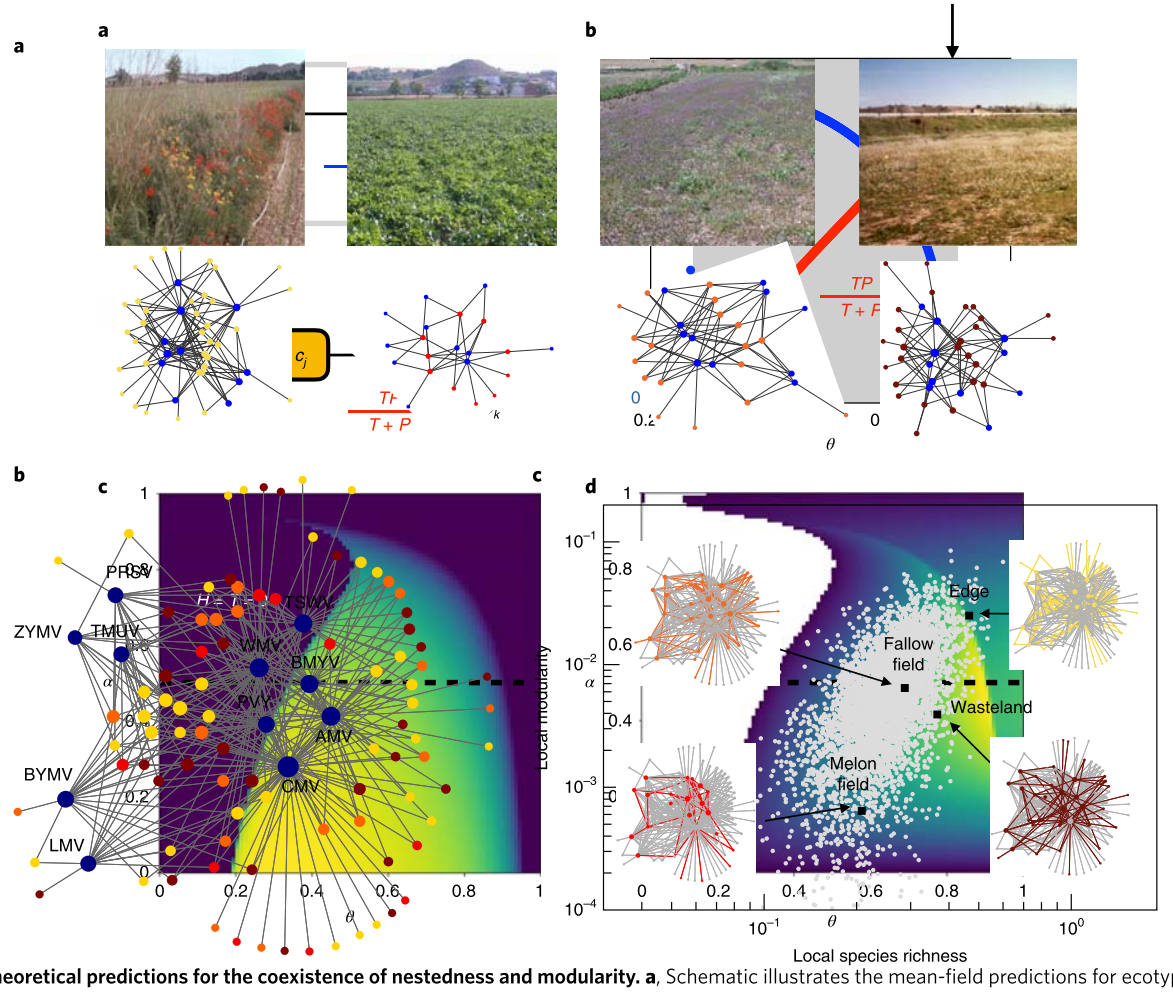


Fig. 4 | Theoretical predictions for the coexistence of nestedness and modularity. a, Schematic illustrates the mean-field predictions for ecotype–ecotype connectivity bounds as a function of number of habitats C, host species richness H, pathogen species richness P and ecotype richness T. Relationships between these variables can be used to map structural properties: $T/(H+C)$ is the average number of intra-habitat connections and $TP/(T+P)$ corresponds to the average degree of a fully connected ERN. **b**, Average intra-habitat connections (blue) and maximal average degree (red) as a function of intraspecific competition. The arrow points to the local maximum for the coexistence of nestedness and modularity (at $\alpha=0.5$). **c**,**d**, Numerical approximation to the mean-field model in the (α,θ) space for average intra-habitat connections (**c**) and maximal average degree (**d**). Dotted black line corresponds to the region where $\mu=0.5$. Parameters: $C=4$; $\delta=0.2$.

Fig. 5 | Modularity of plant–virus infections from 2000 to 2002 in an agricultural landscape in central Spain. a, ERN represents interactions between pathogens (blue nodes) and hosts (coloured nodes) in a single habitat. We can define four habitat subnetworks (from left to right: edge (yellow), melon field (red), fallow field (orange) and wasteland (brown)). This partitioning assumes that evolutionary and ecological dynamics of habitats are largely independent from each other. **b**, Full ERN merging habitat interactions is significantly nested (see Supplementary Discussion, section 4 for a detailed analysis). The modularity of this network is low, which invalidates the ERN assumption of independent habitats. In the network, blue nodes represent pathogens and host ecotype colour encodes the habitat where the host species was infected. AMV, alfalfa mosaic virus; BWYV, beet western yellow virus; BYMV, bean yellow mosaic virus; CMV, cucumber mosaic virus; LMV, lettuce mosaic virus; PRSV, papaya ringspot virus; PVY, potato virus Y; TSWV, tomato spotted wilt virus; TUMV, turnip mosaic virus; WMV, watermelon mosaic virus; ZYMV, zucchini yellow mosaic virus. **c**, Scaling of local habitat modularity (grey circles) with local species richness in the neutral model (see Supplementary Table 1 for full parameter values). Habitat nodes and links in the subgraphs have been highlighted using the same colours as in **a**. The neutrality hypothesis fits the fallow field subgraph, whereas the other habitat networks involve additional mechanisms (see text).

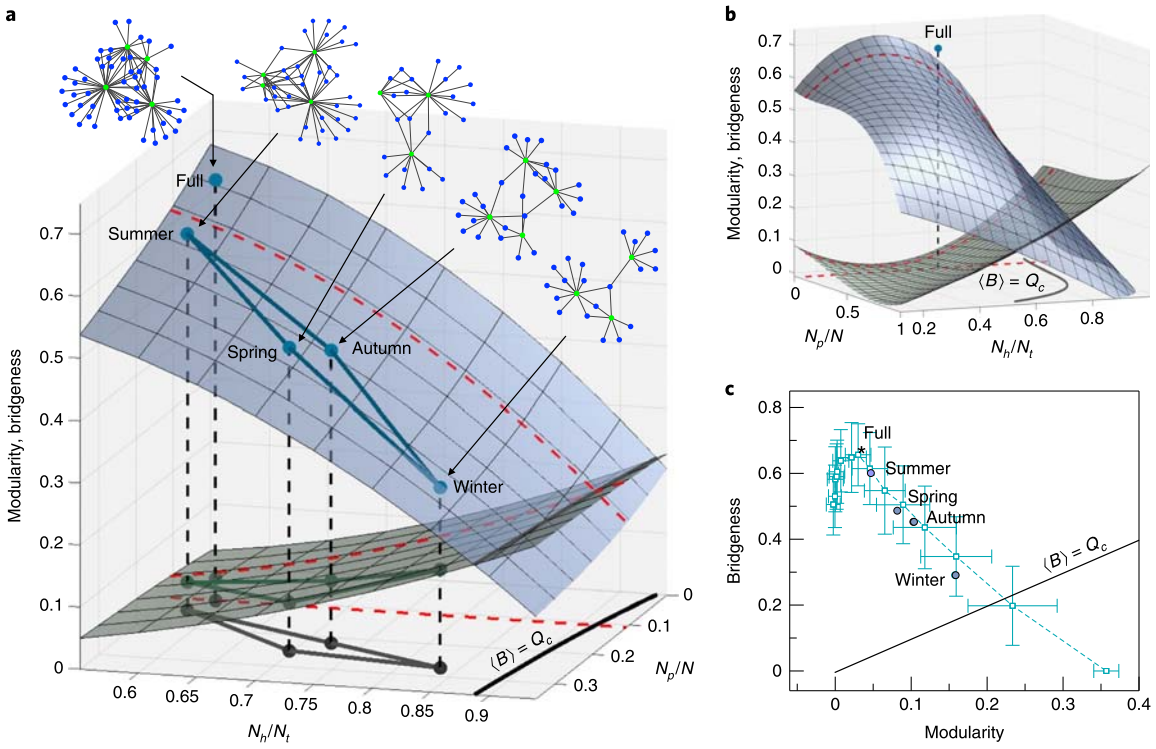
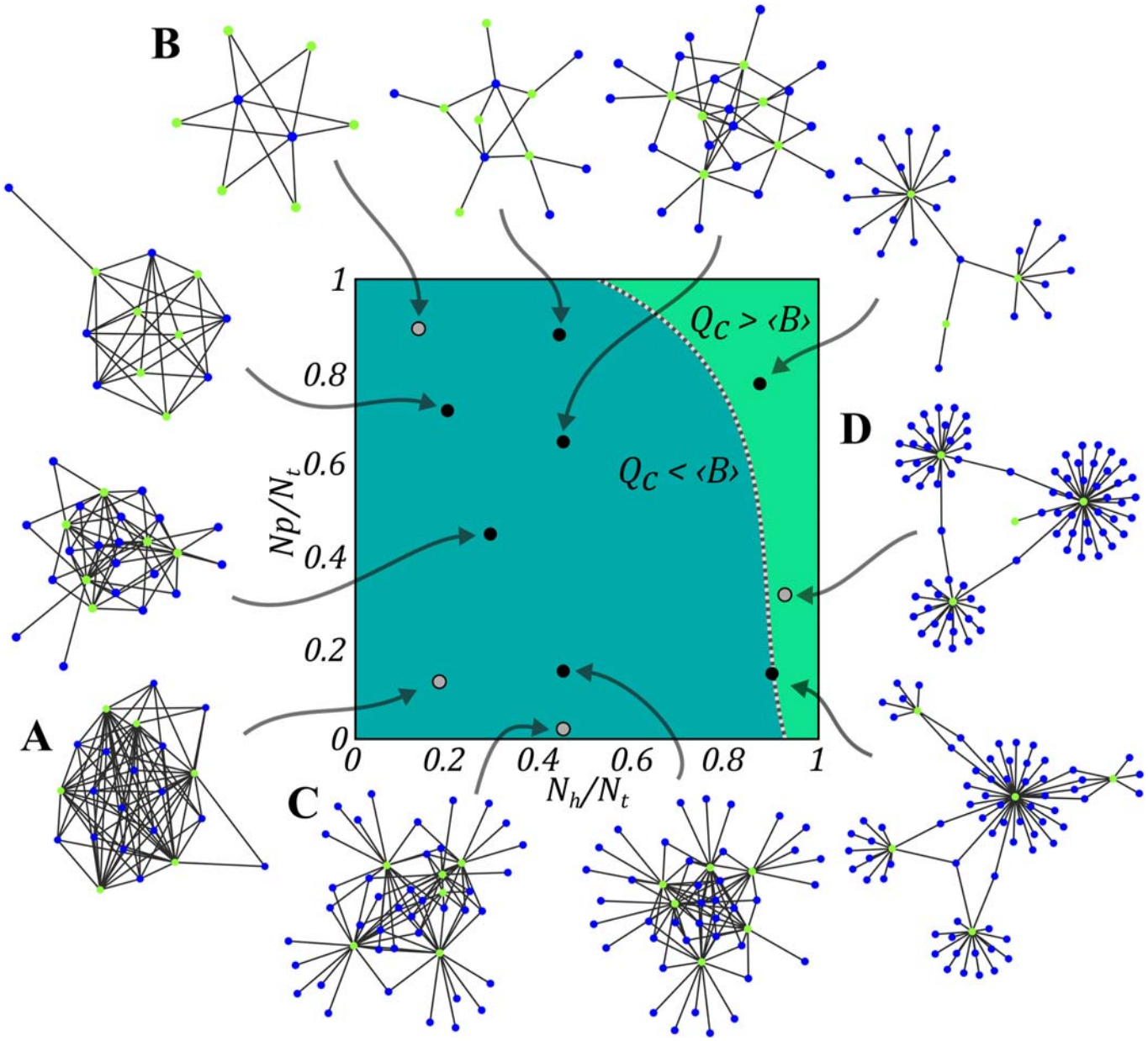


Fig. 6 | Seasonal variation of bridgeness and modularity in empirical EPNs. a, Expected modularity (grey surface) and average bridgeness (blue surface) in the subspace of seasonal networks. Modularity (average bridgeness) of EPNs increases (decreases) in colder seasons. This behaviour is consistent with the empirical host–habitat networks (top). **b**, Estimated location of the full network in the parameter space. Red dashed line corresponds to the estimated pathogen speciation rate $\alpha = 0.035$. The space of possible networks is far richer than observed empirical structures (see also Extended Data Fig. 3). Nestedness is a universal feature of this space, whose statistical significance is size-dependent. **c**, Seasonal trajectory of the system in the bridgeness–modularity space. Distance between predictions (dashed blue line) and data values (circles and asterisk) is within one standard deviation. Statistical significance decreases in winter because the neutral model underestimates seasonal intra-habitat competition (see Supplementary Discussion, section 5 for details). In all panels, intersection between modularity and bridgeness, that is, black line $\langle B\rangle = Q_c$, divides the space according to the relative ecotype richness (see text). See Supplementary Table 1 for full model parameters.



Extended Data Fig. 3 | Exploration of the theoretical morphospace of ecotype-pathogen networks. Nestedness is an invariant feature of this morphospace, while modularity is not and depends on speciation rates. This can be appreciated in the structure of networks generated for different combinations of pathogen and host speciation rates. The model predicts habitat modularity increases with habitat specificity. For example, (D) is a highly modular network, while (A) is not because only a few host species (blue nodes) in the former network share multiple habitats (green balls). The black line depicts the intersection of average bridgeness and habitat modularity, which separates low and intermediate modular networks (dark green region) from highly modular networks (light green) (see Supplementary Section 5 for details).