

The transition between the niche and neutral regimes in ecology

Charles K. Fisher and Pankaj Mehta¹

Department of Physics, Boston University, Boston, MA 02215

Edited by David Tilman, University of Minnesota, St. Paul, MN, and approved July 10, 2014 (received for review March 26, 2014)

An ongoing debate in ecology concerns the impacts of ecological drift and selection on community assembly. Here, we show that there is a transition in diverse ecological communities between a selection-dominated regime (the niche phase) and a drift-dominated regime (the neutral phase). Simulations and analytic arguments show that the niche phase is favored in communities with large population sizes and relatively constant environments, whereas the neutral phase is favored in communities with small population sizes and fluctuating environments. Our results demonstrate how apparently neutral populations may arise even in communities inhabited by species with varying traits.

neutral theory | disordered systems | phase transitions | niche theory

The success of the neutral theory of biodiversity and biogeography (1, 2) at explaining patterns in biodiversity has resulted in a vigorous debate on the processes underlying the assembly, dynamics, and structure of ecological communities (1, 3–12). Starting with the pioneering work of MacArthur (13–15), ecologists have emphasized the roles of interspecific competition and environmental interactions in community assembly and dynamics. These niche-based models emphasize ecological selection as the driving force of community assembly, whereas neutral models of biodiversity assume a functional equivalence between species and emphasize the role of ecological drift (i.e., stochasticity) in community dynamics (1, 2, 16, 17). The success of both types of models at explaining ecological data highlights the crucial need for understanding the impacts of ecological drift and selection in community ecology (18).

Hypothesis

We begin with a hypothesis that a diverse ecological community with many species can be either neutral or nonneutral, depending on the state of its environment. We call the regime in which a community is well described using neutral models the “neutral phase” and the regime in which the community behaviors are inconsistent with neutrality the “niche phase.” The dynamics in the neutral phase are dominated by stochasticity whereas the dynamics in the niche phase are dominated by selection. Our goal in this paper is to demonstrate that these two phases naturally emerge from simple probabilistic models of ecological dynamics and that a community may transition from one phase to the other as its environment is altered (Fig. 1).

Historically, ecological neutrality is based on the assumption of functional equivalence, which states that trophically similar species are essentially identical in terms of their vital characteristics, such as birth and death rates (19). Ecological neutrality, however, is generally not a measurable feature of a community. Therefore, we adopt a pragmatic definition of neutrality: We say that a community is “statistically neutral” if its multivariate distribution of species abundances cannot be distinguished from a distribution constructed under the assumption of ecological neutrality. In other words, the multivariate species abundance distributions of statistically neutral communities are indistinguishable from those of communities of functionally identical species. Note that ecological neutrality implies statistical neutrality, but statistical neutrality does not necessarily imply ecological neutrality. We

can now restate our hypothesis more precisely: As the characteristics of an ecosystem change (e.g., carrying capacity, immigration rate), there will be a transition between a neutral phase where the ecosystem behaves as if it is effectively neutral and a niche phase where the multivariate species abundance distribution is inconsistent with statistical neutrality.

Background on Phase Transitions in Disordered Systems

Our hypothesis that ecological systems are likely to exhibit multiple phases is based on an analogy with disordered systems in physics. For this reason, we briefly provide some background on phase transitions in disordered systems. A phase transition refers to an abrupt change in the qualitative behavior of a system as one of its characteristics, or a characteristic of its environment, is altered (20). The most well-known example may be the behavior of water, which can be found as a solid, liquid, or gas depending on temperature and pressure. Disordered systems often display a more complicated type of phase transition, labeled the freezing transition, where the system configuration gets “frozen” into a particular state.

One illustrative example of a disordered physical system is a protein (21). A protein can be thought of as a disordered system in which the different amino acids along the protein chain interact heterogeneously. The diversity of interactions in a protein distinguishes natural proteins from homopolymers and is what allows some proteins to fold to a stable native structure, while causing others (like prions or amyloids) to misfold. In ecology, this is analogous to the observation that the diversity of interactions between the species in a community distinguishes niche-like communities from neutral communities. To continue our analogy, at high temperatures, a typical polypeptide sequence will be in an unfolded phase where it samples different configurations randomly. If the temperature is lowered below a critical value, the polypeptide will freeze into a single structure (the folded state). This phase transition occurs when the stochasticity

Significance

In recent years, there has been a vigorous debate among ecologists over the merits of two contrasting models of biodiversity: the niche and neutral theories of ecology. Using two different theoretical models of ecological dynamics, we show that there is a transition between a selection-dominated regime (the niche phase) and a drift-dominated regime (the neutral phase). This is analogous to the phase diagram of water, which can be in the solid, liquid, or gas phase, depending on the temperature and pressure. Our results demonstrate how the niche and neutral theories both emerge from the same underlying ecological principles.

Author contributions: C.K.F. and P.M. designed research; C.K.F. performed research; C.K.F. analyzed data; and C.K.F. and P.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. Email: pankajm@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1405637111/-DCSupplemental.

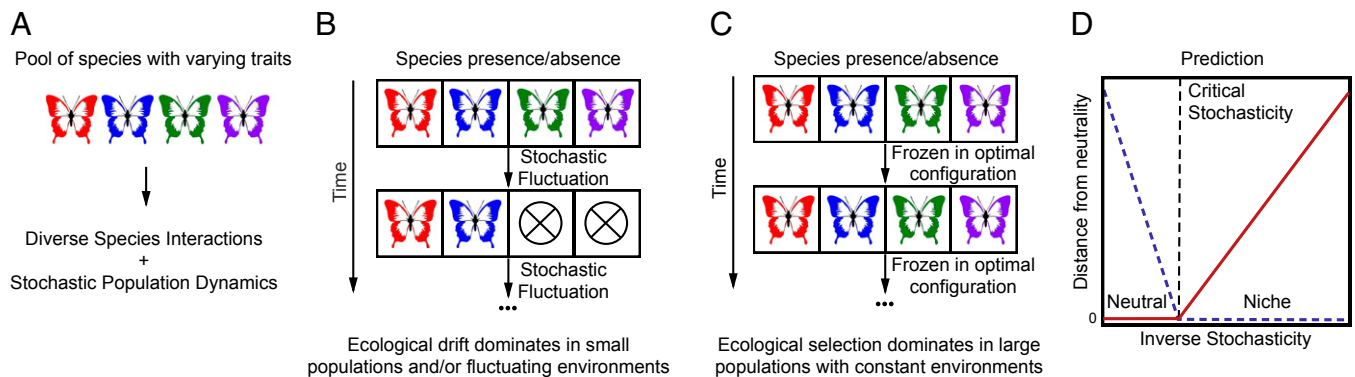


Fig. 1. A schematic illustrating the intuition underlying our hypothesis for a phase transition between the neutral and niche regimes in ecology. (A) The important ingredients of our model are a large pool of diverse species, implying a diversity of species interactions, subject to stochastic population dynamics. (B) Stochastic ecological drift will dominate the dynamics of communities with small population sizes and/or fluctuating environments. (C) By contrast, stabilizing selective forces will cause a community with a large population size and a constant environment to freeze into a unique, optimal configuration. (D) We predict that there is a transition between a drift-dominated (neutral) phase and a selection-dominated (niche) phase. That is, the community behaves exactly neutral when the inverse stochasticity is less than a critical threshold, and the deviation from neutrality rises quickly once the inverse stochasticity is larger than the critical threshold. The red line represents an order parameter based on the distance from neutrality, the dashed blue line represents an order parameter based on the niche phase, and the dashed black line denotes the critical stochasticity.

impacting the dynamics (i.e., the temperature) is smaller than the energetic differences in the interactions of the amino acids along the chain. If we take this analogy seriously, we should expect to find a critical amount of stochasticity, compared with the diversity of species traits, that separates neutral and niche communities.

Theoretical Models for Studying Community Assembly

To test our hypothesis regarding the niche-to-neutral transition, we analyzed two models of ecological dynamics: (i) a generalized Lotka–Volterra (LV) system including immigration and stochasticity and (ii) a binary model for the presence/absence (PA) of the species in a community. Each of these models has advantages and disadvantages. The LV model is a widely used and interpretable model of many ecological phenomena. However, in general, it is intractable to perform analytic calculations using the LV model and one must rely on numerical simulations. In contrast to the LV system, the PA model is amenable to analytical arguments but this comes at the expense of ignoring species abundances. Both models assume well-mixed populations, although relaxing this assumption is an important avenue of future research. These two models correspond to extreme cases of functional responses (22, 23). The functional response in the PA model is essentially a step function in which species interact only when their abundances are above a threshold. By contrast, the LV model corresponds to linear functional responses. We expect that real communities lie somewhere in between these models.

Parameterizing Ecosystem Characteristics

To construct ecological phase diagrams, it is necessary to parameterize ecosystem characteristics. Because we are interested in stochastic community assembly, we must introduce parameters that reflect the impact of stochasticity as well as parameters that capture variation in species traits. Due to the similarity of the two models, we use the same symbols for analogous parameters with an added tilde for parameters in the PA model (e.g., K denotes carrying capacity in the LV model and \tilde{K} denotes carrying capacity in the PA model).

There are two potential sources of stochasticity in the ecological dynamics: “demographic stochasticity” resulting from random births and deaths in small populations and “environmental stochasticity” caused by random variations in the environmental conditions. Although there is no doubt that the origin of the stochasticity is important for making quantitative ecological

predictions (24), extensive numerical simulations suggest that the qualitative phase diagrams are insensitive to these details (*SI Appendix*). For this reason, we parameterize the amount of stochasticity by a single parameter, the noise strength $\tilde{\omega}$.

We must also introduce parameters describing species traits. In principle, each species in the community has a unique immigration rate, a unique carrying capacity, and some set of parameters that describes how it interacts with other species. In the main text, we restrict ourselves to the case where all species have the same immigration rate, λ ($\tilde{\lambda}$), and the same carrying capacity K (\tilde{K}) (see *SI Appendix* for relaxation of these assumptions). We assume that the regional species pool is large so that there is a separation of timescales between the dynamics of the local community and those of the regional species pool. As a result our model does not explicitly include speciation, even though speciation is ultimately required to maintain diversity over longer timescales relevant to the species pool. Following May’s seminal work (25), we randomly draw symmetric interaction coefficients from a probability distribution and focus on describing the average behavior of ecosystems. Specifically, the interaction matrix C (\tilde{C})—with element c_{ij} (\tilde{c}_{ij}) characterizing the strength of interaction between species i and j —is drawn from a Gamma distribution with mean μ/S ($\tilde{\mu}/S$) and variance or “interaction diversity”, σ^2/S ($\tilde{\sigma}^2/S$), where S is the number of species (see *SI Appendix* for results with other distributions).

Stochastic Lotka–Volterra Dynamics

The first model that we analyze is a system of stochastic LV equations including immigration. Niche-based models of community assembly frequently use LV equations as a simplified description of ecological dynamics within a well-mixed community (13, 26–28). Here, we study a system of LV equations incorporating immigration and multiplicative noise (i.e., stochasticity). The rate of change in the abundance (x_i) of species $i = 1, \dots, S$ is

$$\frac{dx_i}{dt} = \lambda + x_i(K - x_i) - \sum_{j \neq i} c_{ij} x_i x_j + \sqrt{\omega x_i} \eta_i(t). \quad [1]$$

The first term (λ) is the rate that species i immigrates into the local community from an infinitely large regional species pool. The second term ($x_i(K - x_i)$) limits the population of species i to its carrying capacity (K) in the absence of immigration and

species interactions. The third term ($\sum_{j \neq i} c_{ij} x_j$) describes the effects that other species in the community have on species i according to their interaction coefficients (c_{ij}). All of these deterministic terms (i.e., λ , K , and c_{ij}) collectively represent the effects of ecological selection on the abundance of species i . Ecological drift is incorporated into our model through the last term ($\sqrt{\omega} \tilde{x}_i \eta_i(t)$), which represents stochasticity using a Gaussian “white noise” $\eta_i(t)$, with mean $\langle \eta_i(t) \rangle = 0$, variance $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$, and strength $\sqrt{x_i \omega}$.

Dynamics of Presence/Absence Model

In the PA model, a species i is described by a binary variable s_i with $s_i = 1$ if species i is present in a community and $s_i = 0$ if it is absent. The stochastic dynamics of species PA are defined by two rates: the rate at which a species immigrates into a community (i.e., the rate that $s_i = 0$ becomes $s_i = 1$) and the rate at which a species becomes extinct once it is in the community (i.e., the rate that $s_i = 1$ becomes $s_i = 0$). Thus, a species immigrates into a community and lives there for some time before it dies out, only to reimmigrate back into the community later, and so on. We assume that the rate of immigration is simply $\tilde{\lambda}$, and we model the rate of extinction as $\exp(-\tilde{K}(1 - \sum_j \tilde{c}_{ij} s_j) / \tilde{\omega})$. Therefore, in the absence of any interactions a species goes extinct at a rate that is exponentially slow in its carrying capacity $\exp(-\tilde{K} / \tilde{\omega})$, and competitive species interactions effectively decrease carrying capacity through $\tilde{K}(1 - \sum_j \tilde{c}_{ij} s_j)$ (13). The master equation describing the dynamics of \vec{s} with these rates is discussed in detail in *SI Appendix*. After an initial transient period, the community reaches a steady state where the immigration and extinction processes are balanced. Due to the simplicity of this model, we can derive an analytic expression for the steady-state probability distribution:

$$P_{PA}(\vec{s}) = \frac{\exp\left(\sum_i (\tilde{K} / \tilde{\omega} + \ln \tilde{\lambda}) s_i - (\tilde{K} / 2\tilde{\omega}) \sum_{(i,j)} \tilde{c}_{ij} s_i s_j\right)}{Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})}. \quad [2]$$

Here, $Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})$ is a normalizing constant such that the total probability sums to one.

Measuring the Neutrality of a Community

To test our hypothesis that communities can exhibit a niche-to-neutral transition, it is necessary to define “order parameters” that distinguish the niche and neutral phases. By convention, an order parameter is chosen so that it is zero in one phase and greater than zero in the other. Recall that the dynamics in the neutral phase are dominated by stochasticity and multivariate species abundance distributions in this phase are indistinguishable from those obtained from a neutral model with functionally equivalent species. By contrast, the niche phase is dominated by interactions and multivariate species abundance distributions are peaked around the equilibrium value they would have in the absence of stochasticity.

Using these intuitions we can define order parameters for both the LV model and the PA model. In the LV model, we define an order parameter that measures the distance (i.e., Kullback–Leibler divergence) between the multivariate species abundance distribution resulting from LV dynamics and the multivariate species abundance distribution resulting from purely neutral dynamics (see below). This order parameter is zero in the neutral phase and nonzero in the niche phase. For the PA model, it is convenient to consider a different order parameter, the Shannon entropy, of the steady-state PA probability distribution. The Shannon entropy is zero in the niche phase and nonzero in the neutral phase. We now discuss both of these order parameters in more detail.

Measuring Neutrality in LV Models. Early studies attempting to quantify the neutrality of a community focused on the shape of the marginal species abundance distribution, i.e., a histogram indicating the number of species with 10 individuals in the community, the number of species with 20 individuals in the community, and so on. However, recent studies have shown that both and neutral and nonneutral ecological models give rise to similar marginal species distributions (12). For this reason, to measure neutrality in the LV model we use the multivariate species abundance distribution. In contrast to previous studies on marginal species abundance distributions in niche and neutral communities (e.g., ref. 12), using the multivariate species abundance distribution allows us to study the effects of different species interactions on correlations in species abundances.

In particular, we quantify statistical neutrality in our LV simulations by measuring the distance between the steady-state distributions of species abundances obtained from the LV model ($P_{LV}(\vec{x})$) and purely neutral dynamics ($P_N(\vec{x})$). The measure of distance that we use is called the Kullback–Leibler divergence, $\text{KL}(P_{LV} \| P_N) = \int d\vec{x} P_{LV}(\vec{x}) \ln P_{LV}(\vec{x}) / P_N(\vec{x})$ (29). One interpretation of $\text{KL}(P_{LV} \| P_N)$ is defined as the amount of information about the true multivariate species abundance distribution [i.e., $P_{LV}(\vec{s})$] that is lost by approximating the distribution with one obtained from a neutral model [i.e., $P_N(\vec{x})$]. The KL divergence ranges from zero to infinity, with $\text{KL}(P_{LV} \| P_N) = 0$ implying that the simulated distribution is identical to the distribution obtained under the assumption of neutrality. We study the average of the KL divergence over many random realizations of the species interactions, i.e., $\langle \text{KL}(P_{LV} \| P_N) \rangle$. We expect $\langle \text{KL}(P_{LV} \| P_N) \rangle \sim 0$ in the neutral phase, whereas $\langle \text{KL}(P_{LV} \| P_N) \rangle \gg 0$ in the niche phase. Similar results are obtained with distance measures other than the KL divergence (*SI Appendix*).

In principle, it is possible to use an explicit formula for $P_N(\vec{x})$ from a specific neutral ecological model. However, many variations of neutral ecological models have been proposed and it is unclear which neutral model to use to calculate our order parameter. To circumvent this problem, we exploit the observation that the multivariate species abundance distributions of all neutral models share several features. Because we have restricted ourselves to considering LV systems where all species have the same immigration rate, we also restrict ourselves to considering neutral models where this assumption holds. The implications of nonuniform immigration rates are discussed in *SI Appendix*. With this caveat in mind, we observe that ecologically neutral models are also statistical neutral. Namely, the time-averaged moments of the abundance of species i are the same as the time-averaged moments of the abundance of species j . Moreover, the correlation in the abundances of species i and j is the same as the correlation in the abundances of species k and l (*SI Appendix*). Simulations shown in Fig. 2A demonstrate that this is the case, at least for Hubbell’s neutral model (*Materials and Methods*), where the KL divergence equals zero for all positive immigration rates. Finally, we note that although ecological neutrality implies statistical neutrality, statistical neutrality does not necessarily imply ecological neutrality. Thus, our use of statistical neutrality is consistent with the interpretation of ecological neutrality as a type of null model that allows one to identify communities in which selection is important.

Measuring Neutrality in the Presence/Absence Model. In the PA model, we do not have access to species abundances. For this reason, it is convenient to define a different order parameter that measures the fluctuations in the binary vector of community composition, \vec{s} : the entropy, $H[P_{PA}] = -\sum_{\vec{s}} P_{PA}(\vec{s}) \ln P_{PA}(\vec{s})$, of the steady-state probability distribution $P_{PA}(\vec{s})$. In the absence of stochasticity, \vec{s} will “freeze” into a unique configuration resulting from ecological selection and $H[P_{PA}] = 0$. In contrast, if the dynamics are entirely random, then each species will randomly

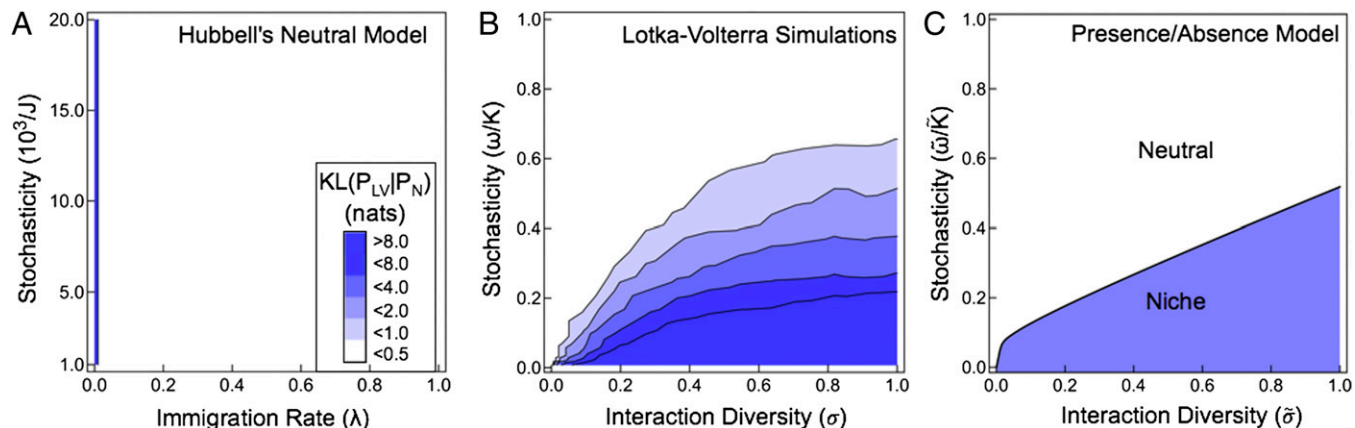


Fig. 2. Phase diagram of neutral and competitive ecosystems. (A) Communities simulated according to Hubbell's neutral model are statistically neutral with a KL divergence equal to zero for all positive immigration rates and community sizes (J). (B) Simulations of competitive LV communities with immigration display two phases: a statistically neutral phase with $\langle \text{KL}(P_{LV}||P_N) \rangle \sim 0$ and a niche phase with $\langle \text{KL}(P_{LV}||P_N) \rangle \gg 0$. Note that the colors represent exponential growth in the KL divergence. The critical stochasticity defining the phase boundary scales with interaction diversity (σ). Simulations were performed with $\mu = 1.0$ and $\lambda = 0.01$. (C) The phase diagram calculated from the presence/absence model has a statistically neutral phase and a niche phase and a phase boundary that scales with interaction diversity ($\tilde{\sigma}$). The phase diagram was calculated with $\tilde{\mu} = 1.0$ and $\tilde{\omega} \ln \tilde{\lambda} = K/2$.

flip between being absent and present in the community and $H[P_{PA}] = S \ln 2$. For diverse ecosystems with $S \gg 1$, we can define the boundary between the neutral phase and the niche phase as the points where $\langle H[P_{PA}]/S \rangle = 0$, with angular brackets denoting averaging over random realizations of interaction coefficients.

Phase Diagrams for Ecological Dynamics

Phase Diagrams. Armed with the order parameters discussed in the previous section, we can construct phase diagrams for both the LV and PA models. Fig. 2 shows the KL divergence and entropy as a function of stochasticity and interaction diversity for the two models. First, we note that the phase diagram determined using LV simulations is remarkably similar to the phase diagram calculated using our PA model (compare Fig. 2B and C), which suggests that our results are fairly robust to model details. Fig. 2B shows that there is a large neutral regime in which $\langle \text{KL}(P_{LV}||P_N) \rangle \sim 0$ in the LV simulations. The distance from neutrality rises once the stochasticity is lowered below a critical value. That is, $\langle \text{KL}(P_{LV}||P_N) \rangle$ increases for small ω/K ; note that the colors in Fig. 2B represent exponential growth in $\langle \text{KL}(P_{LV}||P_N) \rangle$.

Fig. 2C shows the phase diagram for the PA model. In the limit the number of species S becomes large, and the entropy is strictly zero in the niche phase (blue shaded area) and different from zero in the neutral phase (white area). In particular, we find that the PA of the species in a community freezes into a small number of configurations determined by the species traits if the stochasticity ($\tilde{\omega}$) is lowered below a critical value. This freezing is indicative of a phase transition from neutrality to niche-dominated ecological dynamics in the PA model.

It is important to recall that the neutral phase refers to a regime in which the multivariate species abundance distribution is well described by a neutral model, even if the underlying community is not ecologically neutral. For example, one special case that illustrates this relation is a community of species that have become so differentiated that they do not interact at all; i.e., $\mu = 0$ and $\sigma = 0$. Because there is no disorder, this community will reside in the statistically neutral phase even though the species are all highly differentiated. Nevertheless, the species in a community must be differentiated for the community to reside in the niche phase.

Scaling Relation for the Niche–Neutral Phase Boundary. We can explicitly calculate the phase boundary separating the niche and neutral phases, using the PA model. For diverse ecosystems with

many species $S \gg 1$, the relation defining the phase boundary can be derived by mapping the problem to the random energy model in physics (30, 31) (SI Appendix). Using this mapping we can derive a simple scaling relation that indicates when an ecological community will transition between the niche and neutral phases (SI Appendix):

$$\frac{\text{stochasticity}}{\text{carrying capacity}} \sim \frac{\text{immigration} \times \text{interaction diversity}}{\text{mean interaction strength}}.$$

The niche phase is favored when the interaction diversity is large relative to the impact of stochasticity on the dynamics of the population. By contrast, the neutral phase is favored when the interaction diversity is small relative to the impact of stochasticity on the dynamics of the population. This confirms the basic intuitions about ecological dynamics that were suggested by the analogy with protein folding discussed in the Introduction.

On the Nature of the Transition. The transition between the niche and neutral phases in the PA model is sharp (Fig. 3). In the LV model the distance from neutrality ($\langle \text{KL}(P_{LV}||P_N) \rangle$) increases when the stochasticity is lowered below the critical value. However, in the PA model, the derivative of the entropy with respect to stochasticity is undefined along the phase boundary, the signature of a freezing phase transition in the theory of disordered systems. Comparing the two models, the niche-to-neutral transition in the PA model appears to be sharper than in the LV model. Technically, a phase transition occurs at a point with an undefined derivative, whereas the term “crossover” is used to describe transitions between qualitative regimes without this feature. Our numerical results do not distinguish whether the transition in the LV model is a crossover or a true phase transition. This difference arises due to the differences in the functional responses of the two models. These two models were chosen, in part, because they represent the two extremes of possible species functional responses (linear vs. step function). We expect the functional responses of real ecological communities lie somewhere in between these two models. For this reason, we expect that real ecological communities will also exhibit a transition between the niche and neutral phases.

Ecological Implications

The Prevalence of Neutral and Niche Communities. Our model suggests that neutral communities and niche-like communities both

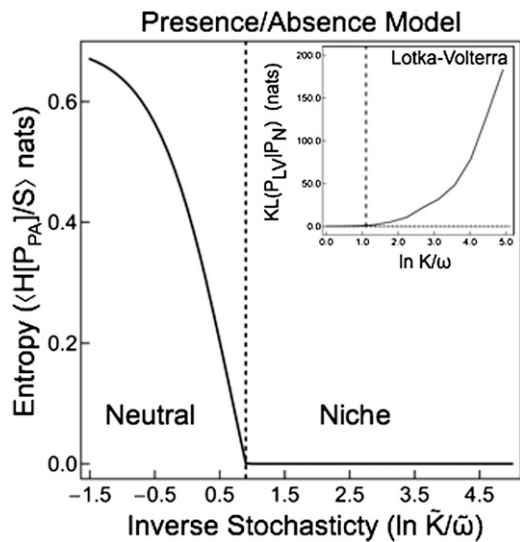


Fig. 3. The nature of the transition between the niche and neutral phases. Note that the order parameters for the two models are different: The order parameter for the PA model is zero in the niche phase and greater than zero in neutral phase, whereas the order parameter for the LV model is greater than zero in the niche phase and zero in the neutral phase. The average entropy $\langle H[P_{PA}/S] \rangle$, which is a measure of fluctuations in the community composition, is positive in the neutral phase and zero in the niche phase, illustrating the freezing transition in the PA model. (Inset) In LV simulations the distance from neutrality $\langle KL(P_{LV}/P_N) \rangle$ is essentially zero in the neutral phase and rises to large values in the niche phase. Parameters: $\bar{\mu} = 1.0$, $\bar{\sigma} \approx 0.4$, $\bar{\omega} \ln \lambda = 3K/2$, $\mu = 1.0$, $\sigma \sim 0.4$, $\lambda = 0.01$.

correspond to large volumes of the ecological phase diagram. Moreover, our model is such that species always have real differences in traits, but these differences in species traits leave no trace on the equilibrium multivariate species abundance distributions in the statistically neutral regime. This does not preclude the possibility that one could observe the effects of species trait variation on other types of observations. This may explain the success of neutral models at explaining many large-scale patterns in ecology, even though selective forces are well documented, and ubiquitous, on local scales. Furthermore, the crossover region surrounding the phase boundary corresponding to “nearly neutral”

communities occupies only a small volume of the phase diagram. As a result, we predict that nearly neutral communities should actually be quite rare, as long as there is not an external force (e.g., group selection) driving communities toward the niche-neutral boundary.

Ecological Disturbances. One of our main predictions is that the apparent neutrality of an ecological community is a function of both the inhabiting species and the environment. As a result, it is possible to drive a community between the niche and neutral phases by changing the environmental conditions. As an example, we consider the effects of selective logging on a population of butterflies in a tropical forest on Buru, Indonesia (32). Through habitat destruction, logging essentially moves the butterfly community from a position with high K/ω to one with low K/ω , tracing a path along the stochasticity axis in the phase diagram (Fig. 4A). Our model predicts that when a diverse community within the niche phase is placed under a stress that lowers K/ω to the critical value, it will undergo a transition to the neutral regime. LV simulations show that this transition results in a collapse of biodiversity and leads to an increase in the skewness of the species abundance distribution (SI Appendix). The increase in skewness of the species abundance distribution calculated from LV simulations is evident in a steeper curve in the rank-abundance plot for low K/ω compared with high K/ω (Fig. 4B). Similarly, the observed data display an increase in the skewness of the rank-abundance curve of the logged forest relative to the unlogged forest, consistent with a loss of biodiversity accompanying a niche-to-neutral transition (Fig. 4C). As our model predicts a transition in the shape of the rank-abundance curve as a function of increasing stochasticity (or decreasing carrying capacity), observations of rank-abundance curves as a function of deforestation, i.e., where the amount of deforestation varies continuously, could provide a more stringent test of our model than only the two endpoints discussed here. Nevertheless, this example demonstrates the potential of ecological phase diagrams for predicting the qualitative effects of community-wide disturbances and for capturing the characteristics that contribute to community resilience.

Conclusion

In summary, we have argued that the niche and neutral perspectives of ecology naturally emerge from stochastic models for the dynamics of diverse populations as distinct phases of an ecological community. Population dynamics in the niche phase

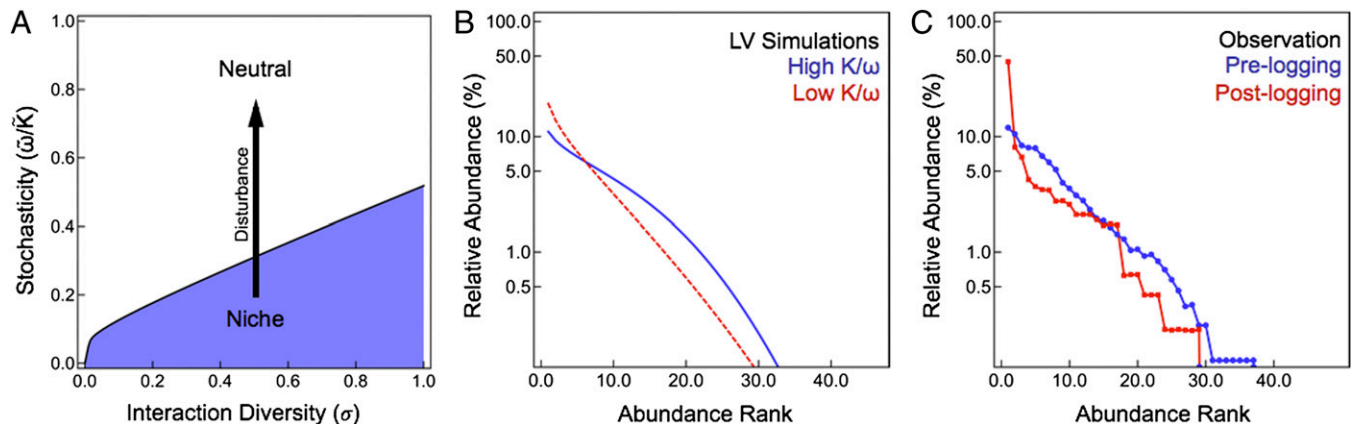


Fig. 4. Temporal variation in stochasticity and biodiversity in disturbed habitats. (A) An environmental disturbance that decreases carrying capacity may cause a community to shift from the niche phase to the neutral phase. (B) A community with a high carrying capacity ($K = 1.0$; blue) has a less skewed species abundance distribution than a community with a low carrying capacity ($K = 0.1$; red), as shown by the steeper red curve in the rank-abundance plot obtained from LV simulations. Simulations were run with $\mu = 1.0$, $\sigma = 0.5$, $\lambda = 0.02$, and $\omega = 0.6$. (C) Similarly, rank-abundance plots of butterfly species in a tropical Indonesian forest before (blue) and after (red) logging reflect an increase in skewness of the species abundance distribution following the disturbance (32).

are dominated by ecological selection, whereas population dynamics in the neutral phase are dominated by ecological drift. Furthermore, we have derived a simple scaling relation for determining whether an ecological community will be well described by neutral models.

Our hypothesis can be experimentally tested using synthetic microbial communities in which the immigration rates, carrying capacities, and interaction coefficients can be controlled to search for a transition as one moves from one region of the phase diagram to another (33). Alternatively, connections to island biogeography discussed in *SI Appendix* suggest that our hypothesis could be tested by calculating the KL divergence from the multivariate species abundance distributions on a chain of islands as a function of their distance to the mainland (34). Observation of a transition in the shape of the rank-abundance curve of a community along a disturbance gradient would also provide evidence of the niche–neutral transition.

In this work, we made some simplifications that are unrealistic for natural ecological communities. For example, we restricted our analysis to well-mixed communities with purely competitive interactions. It will be necessary to generalize our results to include the effects of dispersal, mutualism, predator–prey interactions, etc., to obtain a more quantitative model of natural communities. Nevertheless, we conjecture that the presence of

a niche–neutral phase transition is robust to these model perturbations. However, disordered systems with complex interactions display additional phases (35), which suggests that more complex ecological communities may also exhibit additional phases with novel characteristics.

Materials and Methods

We simulated Hubbell's neutral model with a local community of J individuals connected to an infinitely large metacommunity containing $S = 50$ equally abundant species. In each time step, with probability λ , an individual randomly drawn from the metacommunity replaced a randomly chosen individual in the local community, or with probability $1 - \lambda$, one randomly chosen individual in the local community replaced another randomly chosen individual in the local community. The simulations were run for 5×10^7 steps. Ten simulations were run for each set of parameters, and the results were averaged.

LV simulations with $S = 50$ species were performed over the parameter ranges specified in Figs. 1–4. In each case, the competition coefficients were sampled randomly, and then the stochastic Lotka–Volterra equations (Eq. 1) were forward integrated for 5×10^7 steps of size $\delta t = 0.005$, using the Milstein method. Ten simulations were run for each set of parameters, and the results were averaged.

ACKNOWLEDGMENTS. We thank Jeff Gore, Alex Lang, Javad Noorbakhsh, Daniel Segre, and Les Kauffman for useful discussion and the Alfred Sloan Foundation for funding.

- Rosindell J, Hubbell SP, He F, Harmon LJ, Etienne RS (2012) The case for ecological neutral theory. *Trends Ecol Evol* 27(4):203–208.
- Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (MPB-32) (Princeton Univ Press, Princeton).
- Ricklefs RE, Renner SS (2012) Global correlations in tropical tree species richness and abundance reject neutrality. *Science* 335(6067):464–467.
- Wootton JT (2005) Field parameterization and experimental test of the neutral theory of biodiversity. *Nature* 433(7023):309–312.
- McGill BJ (2003) A test of the unified neutral theory of biodiversity. *Nature* 422(6934):881–885.
- Ricklefs RE (2006) The unified neutral theory of biodiversity: Do the numbers add up? *Ecology* 87(6):1424–1431.
- Dornelas M, Connolly SR, Hughes TP (2006) Coral reef diversity refutes the neutral theory of biodiversity. *Nature* 440(7080):80–82.
- Jeraldo P, et al. (2012) Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proc Natl Acad Sci USA* 109(25):9692–9698.
- Tilman D (2004) Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proc Natl Acad Sci USA* 101(30):10854–10861.
- Volkov I, Banavar JR, Hubbell SP, Maritan A (2009) Inferring species interactions in tropical forests. *Proc Natl Acad Sci USA* 106(33):13854–13859.
- Haegeman B, Loreau M (2011) A mathematical synthesis of niche and neutral theories in community ecology. *J Theor Biol* 269(1):150–165.
- Chisholm RA, Pacala SW (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proc Natl Acad Sci USA* 107(36):15821–15825.
- MacArthur R, Levins R (1967) The limiting similarity, convergence, and divergence of coexisting species. *Am Nat* 101:377–385.
- Chase JM, Leibold MA (2003) *Ecological Niches: Linking Classical and Contemporary Approaches* (Univ of Chicago Press, Chicago).
- Tilman D (1982) *Resource Competition and Community Structure* (Princeton Univ Press, Princeton), Vol 17.
- Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424(6952):1035–1037.
- Azaele S, Pigolotti S, Banavar JR, Maritan A (2006) Dynamical evolution of ecosystems. *Nature* 444(7121):926–928.
- Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85(2):183–206.
- Hubbell SP (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Funct Ecol* 19(1):166–172.
- Kadanoff LP (2000) *Statistical Physics: Statics, Dynamics and Renormalization* (World Scientific, Singapore).
- Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA* 84(21):7524–7528.
- Holling CS (1959) Some characteristics of simple types of predation and parasitism. *Can Entomol* 91:385–398.
- Abrams P (1987) The functional responses of adaptive consumers of two resources. *Theor Popul Biol* 32:262–288.
- Lande R, Engen S, Sæther BE (2003) *Stochastic Population Dynamics in Ecology and Conservation* (Oxford Univ Press, Oxford).
- May RM (1972) Will a large complex system be stable? *Nature* 238(5364):413–414.
- MacArthur R (1970) Species packing and competitive equilibrium for many species. *Theor Popul Biol* 1(1):1–11.
- Chesson P (1990) MacArthur's consumer–resource model. *Theor Popul Biol* 37(1):26–38.
- Chesson P, Kuang JJ (2008) The interaction between predation and competition. *Nature* 456(7219):235–238.
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86.
- Derrida B (1980) Random-energy model: Limit of a family of disordered models. *Phys Rev Lett* 45(2):79–82.
- Derrida B (1981) Random-energy model: An exactly solvable model of disordered systems. *Phys Rev B* 24:2613–2626.
- Hill JK, Hamer KC, Lacey LA, Banham WMT (1995) Effects of selective logging on tropical forest butterflies on Buru, Indonesia. *J Appl Ecol* 32:754–760.
- Hekstra DR, Leibler S (2012) Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* 149(5):1164–1173.
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17(4):373–387.
- Kirkpatrick S, Sherrington D (1978) Infinite-ranged models of spin-glasses. *Phys Rev B* 17:4384–4403.

Supporting Information

Fisher and Mehta 10.1073/pnas.1405637111

SI Appendix

Presence/Absence Model

Dynamics: Master Equation. In this section, we introduce a phenomenological model describing the probability of observing various combinations of species in a local community, which we assume is attached to a large regional species pool. The presence (or absence) of species $i \in \{1, \dots, S\}$ is described by a binary variable s_i , with $s_i = 1$ if the species is present and $s_i = 0$ if it is absent. The probability of observing a particular set of present/absent species \vec{s} at time t is described by a probability distribution $P_t(\vec{s})$. The probability distribution is described by a differential equation called a master equation.

The master equation describing the time evolution of $P_t(\vec{s})$ requires us to specify two types of rates: the rate of immigration at which $s_i = 0 \rightarrow s_i = 1$ and the rate of extinction at which $s_i = 1 \rightarrow s_i = 0$. The rate of immigration of species i , $R_i^I(\vec{s})$, is given by

$$R_i^I(\vec{s}) = \tilde{\lambda}_i.$$

The rate of extinction of species i , $R_i^E(\vec{s})$, is given by

$$R_i^E(\vec{s}) = \exp\left(-\frac{1}{\omega} \left(\tilde{K}_i - \sum_j \tilde{K}_j \tilde{c}_{ij} s_j \right)\right).$$

Here, $\tilde{\lambda}_i$ is the rate of immigration of species i , \tilde{K}_i is the carrying capacity of species i , \tilde{c}_{ij} is an interaction coefficient that describes the influence that species j has on species i , and ω describes the impact of stochasticity on species extinction. Throughout this article we use a convention in which $\tilde{c}_{ii} = 0$ for the presence/absence model and $c_{ii} = 0$ for Lotka–Volterra (LV) models. The units of time have been set to the rate of extinction in the limit that $\omega \rightarrow \infty$.

The time evolution of the probability distribution is described by a master equation:

$$\begin{aligned} \frac{dP_t(\vec{s})}{dt} = \sum_i \left\{ \left(R_i^E(\vec{s} + \vec{e}_i) P_t(\vec{s} + \vec{e}_i) - R_i^I(\vec{s}) P_t(\vec{s}) \right) (1 - s_i) \right. \\ \left. + \left(R_i^I(\vec{s} - \vec{e}_i) P_t(\vec{s} - \vec{e}_i) - R_i^E(\vec{s}) P_t(\vec{s}) \right) s_i \right\}. \end{aligned} \quad [\text{S1}]$$

Here, \vec{e}_i is a vector with element $i = 1$ and all other elements equal to zero.

The first line of Eq. S1 describes the rate of change in the probability of a community with $s_i = 0$. The positive term, $R_i^E(\vec{s} + \vec{e}_i) P_t(\vec{s} + \vec{e}_i)$, reflects the process by which the probability of this state increases due to extinction events where $s_i = 1 \rightarrow s_i = 0$. The negative term, $R_i^I(\vec{s}) P_t(\vec{s})$, reflects the process by which the probability of this state decreases due to immigration events where $s_i = 0 \rightarrow s_i = 1$.

The second line of Eq. S1 describes the rate of change in the probability of a community with $s_i = 1$. The positive term, $R_i^I(\vec{s} - \vec{e}_i) P_t(\vec{s} - \vec{e}_i)$, reflects the process by which the probability of this state increases due to immigration events where $s_i = 0 \rightarrow s_i = 1$. The negative term, $R_i^E(\vec{s}) P_t(\vec{s})$, reflects the process by which the probability of this state decreases due to extinction events where $s_i = 1 \rightarrow s_i = 0$.

There is no general method for solving this equation, although it can be simulated using Gillespie's algorithm (1). Below, we show that an equilibrium solution for $P_t(\vec{s})$ can be obtained in the limit that $t \rightarrow \infty$.

Connection to MacArthur–Levins. Before describing the solution for the equilibrium distribution of Eq. S1 we first explain the intuition for the rates of our phenomenological presence/absence model. First, it is helpful to recount the famous results on species invasion in Lotka–Volterra communities due to MacArthur and Levins (2).

Deterministic models of ecological dynamics typically take the form $dx_i/dt = \lambda_i + x_i f_i(\vec{x})$. Here, λ_i is the rate of immigration and $f_i(\vec{x})$ is the ecological fitness of species i , which is a function of the population \vec{x} . The LV equations describe communities with linear ecological fitness $f_i(\vec{x}) = K_i - x_i - \sum_j c_{ij} x_j$. In general, $f_i(\vec{x})$ may be a complicated function due to nonlinear functional responses or other phenomena. Regardless of the exact form of $f_i(\vec{x})$, the ecological fitness can always be linearized near an equilibrium point, \vec{x}^* , in which case the ecological dynamics are approximately described by LV equations.

MacArthur and Levins (2) considered an equilibrium point (\vec{x}^*) where $x_i \sim 0$ and the linearized ecological fitness of species i is $f_i(\vec{x}^*) \approx K_i - \sum_j c_{ij} x_j^*$. If a small number of species i attempt to invade the community, they will be successful if $K_i - \sum_j c_{ij} x_j^* > 0$ but will be unsuccessful if $K_i - \sum_j c_{ij} x_j^* < 0$. If the community has only two species, then $x_j^* = K_j$ and the relationships are $K_i - K_j c_{ij} > 0$ for successful invasion and $K_i - K_j c_{ij} < 0$ for unsuccessful invasion.

The qualitative results should be similar in the case of a stochastic LV dynamics. In the absence of species i the species abundances fluctuate about their mean values (\vec{x}^*). If a small number of species i attempt to invade the community, they are most likely to be successful if $K_i - \sum_j c_{ij} x_j^* > 0$, i.e., the mean extinction time will be long, but will be unlikely to be successful if $K_i - \sum_j c_{ij} x_j^* < 0$, i.e., the mean extinction time will be short.

Now, let us take a look at our proposed immigration and extinction rates. That the rate at which an absent species with $s_i = 0$ becomes a present species with $s_i = 1$ is simply the rate of immigration $\tilde{\lambda}_i$ is straightforward. In the case of extinction, the rate of extinction is controlled by $\tilde{K}_i - \sum_j \tilde{K}_j \tilde{c}_{ij} s_j$, for fixed ω . This can be compared with $K_i - \sum_j c_{ij} x_j^*$, which controls the rate of extinction in the MacArthur–Levins model. If we make a conjecture that, all other things being equal, x_j^* scales proportionally with K_j , then we can write $x_j^* = K_j \gamma_j c_{ij} s_j$ where γ_j is simply the constant of proportionality describing the scaling of x_j^* with K_j . Finally, we simply define $\tilde{c}_{ij} = \gamma_j c_{ij}$ to arrive at our proposed extinction rate. The behavior is such that if $\tilde{K}_i - \sum_j \tilde{K}_j \tilde{c}_{ij} s_j \gg 0$, then the rate at which species i goes extinct in the local community will be very slow, whereas if $\tilde{K}_i - \sum_j \tilde{K}_j \tilde{c}_{ij} s_j \ll 0$, it will be fast. This behavior is qualitatively similar to the MacArthur and Levins analysis of species invasion in LV communities (2).

Equilibrium Distribution. In general, there is no way of solving Eq. S1 to find $P_t(\vec{s})$. However, it is possible to solve for the steady-state equilibrium distribution $P_{\text{PA}}(\vec{s})$ (PA, presence/absence), i.e., the distribution such that $dP_t/dt = 0$, if we impose the following symmetry conditions: $\tilde{\lambda}_i = \tilde{\lambda}$, $\tilde{K}_i = \tilde{K}$, and $\tilde{c}_{ij} = \tilde{c}_{ji}$. The symmetry conditions ensure that transition rates satisfy the principle of detailed balance (Fig. S1).

Mathematically, it is clear from Eq. S1 that if we can find a distribution $P_{\text{PA}}(\vec{s})$ such that $R_i^E(\vec{s} + \vec{e}_i) P_{\text{PA}}(\vec{s} + \vec{e}_i) = R_i^I(\vec{s}) P_{\text{PA}}(\vec{s})$

if $s_i = 0$ and $R_i^E(\vec{s} - \vec{e}_i)P_{PA}(\vec{s} - \vec{e}_i) = R_i^E(\vec{s})P_{PA}(\vec{s})$ if $s_i = 1$, then dP/dt evaluated at $P_{PA}(\vec{s})$ equals zero. Thus, $P_{PA}(\vec{s})$ is a fixed point of the master equation. Moreover, it is possible to show that this fixed point is unique.

This equilibrium distribution is given by

$$P_{PA}(\vec{s}) = \frac{\exp\left(\sum_i (\tilde{K}/\tilde{\omega} + \ln \tilde{\lambda})s_i - (\tilde{K}/2\tilde{\omega})\sum_{(i,j)} \tilde{c}_{ij}s_i s_j\right)}{Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})}. \quad [\text{S2}]$$

Here, $Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})$ is a normalizing constant such that the total probability sums to one. It is straightforward to verify the detailed balance relations, e.g., $R_i^E(\vec{s} + \vec{e}_i)P_{PA}(\vec{s} + \vec{e}_i) = R_i^E(\vec{s})P_{PA}(\vec{s})$, by plugging in Eqs. S1 and S2. For example,

$$\begin{aligned} R_i^E(\vec{s} + \vec{e}_i)P_{PA}(\vec{s} + \vec{e}_i) &= R_i^E(\vec{s})P_{PA}(\vec{s}) \\ \Rightarrow \\ \exp\left(-(\tilde{K} - \sum_j \tilde{K}\tilde{c}_{ij}s_j)/\tilde{\omega}\right) \exp\left(\sum_i (\tilde{K}/\tilde{\omega} + \ln \tilde{\lambda})s_i - (\tilde{K}/2\tilde{\omega})\sum_{(i,j)} \tilde{c}_{ij}s_i s_j\right) \\ &= \frac{\tilde{\lambda} \exp\left(\sum_{j \neq i} (\tilde{K}/\tilde{\omega} + \ln \tilde{\lambda})s_j - (\tilde{K}/2\tilde{\omega})\sum_{(k \neq i,j)} \tilde{c}_{kj}s_k s_j\right)}{Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})} \\ \Rightarrow \\ \exp\left(-\frac{(\tilde{K} - \sum_j \tilde{K}\tilde{c}_{ij}s_j)}{\tilde{\omega}} + \sum_i \left(\frac{\tilde{K}}{\tilde{\omega}} + \ln \tilde{\lambda}\right)s_i - \frac{\tilde{K}}{2\tilde{\omega}} \sum_{(i,j)} \tilde{c}_{ij}s_i s_j\right) \\ &= \exp\left(\ln \tilde{\lambda} + \sum_{j \neq i} \left(\frac{\tilde{K}}{\tilde{\omega}} + \ln \tilde{\lambda}\right)s_j - \frac{\tilde{K}}{2\tilde{\omega}} \sum_{(k \neq i,j)} \tilde{c}_{kj}s_k s_j\right) \\ \Rightarrow \\ \exp\left(\ln \tilde{\lambda} + \sum_{j \neq i} \left(\frac{\tilde{K}}{\tilde{\omega}} + \ln \tilde{\lambda}\right)s_j - \frac{\tilde{K}}{2\tilde{\omega}} \sum_{(k \neq i,j)} \tilde{c}_{kj}s_k s_j\right) \\ &= \exp\left(\ln \tilde{\lambda} + \sum_{j \neq i} \left(\frac{\tilde{K}}{\tilde{\omega}} + \ln \tilde{\lambda}\right)s_j - \frac{\tilde{K}}{2\tilde{\omega}} \sum_{(k \neq i,j)} \tilde{c}_{kj}s_k s_j\right). \end{aligned}$$

Generality of the Presence/Absence Model. Our analysis of our presence/absence model in this work depends only on the properties of the steady-state distribution, $\tilde{P}_{PA}(\vec{s})$. Under our symmetry assumptions $\tilde{P}_{PA}(\vec{s}) = P_{PA}(\vec{s})$, although this is not the case in general. Nevertheless, we can generally expand the logarithm of the steady-state distribution in terms of its moments as

$$\ln \tilde{P}_{PA}(\vec{s}) = \ln Z + \sum_i h_i s_i + \sum_{ij} J_{ij} s_i s_j + \sum_{i,j,k} J_{ijk} s_i s_j s_k + \dots,$$

where the h s and J s are appropriately chosen coefficients. Thus, one way to view our analysis is that we have taken such a series and truncated it after the pairwise term; although there are some subtleties that arise in using statistical mechanics to analyze a steady-state distribution rather than an equilibrium distribution. Moreover, the form of the equilibrium distribution that we analyze (Eq. S2) can be viewed as a statistical maximum entropy model, also known as binary Markov random fields in the statistics and machine-learning literature (3–6). These statistical modeling approaches have been used to model phenomena across many

different fields. The appearance of phase transitions in these types of models is quite general.

Quenched Disorder and the Average Free Energy. For a particular choice of $\tilde{\lambda}$, \tilde{K} , $\tilde{\omega}$, and \tilde{c}_{ij} , the distribution of the presence/absence of the species in a community is described by the equilibrium probability distribution

$$P_{PA}(\vec{s}) = \frac{\exp\left(\sum_i (\tilde{K}/\tilde{\omega} + \ln \tilde{\lambda})s_i - (\tilde{K}/2\tilde{\omega})\sum_{(i,j)} \tilde{c}_{ij}s_i s_j\right)}{Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})},$$

where $Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})$ is a normalizing constant such that the total probability sums to one. However, in this work we are not interested in understanding the behavior of the system for a particular choice of parameters, but we instead want to understand the average behavior of the system over many random realizations of the parameters. We approach this problem by studying the average behavior of $\ln Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})$ in the limit that the number of species S goes to infinity.

In physics, the logarithm of the normalization constant of an equilibrium distribution, e.g., $\ln Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})$, is called the free energy of the system. The free energy is an important, and useful, quantity because its derivatives provide the moments of the distribution (i.e., average, variance, etc.). Phase transitions correspond to points where one of the derivatives of the free energy (often with respect to stochasticity) is undefined (7).

There are two ways that one could compute the average free energy: the “annealed” average equal to $\ln \langle Z \rangle$ or the “quenched” average given by $\langle \ln Z \rangle$. Jensen’s inequality immediately implies that $\ln \langle Z \rangle \geq \langle \ln Z \rangle$. Analyzing the quenched average of the free energy corresponds to analyzing the properties of typical communities for which the species interactions are drawn from a given probability distribution. Thus, our goal is to calculate the quenched average of $\ln Z(\tilde{\lambda}, \tilde{K}, \tilde{C}, \tilde{\omega})$.

In principle, the quenched average free energy can be calculated using “the replica trick” (8):

$$\lim_{S \rightarrow \infty} \frac{1}{S} \langle \ln Z \rangle = \lim_{S \rightarrow \infty} \frac{1}{S} \lim_{n \rightarrow \infty} \frac{\langle Z^n \rangle - 1}{n}.$$

However, it is sufficient for our purposes to calculate the quenched average free energy using a random energy model approximation (9–11).

Deriving the Phase Diagram with the Random Energy Model. Just to be precise, the goal of our random energy model calculation is to compute the quenched average free energy,

$$\lim_{S \rightarrow \infty} \frac{1}{S} \langle F \rangle = \lim_{S \rightarrow \infty} \frac{1}{S} \langle \ln Z \rangle,$$

where the angular brackets denote an average over random realizations of the species interactions. The definition of free energy in thermodynamics is $F = U - TH$, where U is the “internal energy” (described below), H is the “entropy,” and T is the “temperature.” For a discrete probability distribution, the entropy can be calculated using the Shannon–Gibbs formula $H = -\sum_i p_i \ln p_i$, where p_i is the probability of the i th state. That is, the entropy is a measure of fluctuations. We show that there is a freezing transition in the random energy model, where

$$\lim_{S \rightarrow \infty} \frac{1}{S} \langle H \rangle = 0.$$

The calculation follows that in refs. 9–11.

The random energy approximation. First, we define a function called the internal energy (or just the energy) as

$$U(\vec{s}) = - \sum_i (1 + \tilde{\Lambda}) s_i + \frac{1}{2} \sum_{(i,j)} \tilde{c}_{ij} s_i s_j, \quad [\text{S3}]$$

where $\tilde{\Lambda} = \tilde{\omega} \ln \tilde{\lambda} / \tilde{K}$ is called the immigration potential. This allows us to rewrite the equilibrium distribution as $P_{\text{PA}}(\vec{s}) = Z^{-1} \exp(-\tilde{K}U(\vec{s})/\tilde{\omega})$. So far, we have done nothing except rewrite the equilibrium distribution, i.e., Eq. S2, in a different form that is the typical convention in physics with the identification of temperature $T = \tilde{\omega} / \tilde{K}$.

Each of the interaction coefficients, \tilde{c}_{ij} , is treated as an independent random variable with mean $\langle \tilde{c}_{ij} \rangle = \tilde{\mu} / S$ and finite variance $\langle \tilde{c}_{ij}^2 \rangle - \langle \tilde{c}_{ij} \rangle^2 = \tilde{\sigma}^2 / S$. For example, the interaction coefficients could be randomly drawn from a Gamma distribution as in the LV simulations presented in the main text. Note, however, that the precise form of the distribution does not matter for the random energy model calculation. Each of the interaction coefficients in the energy is a random variable and, as a result, the energy itself is also a random variable. In fact, because the energy is a sum of many independent and identically distributed random variables with finite variances, we know that it must be approximately normally distributed according to the central limit theorem. Let $M = \sum_i s_i$ be the number of species present in the community. Because we are interested in the limit where S is large, we keep only terms $O(S)$. The moments of Eq. S3 are

$$\langle U \rangle \simeq -(\tilde{\Lambda} + 1)M + \frac{1}{2} \tilde{\mu} \frac{M^2}{S} \quad [\text{S4}]$$

and

$$\langle U^2 \rangle - \langle U \rangle^2 \simeq \frac{1}{4} \tilde{\sigma}^2 \frac{M^2}{S}. \quad [\text{S5}]$$

It is helpful to simplify our notation by introducing related quantities that do not scale with the number of species, S . These “intensive” quantities are

$$m = \frac{M}{S} \quad [\text{S6}]$$

$$u(m) = \frac{\langle U \rangle}{S} = -(\tilde{\Lambda} + 1)m + \frac{1}{2} \tilde{\mu} m^2 \quad [\text{S7}]$$

$$u'(m) = \frac{du}{dm} = -(\tilde{\Lambda} + 1) + \tilde{\mu} m$$

$$v(m) = \frac{\langle U^2 \rangle - \langle U \rangle^2}{S} = \frac{1}{4} \tilde{\sigma}^2 m^2$$

$$v'(m) = \frac{dv}{dm} = \frac{1}{2} \tilde{\sigma}^2 m.$$

Here, $m = M/S$ is the “species saturation”; i.e., the fraction of species in the regional species pool that are present in the community.

The energies are approximately normally distributed, so they have a probability density function given by

$$f(U|m) = \frac{1}{\sqrt{2\pi S v(m)}} \exp\left(-\frac{(U - S u(m))^2}{2 S v(m)}\right). \quad [\text{S8}]$$

The approximation in using the random energy model to derive the phase diagram for our presence/absence model comes from an assumption that all of the energies are independently distrib-

uted. In reality, the correlation between the energies of two configurations of the community, $\vec{s}^{(1)}$ and $\vec{s}^{(2)}$, is a function of their “overlap,” $\vec{s}^{(1)} \cdot \vec{s}^{(2)}$. These correlations decay as the order of the species interactions is increased; e.g., the correlation is smaller if we include interactions between three species than if we consider only pairwise interactions. Thus, our approximation is to neglect these correlations even though we have included only pairwise species interactions. This approximation has been used extensively to model protein folding.

The entropy. The next step in the calculation is to compute the average number of energy levels, $\langle n(U) \rangle$ in the region $[U, U + dU]$. This is given by

$$\langle n(U) \rangle = \sum_{M=0}^{M=S} \binom{S}{M} f\left(U \mid \frac{M}{S}\right). \quad [\text{S9}]$$

Now, we take the limit of a large number of species. When S is large, the binomial coefficient becomes peaked around a maximum value and the sum is dominated by a single term. We take the logarithm of Eq. S9, using Stirling’s approximation (e.g., $\ln S! \sim S \ln S$), drop all terms $O(\ln S)$, and pull the S out front,

$$\ln \langle n(U) \rangle = S \max_{0 < m < 1} \left\{ -m \ln m - (1 - m) \ln(1 - m) - \frac{(u - u(m))^2}{2v(m)} \right\}, \quad [\text{S10}]$$

where $u = U/S$. The maximization over m results from the statement that a single term dominates the sum in Eq. S9.

The Boltzmann formula for the microcanonical entropy, $H(U)$ tells us that $H(U) = \ln \langle n(U) \rangle$. Combining this with the thermodynamic relation

$$\frac{\partial H}{\partial U} = \frac{1}{T} = \frac{\tilde{K}}{\tilde{\omega}} \quad [\text{S11}]$$

allows us to determine the free energy. Specifically, we have

$$\frac{\partial H}{\partial U} = -\frac{(u - u(m))}{v(m)} = \frac{1}{T}. \quad [\text{S12}]$$

Solving for u gives $u = u(m) - v(m)/T$. Plugging this result into Eq. S10, we obtain the entropy

$$\frac{H(m)}{S} = -m \ln m - (1 - m) \ln(1 - m) - \frac{1}{2} \frac{v(m)}{T^2} \quad [\text{S13}]$$

and the free energy

$$\frac{F(m)}{S} = u(m) - \frac{1}{2} \frac{v(m)}{T} + \frac{1}{T} (m \ln m + (1 - m) \ln(1 - m)). \quad [\text{S14}]$$

The typical species saturation m is found by minimizing the free energy.

The freezing transition. The entropy in the random energy model (Eq. S13) is a sum of two terms: $-m \ln m - (1 - m) \ln(1 - m) \geq 0$ and $-\frac{1}{2} v(m) \leq 0$. Thus, there is a possibility that $H(m)/S \leq 0$ if the variance in the energies (that is the variance in the species interactions) is large enough: $v(m) \geq 2(-m \ln m - (1 - m) \ln(1 - m))$. However, the thermodynamic entropy is a nonnegative quantity, so the line $H(m)/S = 0$ defines a “freezing” phase transition. The phase boundary can be parameterized as a critical temperature:

$$T_c = \sqrt{\frac{v(m)}{2(-m \ln m - (1 - m) \ln(1 - m))}}. \quad [\text{S15}]$$

Above this critical temperature (which is stochasticity/carrying capacity) the presence/absence of the various species in the community fluctuates randomly. Below this critical temperature the presence/absence of the various species in the community freezes into a particular configuration determined by the species traits. Now, we derive the scaling relation presented in the main text from Eq. S15.

Scaling Relation for the Phase Boundary. A simple scaling argument captures the main features of the freezing phase boundary. Before proceeding, it is useful to find a simple scaling for the magnetization (species saturation) m . Differentiating Eq. S14 with respect to m and setting the equation equal to zero yields

$$m = \frac{1}{2} \left(1 + \tanh \left(-\frac{1}{2} \frac{1}{T} u'(m) + \frac{1}{4} \frac{1}{T^2} v'(m) \right) \right).$$

For large $\tilde{\mu}$, we find that $m \sim \tilde{\Lambda}/\tilde{\mu}$. We can combine this relation to derive a scaling relation for the niche-to-neutral phase boundary.

To do so, note from Eq. S11 that we can define a critical stochasticity

$$T_c = \frac{\tilde{\omega}_c}{\tilde{K}}. \quad [\text{S16}]$$

On the other hand, because $v(m) \sim \sigma^2 m^2$ and $-m \log m - (1-m) \log(1-m)$ is order 1, Eq. S15 yields the scaling relation

$$T_c \sim \tilde{\sigma} m.$$

Setting the two equations above equal to each other yields

$$\frac{\tilde{\omega}}{\tilde{K}} \sim \frac{\tilde{\sigma} \times \tilde{\Lambda}}{\tilde{\mu}}$$

or, in words,

$$\frac{\text{stochasticity}}{\text{carrying capacity}} \sim \frac{\text{immigration} \times \text{interaction diversity}}{\text{mean interaction strength}}.$$

This result is illustrated in Fig. S2. In both numerical simulations and analytic calculations, the niche phase is favored when the carrying capacity and interaction diversity are high. By contrast, the neutral phase is favored at low carrying capacities, high stochasticity, and low interaction diversity. These data illustrate that selection dominates community assembly in the niche phase, whereas ecological drift dominates community assembly in the statistically neutral phase. Immigration shifts the phase boundary between the two phases, with high rates of immigration increasing the size of the niche phase relative to the neutral phase.

Freezing in Lotka–Volterra Communities. The Lotka–Volterra model that we study in the main text (and below) is also a stochastic model of population dynamics. It is important to be clear about the effect that the freezing transition has on these dynamics. The abundances of the species in the community always stochastically fluctuate—even in the frozen (niche) phase. However, if the fluctuations in the abundance of a species $\langle x_i^2 \rangle - \langle x_i \rangle^2$ are much smaller than its mean abundance $\langle x_i \rangle$, then the presence/absence of that species is essentially frozen. These effects are clear even when comparing species abundance distributions across different levels of stochasticity in a single-species LV model (Fig. S3), although there are important differences between a community with many species and a community with only one species.

In a diverse community with many species, the niche-to-neutral transition coincides with increase in skewness of the marginal

species abundance distribution as a result of the “unfreezing” of the species presences/absences (Fig. S4). When the stochasticity increases beyond a critical value, the marginal species abundance distribution transitions from a distribution peaked around K to one peaked at 0. As a result, the niche-to-neutral transition is accompanied by a loss of biodiversity when the stabilizing selective forces are dominated by stochastic effects.

Measuring Neutrality from Abundance Distributions

To test our hypothesis that communities can exhibit a sharp niche-to-neutral phase transition, it is necessary to define an “order parameter” that distinguishes the niche and neutral phases. By convention, an order parameter is chosen so that it is zero in one phase and greater than zero in the other. Recall that the dynamics in the neutral phase are dominated by stochasticity and species abundance distributions in this phase are indistinguishable from those obtained from a neutral model with functionally equivalent species. By contrast, the niche phase is dominated by interactions and species abundance distributions are peaked around the equilibrium value they would have in the absence of stochasticity.

We quantify statistical neutrality in our LV simulations by measuring the distance between the steady-state distributions of species abundances obtained from the LV model ($P_{LV}(\vec{x})$) and purely neutral dynamics ($P_N(\vec{x})$). The measure of distance that we use in the main text is called the Kullback–Leibler divergence, $\text{KL}(P_{LV}||P_N) = \int d\vec{x} P_{LV}(\vec{x}) \ln P_{LV}(\vec{x})/P_N(\vec{x})$ (12). One interpretation of $\text{KL}(P_{LV}||P_N)$ is defined as the amount of information about the true multivariate species abundance distribution [i.e., $P_{LV}(\vec{s})$] that is lost by approximating the distribution with one obtained from a neutral model [i.e., $P_N(\vec{x})$]. The KL divergence ranges from zero to infinity, with $\text{KL}(P_{LV}||P_N) = 0$ implying that the simulated distribution is identical to the distribution obtained under the assumption of neutrality. We study the average of the KL divergence over many random realizations of the species interactions, i.e., $\langle \text{KL}(P_{LV}||P_N) \rangle$. We expect $\langle \text{KL}(P_{LV}||P_N) \rangle \sim 0$ in the neutral phase, whereas $\langle \text{KL}(P_{LV}||P_N) \rangle \gg 0$ in the niche phase.

The KL divergence is not a true distance metric because it is asymmetric and does not satisfy the triangle inequality. There are many alternative measures of discrepancy between probability distributions. One example is the Jensen–Shannon divergence (JSD) given by (13)

$$\text{JSD}(P_{LV}, P_N) = \frac{1}{2} \left(\text{KL} \left(P_{LV} \parallel \frac{1}{2} (P_{LV} + P_N) \right) + \text{KL} \left(P_N \parallel \frac{1}{2} (P_{LV} + P_N) \right) \right).$$

The JSD is symmetric and bounded by $0 \leq \text{JSD}(P_{LV}, P_N) \leq 1$, and although it does not directly satisfy the triangle inequality, its square root ($\sqrt{\text{JSD}(P_{LV}, P_N)}$) does (14). To demonstrate that our results are not very sensitive to our choice of metric, Fig. S5 illustrates the phase diagram and the sharpness of the niche-neutral transition, using $\text{JSD}(P_{LV}, P_N)$ instead of $\text{KL}(P_{LV}||P_N)$.

Discussion on Unequal Immigration Rates and Carrying Capacities

In the main text and in our presence/absence model calculations, we have assumed that all of the species in the community have equal immigration rates and carrying capacities so that species interactions are the only source of variation. The assumption of equal immigration rates is equivalent to assuming that all species are equally abundant in the regional species pool. Of course, this is not generally the case; for example, Hubbell’s neutral model results in a regional species pool with abundances that are distributed according to a log-series distribution (15, 16). Additional

simulations discussed below demonstrate that small deviations from the assumption of equal immigration rates and carrying capacities do not affect our main conclusions (Fig. S7). However, it is important to extend our results to apply in the more general case. In this section, we provide a brief sketch of how our results can be generalized, but an extensive treatment of the subject is left for future work.

The order parameter that we proposed for measuring neutrality based on multivariate species abundance distributions, i.e., $\text{KL}(P_{\text{LV}}||P_{\text{N}}) = \int d\vec{x} P_{\text{LV}}(\vec{x}) \ln P_{\text{LV}}(\vec{x})/P_{\text{N}}(\vec{x})$, is applicable regardless of parameters of the model. In principle, one could explicitly calculate $P_{\text{N}}(\vec{x})$ from a specific neutral model (15–18). However, because the parameters of the Lotka–Volterra system and neutral models are different, it is not entirely clear to which neutral distribution the comparison should be made. One potential option for quantifying statistical neutrality in communities with unequal immigration rates is to use the fact that, in a neutral model, the Pearson correlation coefficient between the abundances of species i and j is equal to the correlation coefficient between the abundances of species k and l regardless of the immigration rates. Thus, a distribution with equal correlation coefficients can stand in as a proxy for measuring statistical neutrality.

Alternatively, we can consider the effects of unequal immigration rates or carrying capacities on the phase diagram of species presence/absence. In general, the stochastic process describing the dynamics of the presence/absence of the species in a community in which the species have unequal immigration rates (or carrying capacities) will not satisfy detailed balance and, thus, never reaches equilibrium. Nevertheless, we can typically assume that there is a steady-state distribution $P_{\text{PA}}(\vec{s})$ defined as

$$\tilde{P}_{\text{PA}}(\vec{s}) = \lim_{t \rightarrow \infty} P_{\text{PA}}(\vec{s}; t).$$

In general, we can expand the logarithm of the steady-state distribution in terms of its moments as

$$\ln \tilde{P}_{\text{PA}}(\vec{s}) = \ln Z + \sum_i h_i s_i + \sum_{ij} J_{ij} s_i s_j + \sum_{ijk} J_{ijk} s_i s_j s_k + \dots$$

Thus, we can define something like an “energy” given by

$$U(\vec{s}) = \sum_i h_i s_i + \sum_{ij} J_{ij} s_i s_j + \sum_{ijk} J_{ijk} s_i s_j s_k + \dots$$

and proceed with a calculation analogous to the random energy model. To do so, we have only to relate the means and variances (and potentially covariances) of the h s and J s to the distributions of λ_i , K_i , and c_{ij} and to ω . This is a difficult problem to solve exactly, but we can conjecture on the general behavior. For example, we expect that the mean of h monotonically increases with the mean immigration rate and the mean carrying capacities, but monotonically decreases with increasing ω . Likewise, the variance of h should monotonically increase with the variance in the immigration rates and the variance in the carrying capacities, but should decrease monotonically with increasing ω . Relations of this form provide a general ansatz for studying complex communities and can be verified numerically. This is an important avenue for future work.

Additional Lotka–Volterra Simulations

The qualitative conclusions of our paper hold for a diverse choice of distributions for the LV interaction coefficients. In this section, we present simulations of the LV system (Eq. S17) in which the parameters (K_i , λ_i , and c_{ij}) are quenched random variables drawn from various distributions.

Low-Rank Interaction Matrix. Low-rank approximations are often used for inferring matrices because they are less sensitive to noise and overfitting than their full-rank counterparts. Here, we consider a low-rank interaction matrix $c_{ij} = S^{-1} g_i g_j$, where the g_i are independent, Beta-distributed random variables on the interval (0, 1) with mean $\langle g_i \rangle = \mu$ and variance $\langle g_i^2 \rangle - \langle g_i \rangle^2 = \mu(1-\mu)/(1+\nu)$. With this parameterization $\langle c_{ij} \rangle = S^{-1} \mu^2$ and $\langle c_{ij}^2 \rangle - \langle c_{ij} \rangle^2 \approx S^{-2} 2\mu^3(1-\mu)/(1+\nu)$. Note that the interaction coefficients (c_{ij}) are not independent as in the main text, but are actually correlated. The phase diagrams, as a function of μ , obtained from LV simulations and PA calculations with the low-rank interaction matrix are shown in Fig. S6.

A similar phase diagram is obtained even if we perturb λ_i , K_i , and c_{ij} with multiplicative noise drawn from $\mathcal{N}(1, \epsilon)$. (Fig. S7). In this case, the interaction matrix is not symmetric and not necessarily strictly competitive. Moreover, the immigration rates and carrying capacities are no longer equal; that is, $\lambda_i \neq \lambda_j$ and $K_i \neq K_j$.

Exponentially Distributed Interaction Coefficients. In Fig. S8, we present simulations with constant carrying capacity $K_i = K$, constant immigration rate $\lambda_i = \lambda$, and exponentially distributed interaction coefficients $Sc_{ij} \sim \text{EXPO}(\sigma)$. The interaction coefficients have mean $\langle Sc_{ij} \rangle = \sigma$, variance $\langle Sc_{ij}^2 \rangle - \langle Sc_{ij} \rangle^2 = \sigma^2$, and probability density function

$$f(Sc_{ij}) = \frac{1}{\sigma} e^{-Sc_{ij}/\sigma}.$$

Pareto-Distributed Interaction Coefficients. In Fig. S9, we present simulations with constant carrying capacity $K_i = K$, constant immigration rate $\lambda_i = \lambda$, and Pareto-distributed interaction coefficients $Sc_{ij} \sim \text{PARETO}(\sigma)$. The interaction coefficients have mean $\langle Sc_{ij} \rangle \approx \sigma/(1+\sigma-\sigma^2/2)$, variance $\langle Sc_{ij}^2 \rangle - \langle Sc_{ij} \rangle^2 \approx \sigma^2$, and probability density function

$$f(Sc_{ij}) = \frac{2+\sigma^{-1}-\sigma/2}{(1+Sc_{ij})^{3+\sigma^{-1}-\sigma/2}}.$$

We require $\sigma < \sqrt{2}$ for this distribution to have a finite second moment.

Island Biogeography

The qualitative features of the equilibrium behavior of a community are described by its position on the phase diagram just as the state of water can be characterized as a solid, liquid, or gas, depending on pressure and temperature. The ecological consequences of the niche-to-neutral transition can be illustrated by comparing communities that correspond to different regions of the phase diagram or by the studying how the characteristics of a community change in response to a disturbance that moves the community from one phase to another.

According to the theory of island biogeography, the rate of immigration to an island decreases with its distance from the source population such that $\Lambda \sim -\text{distance}$ (19). Therefore, a chain of islands at different distances from the mainland traces a path along the immigration axis of the phase diagram (Fig. S10A). The percentage of species from the source population that inhabit an island, its “saturation” $100 \times \sum_i s_i / S$, decreases with the distance of the island from the mainland. As shown in Fig. S10B and C, our model can reproduce the qualitative trend in saturation as a function of distance famously observed for islands off the coast of New Guinea (19). In addition, our model predicts that there is a distance that defines a sharp niche-to-neutral transition.

Methods Used in Numerical Simulations

We simulated Hubbell's neutral model with a local community of J individuals connected to an infinitely large metacommunity containing $S = 50$ equally abundant species (15). In each time step, with probability λ , an individual randomly drawn from the metacommunity replaced a randomly chosen individual in the local community or, with probability $1 - \lambda$, one randomly chosen individual in the local community replaced another randomly chosen individual in the local community. The simulations were run for 5×10^7 steps. Ten simulations were run for each set of parameters, and the results were averaged.

We studied well-mixed, competitive communities described by a system of stochastic LV equations given by

$$\frac{dx_i(t)}{dt} = \lambda_i + x_i \left(K_i - x_i - \sum_{j \neq i} c_{ij} x_j \right) + \sqrt{\omega x_i} \eta_i(t). \quad [\text{S17}]$$

LV simulations with $S = 50$ species were performed over the parameter ranges specified in the relevant figure legends. The system of equations (Eq. S17) was forward integrated using the Milstein method, as

$$x_i(t + \delta t) = x_i(t) + \lambda \delta t + x_i(t) \left(K_i - x_i(t) - \sum_{j \neq i} c_{ij} x_j(t) \right) \delta t + \sqrt{\omega x_i(t)} \delta W + \frac{\omega}{4} \left((\delta W)^2 - \delta t \right),$$

where $\delta W \sim \sqrt{\delta t} \mathcal{N}(0, 1)$, and $\mathcal{N}(0, 1)$ denotes a standard normally distributed random variable. In each case, the competition coefficients were sampled randomly, and then the stochastic Lotka–Volterra equations (Eq. S17) were forward integrated for 5×10^7 steps of size $\delta t = 0.005$, using the Milstein method (20). Ten simulations were run for each set of parameters, and the results were averaged.

Quantifying Statistical Neutrality. The degree of statistical neutrality, $\text{KL}(P_{LV} \| P_N)$, was measured using Gaussian approximations for $P_{LV}(\vec{x})$ and $P_N(\vec{x})$. The mean abundance of species i was calculated from a simulation of length τ , using $\langle x_i \rangle = \tau^{-1} \sum_{t=0}^{\tau-1} x_i(t)$, and the covariance between species i and j was $C_{ij} = \tau^{-1} \sum_{t=0}^{\tau-1} (x_i(t) - \langle x_i \rangle)(x_j(t) - \langle x_j \rangle)$. The corresponding statistically neutral moment estimates were $\bar{x} = S^{-1} \sum_{i=1}^S \langle x_i \rangle$, $\bar{C}_{ii} = S^{-1} \sum_{i=1}^S C_{ii}$, and $\bar{C}_{ij} = 2S^{-1}(S-1)^{-1} \sum_{i < j} C_{ij}$ for the mean, variance, and covariance, respectively. Thus, $P_{LV}(\vec{x})$ [or $P_N(\vec{x})$] is a multivariate Gaussian distribution with mean $\langle \vec{x} \rangle$ (or $\bar{\vec{x}}$) and covariance matrix C (or \bar{C}). The KL divergence was calculated as

$$\text{KL}(P_{LV} \| P_N) = \frac{1}{2} \left(\text{tr}(\bar{C}^+ C) - \text{rk}(\bar{C}) + (\bar{\vec{x}} - \langle \vec{x} \rangle)' \bar{C}^+ (\bar{\vec{x}} - \langle \vec{x} \rangle) - \ln |C|_+ + \ln |\bar{C}|_+ \right) \quad [\text{S18}]$$

Here, $|M|_+$ denotes the pseudodeterminant of the matrix M , M' denotes its transpose, M^+ denotes its pseudoinverse, and $\text{rk}(M)$ denotes its rank. The pseudodeterminant and inverse were used because, in Hubbell's model, the constraint that the number of individuals in the local community is fixed reduces the degrees of freedom by 1. To calculate the degree of statistical neutrality for a set of parameters for the competition coefficients, we averaged the KL divergence over 10 independent draws of the interaction coefficients (c_{ij}).

Related Literature

This section provides some related references on phase transitions in populations. In particular, there are a number of papers on phase transitions in systems of replicator equations (which are closely related to Lotka–Volterra models) with disordered species traits (21–26). The random energy model has also been applied to study phase transitions in evolution (for a review see ref. 27).

- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.
- MacArthur R, Levins R (1967) The limiting similarity, convergence, and divergence of coexisting species. *Am Nat* 101:377–385.
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630.
- Mora T, Bialek W (2011) Are biological systems poised at criticality? *J Stat Phys* 144: 268–302.
- Schneidman E, Berry MJ, 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440(7087): 1007–1012.
- Kindermann R, et al. (1980) *Markov Random Fields and Their Applications* (Am Math Soc, Providence, RI), Vol 1.
- Kadanoff LP (2000) *Statistical Physics: Statics, Dynamics and Renormalization* (World Scientific, Singapore).
- Kirkpatrick S, Sherrington D (1978) Infinite-ranged models of spin-glasses. *Phys Rev B* 17:4384–4403.
- Derrida B (1980) Random-energy model: Limit of a family of disordered models. *Phys Rev Lett* 45(2):79–82.
- Derrida B (1981) Random-energy model: An exactly solvable model of disordered systems. *Phys Rev B* 24:2613–2626.
- Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA* 84(21):7524–7528.
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86.
- Lin J (1991) Divergence measures based on the Shannon entropy. *IEEE Trans Information Theory* 37(1):145–151.
- Fuglede B, Topsoe F (2004) Jensen-Shannon divergence and Hilbert space embedding. *IEEE International Symposium on Information Theory* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), p 31.
- Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (MPB-32) (Princeton Univ Press, Princeton).
- Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424(6952):1035–1037.
- Azaele S, Pigolotti S, Banavar JR, Maritan A (2006) Dynamical evolution of ecosystems. *Nature* 444(7121):926–928.
- Chisholm RA, Pacala SW (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proc Natl Acad Sci USA* 107(36):15821–15825.
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17(4):373–387.
- Milstein GN (1995) *Numerical Integration of Stochastic Differential Equations* (Springer, Berlin), Vol 313.
- Diederich S, Oppen M (1989) Replicators with random interactions: A solvable model. *Phys Rev A* 39(8):4333–4336.
- Chawanya T, Tokita K (2002) Large-dimensional replicator equations with antisymmetric random interactions. *J Phys Soc Jpn* 71:429–431.
- De Oliveira V (2003) Replicators with hebb interactions. *Eur Phys J B* 31:259–263.
- Poderoso F, Fontanari JF (2005) The random replicator model at nonzero temperature. *Eur Phys J B* 48:557–565.
- de Oliveira VM, Fontanari JF (2001) Extinctions in the random replicator model. *Phys Rev E Stat Nonlin Soft Matter Phys* 64(5 Pt 1):051911.
- Galla T (2006) Random replicators with asymmetric couplings. *J Phys Math Gen* 39:3853–3869.
- Neher RA, Shraiman BI (2011) Statistical genetics and evolution of quantitative traits. *Rev Mod Phys* 83:1283–1300.

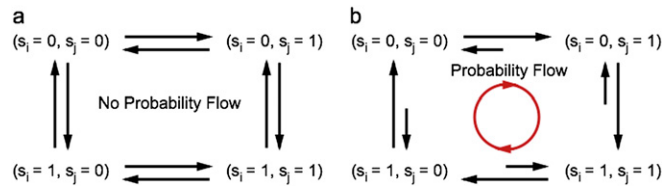


Fig. S1. Equilibrium and detailed balance. (A) In a system that satisfies detailed balance, the flows of probability into and out of all of the states are equal so that the system remains in equilibrium. (B) In a system without detailed balance, there is a nonzero flow between the different configurations of the community, even when the probability distribution reaches a steady state.

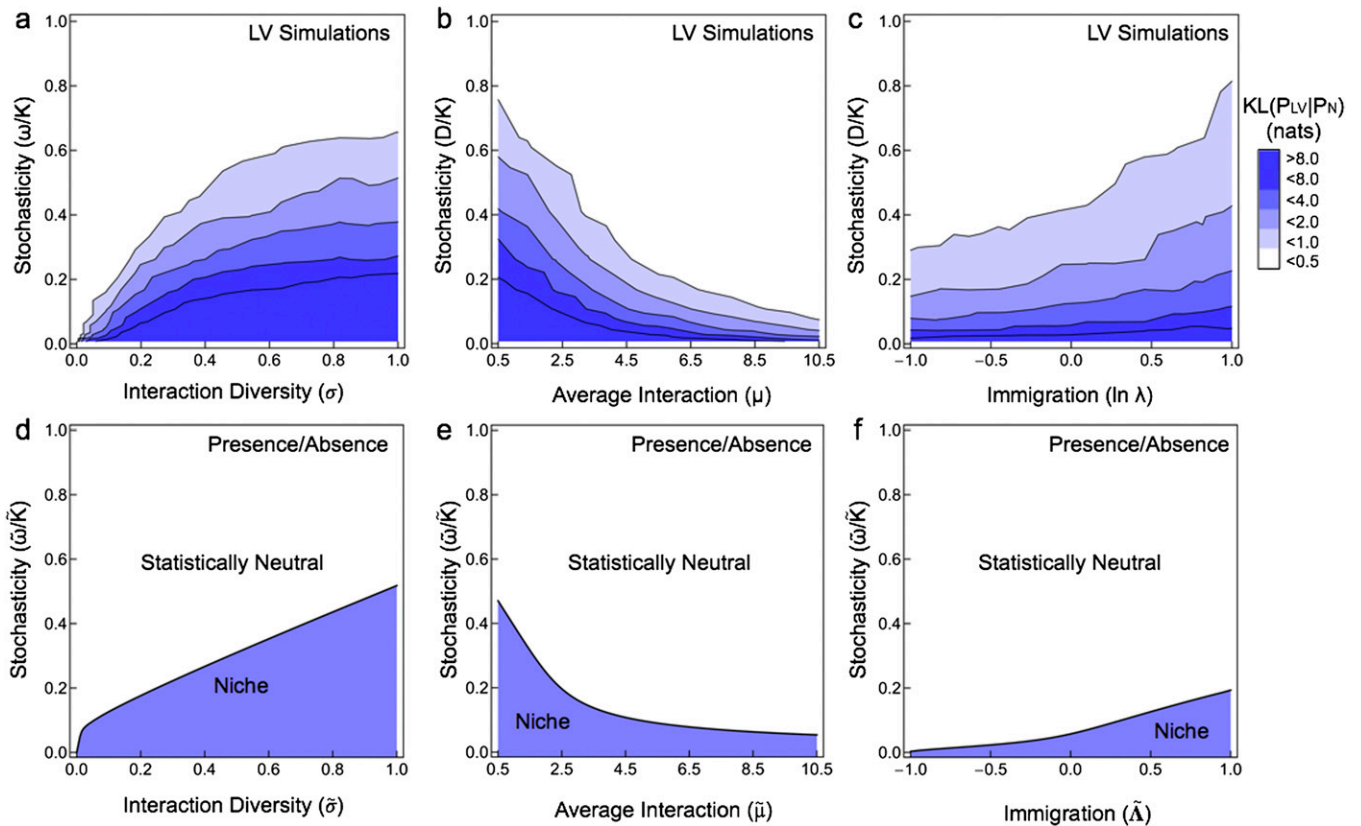


Fig. S2. Slices through the niche–neutral phase diagram demonstrate the scaling relation $\omega \sim K \ln \lambda \sigma / \mu$. (A–C) The KL divergence measured during LV simulations while varying interaction diversity (σ), the mean interaction strength ($\bar{\mu}$), and the immigration rate (λ). (D–F) Phase diagrams calculated from the presence/absence model while varying interaction diversity ($\bar{\sigma}$), the mean interaction strength ($\bar{\mu}$), and the immigration potential ($\bar{\lambda}$).

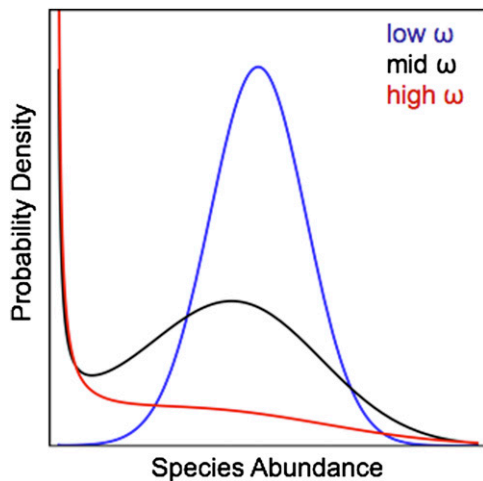


Fig. S3. Species abundance distributions computed from a single-species LV model with different amounts of stochasticity.

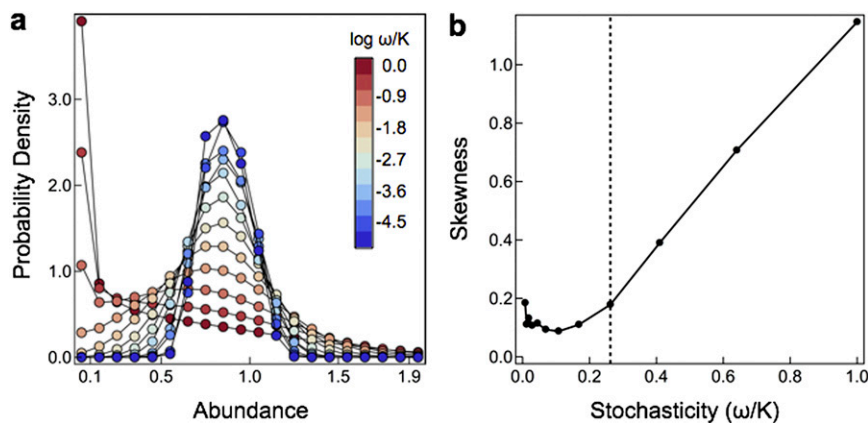


Fig. S4. Species abundance distributions from LV simulations. (A) Species abundance distributions with $\mu = 1.0$, $\sigma = 0.1$, and $\lambda = 0.01$. (B) Skewness of the species abundance distribution increases beyond the critical stochasticity.

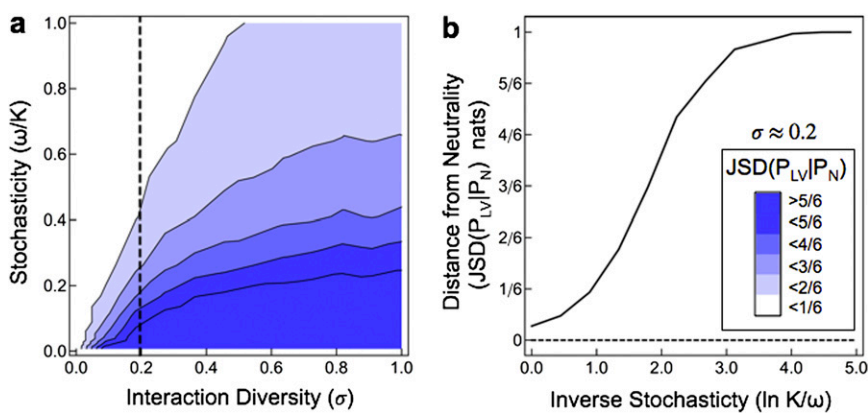


Fig. S5. Phase diagram and phase transition calculated with the Jensen-Shannon divergence in LV communities. (A) Phase diagram as a function of interaction diversity and stochasticity. (B) A slice through the phase diagram demonstrating the sharpness of the transition.

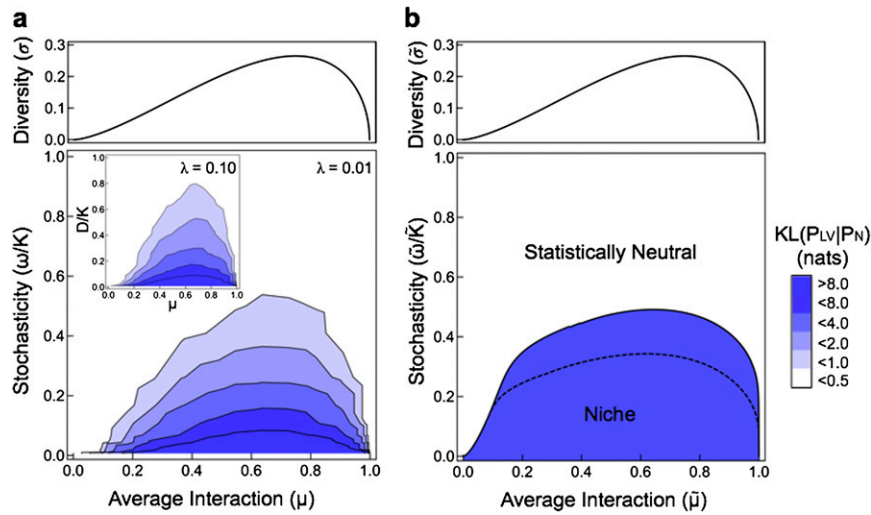


Fig. 56. Phase diagram with low-rank interactions. (A) Simulations of competitive LV communities with immigration rate $\lambda = 0.01$ (inset $\lambda = 0.1$) display two phases: a statistically neutral phase with $\langle \text{KL}(P_{LV}|P_N) \rangle \sim 0$ and a niche phase with $\langle \text{KL}(P_{LV}|P_N) \rangle \gg 0$. The critical stochasticity defining the phase boundary is approximately proportional to interaction diversity (σ). (B) The phase diagram calculated from the PA model with high ($\bar{\lambda} = 1.5$; solid line) and low ($\bar{\lambda} = 0.5$; dashed line) immigration potentials has a statistically neutral phase and a niche phase. The niche phase is largest for communities with high interaction diversity (σ) and high rates of immigration.

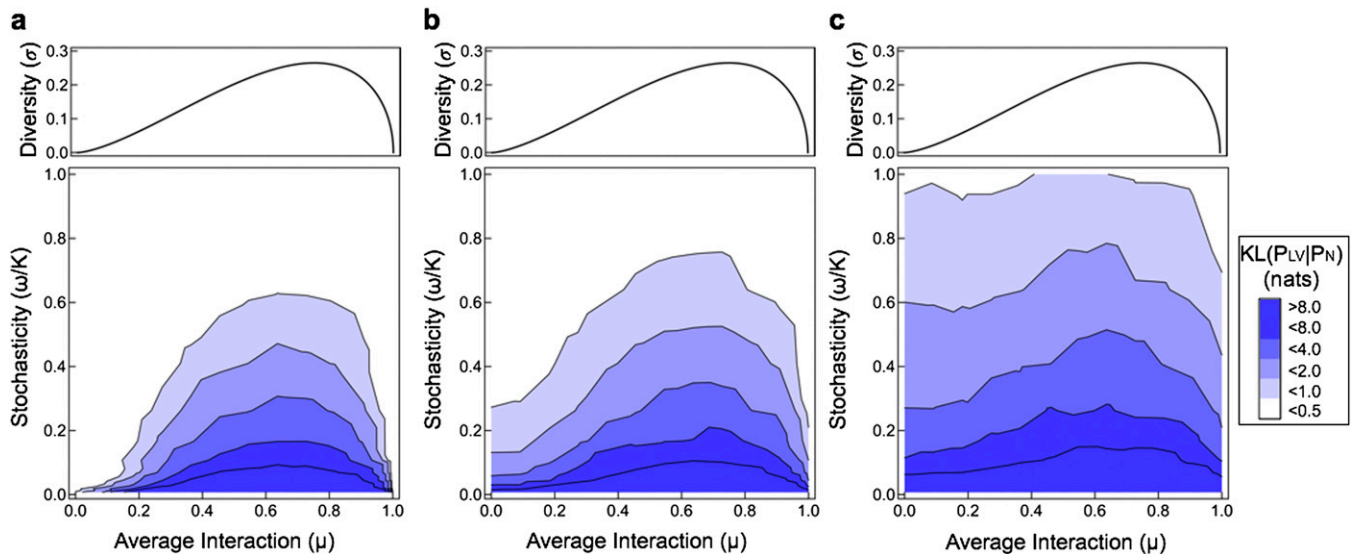


Fig. 57. Phase diagram of noisy low-rank interactions. (A) Simulations of competitive LV communities with 1% multiplicative noise ($\varepsilon = 0.01$). (B) Simulations of competitive LV communities with 5% multiplicative noise ($\varepsilon = 0.05$). (C) Simulations of competitive LV communities with 10% multiplicative noise ($\varepsilon = 0.10$). In all cases, the immigration rate was $\lambda = 0.1$.

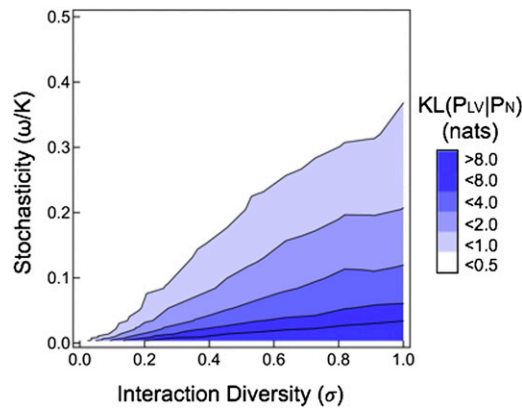


Fig. S8. Phase diagram of competitive ecosystems with exponentially distributed interaction coefficients. The critical stochasticity scales with functional diversity (σ) as predicted from theory, even though the interaction coefficients are drawn from a different distribution. The immigration rate was $\lambda = 0.1$.

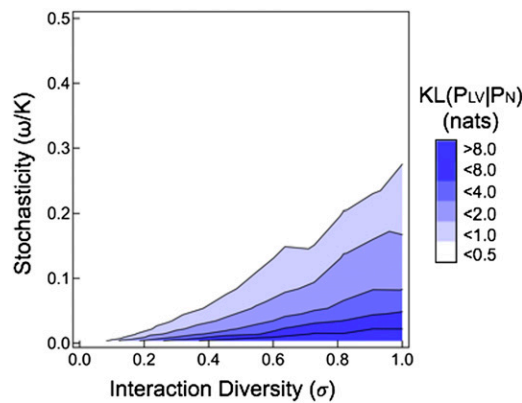


Fig. S9. Phase diagram of competitive ecosystems with Pareto-distributed interaction coefficients. The critical stochasticity scales with functional diversity (σ) as predicted from theory, even though the interaction coefficients are drawn from a different distribution. The immigration rate was $\lambda = 0.1$.

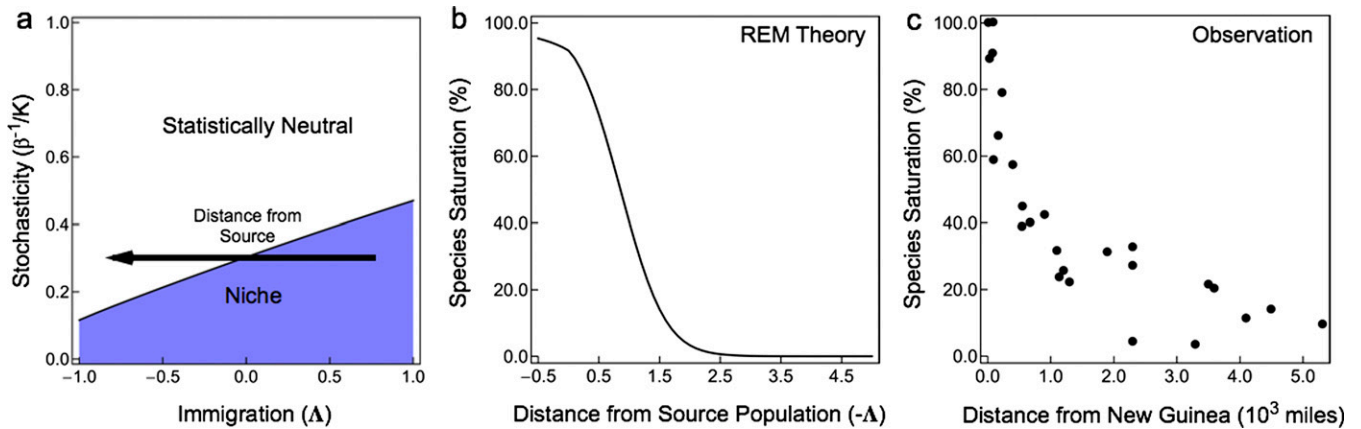


Fig. 510. Spatial variation in immigration and island biogeography. (A) Islands at different distances from the mainland are located in different regions of the phase diagram because the immigration potential ($\bar{\lambda}$) decreases with distance from the source population. (B) The percentage of species inhabiting an island (its saturation) decreases with distance ($-\bar{\lambda}$) from the source population in the PA model. The PA model calculations are for $\bar{\mu} = 0.5$, $\bar{\sigma} = 0.5$, and $\bar{\omega}/K = 0.3$. (C) Similarly, species saturation decreases with distance from the source population in islands off the coast of New Guinea (19).