Generalizing clusters of similar species as a signature of coexistence under competition

Rafael D'Andrea,[1,2,*] Maria Riolo[1,†] & Annette Ostling[1,‡]

1. Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA
2. Plant Biology, University of Illinois, Urbana-Champaign, Illinois, USA
∗ Corresponding author; e-mail: rdandrea@illinois.edu. ORCID 0000-0001-9687-6178
† mariolo@umich.edu.
‡ aostling@umich.edu.

Manuscript elements: Abstract (185 words), Main text (4,270 words), Box 1, Figures 1-4, Online appendices A, B.

Keywords: Lumpy coexistence, Self-organized similarity, Niche differentiation, Stochastic niche assembly, Immigration.

Manuscript type: Research Article.

Author statement: RD and AO designed the study. RD wrote the code, carried out simulations, performed data analysis, and wrote the manuscript. MR developed the Ripley's K metric. All authors contributed to revisions of the manuscript.

Data availability: The code for the metrics can be found at `https://github.com/rafaeldandrea/Clustering-metric`. All code used in the study will be uploaded to GitHub if the manuscript is accepted.

Abstract

Patterns of trait distribution among competing species can potentially reveal the processes that allow them to coexist. It has been recently proposed that competition may drive the spontaneous emergence of niches comprising clusters of similar species, in contrast with the dominant paradigm of greater-than-chance species differences. However, current clustering theory relies largely on heuristic rather than mechanistic models. Furthermore, studies of models incorporating demographic stochasticity and immigration, two key players in community assembly, did not observe clusters. Here we demonstrate clustering under partitioning of resources, partitioning of environmental gradients, and a competition-colonization tradeoff. We show that clusters are robust to demographic stochasticity, and can persist under immigration. While immigration may sustain clusters that are otherwise transient, too much dilutes the pattern. In order to detect and quantify clusters in nature, we introduce and validate metrics which have no free parameters nor require arbitrary trait binning, and weigh species by their abundances rather than relying on a presence-absence count. By generalizing beyond the circumstances where clusters have been observed, our study contributes to establishing them as an update to classical trait patterning theory.

Author Summary

Species traits determine how they compete with each other. As such, patterns in the distributions of traits in a community of competing species may reveal the processes responsible for coexistence. One central idea in theoretical ecology is that the strength of competition relates to similarity in species needs and strategies, and therefore if competition plays out at short timescales, coexisting species should be more different than expected by chance. However, recent theory suggests that competition may lead species to temporarily self-organize into groups with similar traits. Here we show that this clustering is a generic feature of competitive dynamics, which is robust to demographic stochasticity and can be indefinitely maintained by immigration. We show that clustering arises whether species coexist by partitioning resources, environmental preferences, or through tradeoffs in life-history strategies. We introduce and validate metrics that, given species traits and abundances, determine whether they are clustered, and if so, how many clusters occur. By showing the generality of self-organized species clusters and providing a tools for their detection, our study contributes to updating classical ideas about how competition shapes communities, and motivates searches for them in nature.

Introduction

Competition is a driving force in nature, and a central question in ecology is how it shapes community structure. Species traits mediate their interactions with the environment and each other, and therefore determine how they compete. As such, patterns of trait distribution among co-occurring species can give insights into the underlying coexistence mechanisms [1–5]. Theory shows that coexistence is only stable if species differ in their ecological needs and impacts—i.e. they must display 'niche differences' (see Glossary) [6, 7]. These lead to the lessening of competition between species compared to competition within species necessary for each species to grow from low abundance in the presence of the others. Hence when species are arranged

on 'niche axes' so that proximity on those axes indicates similarity in their niche strategies, they are expected to expected to display 'limiting similarity'–greater-than-chance differences on those axes [8–17]. Ecologists look for these patterns in the distribution of species on trait axes reflective of positioning on these niche axes. This classical idea remains the dominant paradigm, despite mixed empirical support [2, 3, 5].

In contrast, recent studies suggest that competition causes the spontaneous emergence of transient clusters of similar species on trait axes [18–20]. Limiting similarity results only if competition plays out to its end result, of all but one species in each cluster being excluded, on short time scales. The species that would remain have the emergent 'optimal niche strategies' (see Glossary). They are a set of species that are niche differentiated enough from one another (i.e. have enough space between them on the niche axis) to all stably coexist together, and have emerged as dominant over other strategies through the competitive process. Their dominance can be influenced in part by the environment (which might favor some strategies over others), or simply arise from the effect of starting conditions or the order of arrival of species to the community on subsequent competitive dynamics. When species outnumber these optimal niches, all other species would eventually be competitively excluded. That exclusion is slow on species near the optimal niche strategies, and fast away from them [21–23]. As a result, the community temporarily self-organizes into clusters of similar species near optimal strategies, with gaps in between. Several mechanisms can make these transient clusters persist, such as specialist enemies [18] and environmental fluctuations [24, 25].

Clustering has traditionally been associated with environmental filters [26], and more recently with one-sided competition without a balancing tradeoff, such as competition for light where taller is better [27]. However, those will tend to produce a single big cluster around the favored trait, whereas partitioning of a niche axis will lead to multiple clusters, one per optimal niche strategy. Patterns of multiple clusters have in fact been reported in empirical studies of body size in phytoplankton communities and certain animal taxa [18, 28–30]. This phenomenon has been interpreted to bridge coexistence through differences (niche theory) and similarity (neutral theory) [31], and as such potentially represents a unification of classical ideas and a generalization of limiting similarity.

However, the generality of clusters as a signature of competition cannot be established without showing that they emerge in communities subject to stochastic processes. In particular, immigration and ecological drift are intrinsic to most communities, and have been amply demonstrated to be important players in nature [31–34]. In fact, models ignoring all of biology but drift and immigration successfully describe observed macroecological patterns [35] (which is not to say, of course, that deterministic forces are unimportant). Yet, clustering remains unseen in competition models incorporating these processes [36–38]. In plankton models where clustering occurs, immigration has as a negative impact [39].

Moreover, clusters have not been widely demonstrated beyond Lotka-Volterra dynamics. Lotka-Volterra competition equations are a heuristic description which does not specify a niche mechanism. They are a special limit [40] of MacArthur's consumer-resource dynamics model [41], and cannot describe all types of competition [42]. While species clusters have been shown to emerge in explicit consumer-resource dynamics [25, 39], these studies ignored the possibility of resource depletion effects. The latter have been shown to greatly affect competitive relations between consumers, for example by violating the assumption that competition always decreases with trait differences [43]. It is also not clear that clustering should emerge among species com-

peting along environmental gradients; indeed, early studies of stochastic niche dynamics [36–38] focused on competition of this type and found no clustering pattern (although these studies, which predate [18], were not specifically looking for clustering). Finally, it is not known whether clusters emerge in communities characterized by hierarchical competition among species that coexist via life-history tradeoffs. These mechanisms, such as the competition-colonization tradeoff [44] and the tolerance-fecundity tradeoff [45], have been chiefly studied in terrestrial plants, but may enable coexistence among other sessile organisms with a dispersive stage, such as coral, coral fishes, and microbes [45].

To determine whether clusters are a general outcome of competition as opposed to an artifact of specific models, and to verify their robustness to stochastic forces, here we use a stochastic niche simulation approach to investigate the emergence of clustering by traits in species assemblages undergoing competitive dynamics and open to immigration.

We start with Lotka-Volterra dynamics, where clusters are known to emerge in the deterministic model [18], to first broach the question of their robustness to stochastic forces. We test for this robustness with and without the confounding influence of environmental filters, which we hypothesize might mask any clustering caused by niche partitioning. Furthermore, we see how the robustness of cluster emergence to stochasticity is influenced by regional diversity and immigration rate. Clustering intrinsically involves many species, and should in principle be stronger, or at least more detectable, under a high species-to-niches ratio. Therefore we expect more clustering under higher regional diversity. As for immigration, we expect that while it may contribute to the persistence of clusters by keeping weak competitors from being excluded, too much immigration will drown any pattern caused by competition.

Next, to determine whether clusters are a general outcome of competition, we analyze models spanning three key niche mechanisms: resource partitioning, habitat partitioning, and a competition-colonization tradeoff. Under resource partitioning, we further explore the differences between scenarios with low and high resource depletion. Under competition for habitat, we also examine the impact of dispersal limitation as opposed to global dispersal. We expect that larger scale dispersal relative to the scale of the environmental gradient could influence cluster emergence by lessening the dominance that emergent optimal strategies might achieve, since their propagules would be more spread to locations with suboptimal habitat. If clusters are a signature of competition and niche differentiation, we expect clustering to emerge generally under all scenarios of all three niche mechanisms.

To determine whether communities are clustered beyond patterns that could arise by chance, we present two metrics: the first uses the k-means clustering algorithm [46] to assign species to clusters by minimizing trait differences within clusters, and then quantifies clustering based on how small the sum of those differences is. The second is based on Ripley's K [47], and quantifies clustering based on how relatively unoccupied the regions between clusters are. Both metrics apply the gap statistic method for comparing data with null models [48]. These metrics improve over existing tools to detect clustering [18, 28–30, 49] as they contain no free parameters and do not arbitrarily bin traits. Furthermore, rather than reducing species to presence-absence counts, our metrics weigh them by their abundance. We validate the metrics by confirming that they detect clusters in Lotka-Volterra communities but not in neutral communities.

Results

All niche mechanisms (Fig. 1) resulted in species clustering (Fig. 2A-G, Fig. 3), demonstrating both the robustness to stochasticity and generality of clustering emergence under niche differentiation. In contrast, neutral communities were not clustered, as expected (Fig. 2H), and communities under no niche mechanism but subject to environmental filters collapsed into a single cluster (results not shown). Our success at detecting clusters, combined with this lack of false detection when niche mechanisms are absent, suggests our metrics are well-suited to quantifying the pattern arising from niche differentiation.

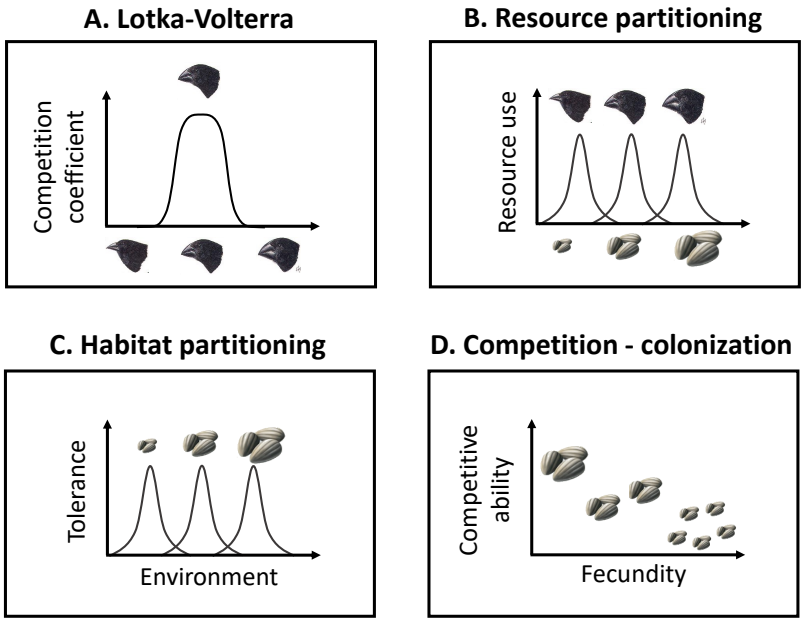


Figure 1: **Schematic representation of niche models.** Panel A illustrates Lotka-Volterra competition, while panels B-D illustrate different niche mechanisms, with the niche axis represented on the abscissa. **A**. Lotka-Volterra: competition coefficients are tied to traits, here beak size. Curve shows competitive impact by each species on the focal species, increasing with trait similarity. **B**. Partitioning resources: We assume a continuum of resources (seeds), and species traits (beak size) determine their use of the different resources. Depending on the degree of specialization, resource depletion may occur (see text). **C**. Partitioning environment: species with different traits (seed size) are optimally adapted to different environments. This model is spatially explicit, and we implement it both with and without dispersal limitation; **D**. Competition-colonization tradeoff: species trade off fecundity (seed output) with ability to win sites against other species (mediated by seed size).

Lotka-Volterra

We implemented a stochastic analogue of classical Lotka-Volterra competition, which assumes species influence each other directly via competition coefficients. We tied those coefficients to

similarity in traits (Fig. 1A), so that trait differences can enable stable coexistence (niche differentiation). In its deterministic form, this type of model is known to produce clusters[18], hence we used this model to consider its robustness to stochastic forces. We also used it to consider whether that robustness of clustering is influenced by the presence of environmental filtering, and by regional diversity and immigration.

Clustering was strong enough in the Lotka-Volterra communities to be easily distinguishable by eye (Fig. 2A-B), with all 100 replicates testing significant at the $p < 0.05$ level by either metric (Fig. 3).

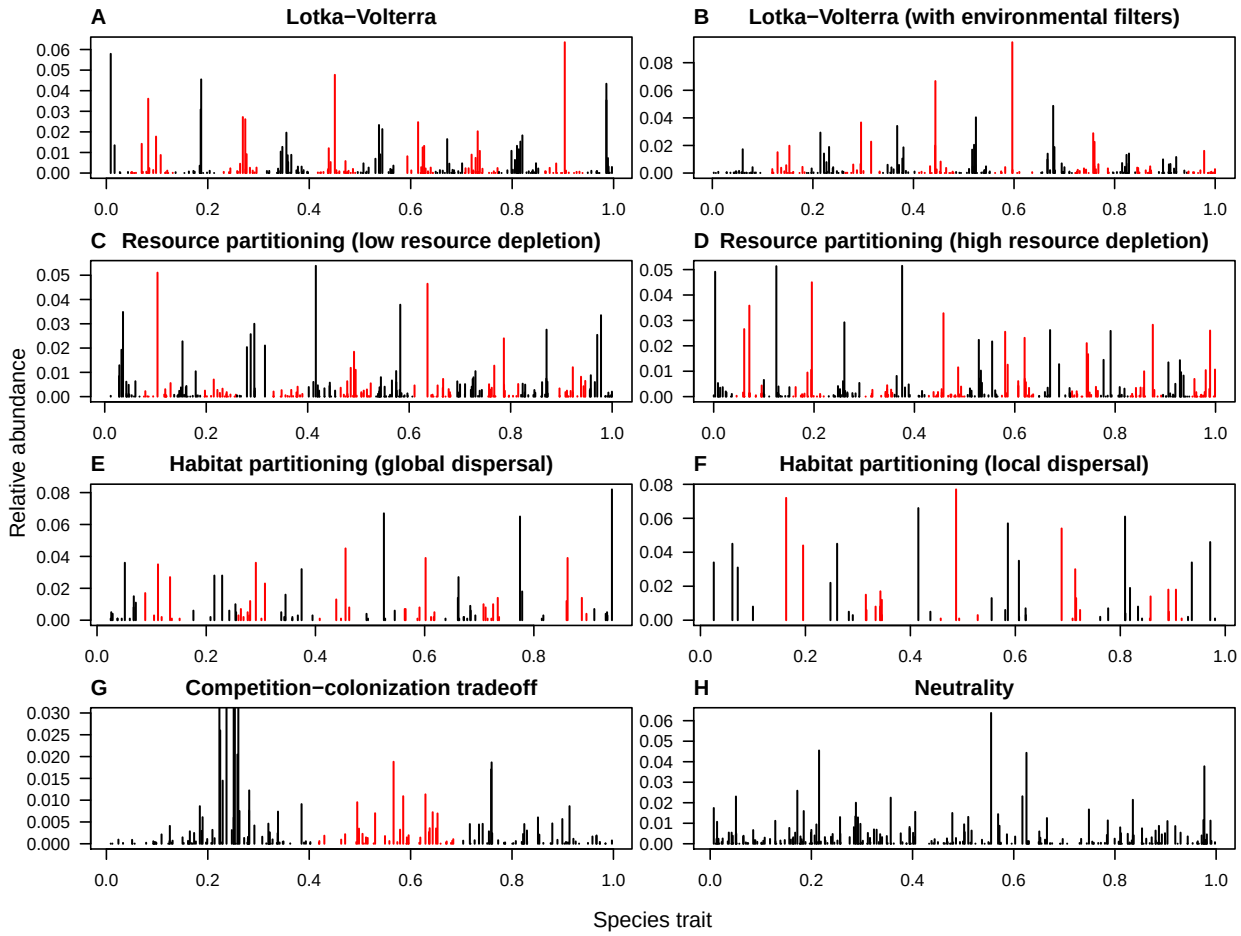


Figure 2: **Abundance-by-trait pattern across niche mechanisms.** For each scenario (A-H) we show the replicate with the highest z-score by either metric. All communities with a niche mechanism (A-G) are clustered at the $p < 0.05$ level, while the neutral community (H) is not ($p = 0.1$). Alternating black and red colors highlight the clusters. We truncate the y-axis in the competition-colonization tradeoff scenario (G) to better show abundance structure among rare species. (Immigration rate $m = 0.08$, regional diversity c. 400 species. All communities had 21,000 individuals, except in the partitioning environment scenario, which had 1,000 individuals.)

Environmental filtering

Environmental filtering [26] or physiological constraints may cause some species to grow faster than others in the absence of competitors. For example, species with intermediate traits may grow faster because extreme trait values may have metabolic costs or are maladapted to the local environment, regardless of competitive interactions. This will affect abundances, and might interfere with the emergence of niche-driven clusters. In our simulations, environmental filters against extreme traits caused species near the center of the trait axis to be more abundant than their counterparts near the edges, as expected (Fig. 2B), but this had minimal impact on clustering relative to no filtering (compare average z-scores, Fig. 3). However, as we increased the strength of environmental filters, they ultimately eclipsed niche partitioning as a driver of pattern, and the community appeared as a single cluster (Fig. S1), consistent with results from environmental filters acting alone.

Regional diversity

In the Lotka-Volterra simulations, clustering increased with regional diversity. As we increased the number of species in the regional pool while keeping the number of clusters fixed, the z-score of the clustering metric rose monotonically (Fig. 4). This reflects the fact that clustered patterns in general are increasingly distinguishable from randomness as the number of items per cluster increases (Fig. S6).

Immigration rates

Clustering had a modal relationship with immigration pressure: as we increased the immigration rate from low values, the clustering z-score first rose, then peaked and fell (Fig. 4). This indicates that while immigration sustains the clusters by replenishing populations that would otherwise be excluded, too much immigration dilutes the effects of local dynamics and adds noise to the pattern (Fig. S6). While this modal shape is presumably general, the position of the mode is specific to the niche mechanism and other factors such as environmental filters.

Communities in the other niche scenarios under immigration rate $m = 0.01$ as opposed to 0.08 generally showed a similar number of clusters but fewer species per cluster, and gaps between clusters were more rarified (Fig.S2). In some cases, this increased clustering (Fig. S3), suggesting that $m = 0.08$ falls further to the right of the mode in those niche models.

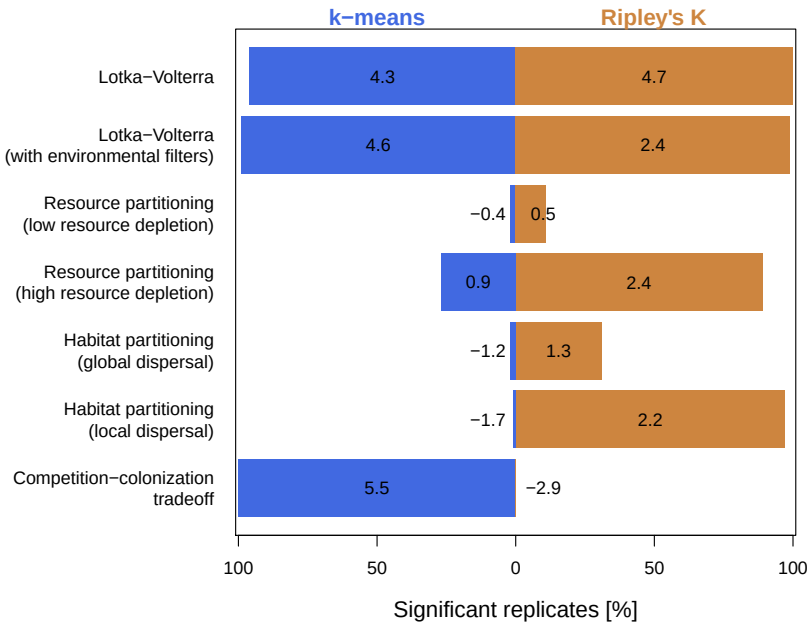


Figure 3: **Clustering across niche mechanisms**. Bars show percentage of replicates that were significantly clustered at level $p < 0.05$ according to the k-means metric (blue) and the Ripley's K metric (orange), out of 10 replicates. Numbers within bars show average z-score. None of the neutral communities were clustered by either metric ($\overline{Z}_{\text{k-means}} = 0.32$, $\overline{Z}_{\text{Ripleys'K}} = 0.39$). ($m = 0.08$, c. 400 regional species across scenarios.)

Resource partitioning

Assuming fast resource dynamics relative to consumers, one can derive the Lotka-Volterra model from resource-consumer dynamics [40, 50], with trait similarity indicating similar resource preferences [8, 41] (Fig. 1B). However, this approach ignores the effects of stochastic fluctuations in resource availability, and particularly resource depletion. The latter has been shown to impact coexistence outcomes [51] and even competition-similarity relationships [43]. These relationships are drivers of pattern [52], and we therefore tested whether incorporating stochastic resource dynamics and depletion affected the emergence of clusters.

We found that clustering occurred under explicit resource-consumer dynamics, but was overall weaker than in our Lotka-Volterra simulations (Fig. 3).

Consumers in the low resource depletion scenario undergo approximate Lotka-Volterra dynamics, but with competition coefficients following a Gaussian function of trait separation [8]. Competitive sorting under Gaussian competition has been shown to be slow (see Fig. S8 in [22]), and our results reflect this slowness relative to immigration, which dilutes the pattern. Indeed, clustering was stronger under lower immigration pressure (Figs. S2B, S3).

Resource depletion breaks the link to Gaussian Lotka-Volterra dynamics, thereby fundamentally changing the competitive interactions between consumers [43, 51]. In our simulations it strengthened clustering (Fig. 3). As resources were extirpated, species ended up clustering based on specialization to the remaining ones (Fig. S4). While resource depletion did not qualitatively

change the emergence of clusters, the fact that it sets the optimal niche strategies (i.e. specialization onto remaining resources) made the pattern stronger.

Habitat partitioning

Habitat is a critical resource for which species compete, and as such the environment is thought of as a key axis of niche differentiation [53–55]. Traits reflect adaptations to different environments, and hence competition for habitat must shape trait distribution.

To test whether niche-differentiation based on environmental preference leads to clustering, we used the model introduced by [37]. We assume a linear landscape on a habitat gradient, e.g. an elevation gradient. Different species are optimally adapted to different habitats, and competition arises from overlap in environmental preference (Fig. 1C). Competition occurs at the recruitment stage, where the probability of recruiting is based on tolerance to the local environment. Because dispersal can play a central role in competition for space [55, 56], we consider two scenarios: global and local dispersal, which differ by whether individuals are more likely to disperse shorter distances from their parents. Since this is an individual-based model, we used a smaller community size (1,000 individuals rather than 21,000) for computational expedience.

We found that clusters also emerge under competition for habitat. Switching dispersal from global to local had a strong impact on cluster shape, and reduced the number of species per cluster (compare Fig. 2E, F). This occurred as dispersal limitation effectively decreases immigration. Moreover, the attending reduction of the diluting effects of immigration (see below) substantially strengthened clustering (Fig. 3).

Competition-colonization tradeoff

All models examined so far describe symmetric competition, whereby the competitive impact of species A on B is similar to that of B on A. Competitive hierarchies stand in contrast to this. That is the case of the competition-colonization tradeoff [44, 57], where propagule production, or colonization ability, trades off with the ability to displace individuals of other species, or competitive ability (Fig. 1D). Even though in this case some species are better competitors than others, the tradeoff is a niche mechanism because it allows for the stable coexistence of multiple species[1] (see Glossary, Fig. S5).

Clusters also emerged under this niche mechanism. The asymmetry in species interactions was reflected in its asymmetric clusters (Fig. 2G): while we found three clusters in all replicates, species in the first cluster always dominated the community (e.g., the most abundant species in the first cluster in Fig. 2G had 4,798 individuals, compared with 395 in the second and 392 in the third). This indicates that species adopting the high-competitiveness strategy (left side of the trait axis) outperform both those who invest in high fecundity and the intermediate group. Dispersal limitation could reduce the asymmetry by augmenting the benefits of high fecundity. We hypothesize that a spatially explicit formulation of this mechanism would produce more similarly sized clusters.

[1] In fact the model can be cast in Lotka-Volterra form, and one can show that the net competitive impact, while asymmetric, is stronger between more similar species (Fig. S5).

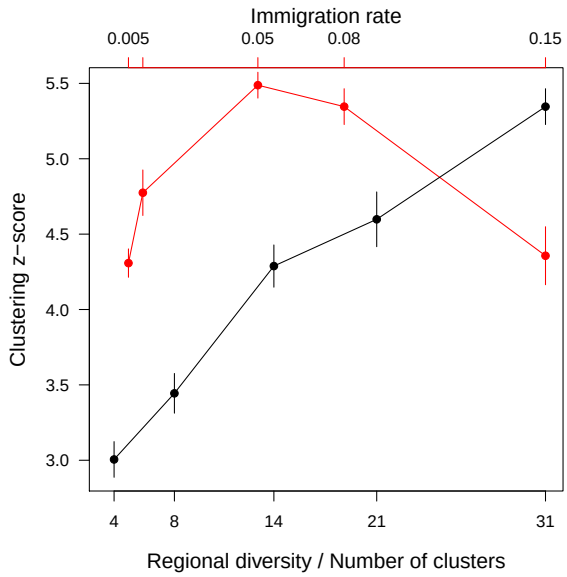


Figure 4: **Impact of immigration and regional diversity**. Clustering increases with regional diversity (black), and has a modal relationship with immigration pressure (red). Bottom axis: number of species in the pool divided by the number of clusters in the community. Top axis: immigration rate. Data points show the mean z-score of the Ripley's K metric across 10 replicates. Error bars show 1.96 standard errors of the mean. ($m = 0.08$ when varying regional diversity, and the latter is fixed at c. 400 regional species when varying immigration.)

Metric performance

While our metrics performed equally well in the Lotka-Volterra communities, they differed in the other niche mechanisms. The Ripley's K metric fared better than the k-means metric in scenarios where species partition resources and habitat (Fig. 3). Ripley's K focuses on the regularity of the gaps between species rather than on the clusters themselves. As such, it is better capable of identifying single-occupancy clusters than the k-means algorithm, as long the clusters are regularly spaced. On the other hand, the k-means metric found the clusters in the competition-colonization tradeoff while Ripley's K missed them (Fig. 3). This is because the k-means algorithm is less sensitive to strong asymmetries between the clusters. In these cases, the k-means metric is a better choice.

Discussion

Ecologists have long sought to understand how competition shapes community structure. While competing species are usually expected to be more different than predicted by chance [26, 58], recent studies suggest that competition may cause species to cluster by traits, such that the community self-organizes into groups of similar species. Scheffer and van Nes [18] first demonstrated the emergence of clusters in Lotka-Volterra dynamics. Later work showed that the shape of the function connecting competition to traits is essential to pattern-forming instabilities in general

[52], and cluster emergence in particular [59]. In deterministic models clustering is a transient phenomenon unless extra sources of intraspecific competition are added [18, 59], or resources fluctuate periodically [25].

Our study verified that clustering transcends Lotka-Volterra dynamics, occurring under a number of niche mechanisms. Further, we showed that clustering is robust to stochastic drift, an intrinsic property of real-life communities. Immigration maintains clusters that are otherwise transient, and the strength of clustering has a modal relationship with immigration pressure. We showed that clustering may be detectable under the confounding influence of environmental filters, and is enhanced by regional diversity. Finally, we provided metrics for detecting and quantifying clusters in nature.

Why do clusters arise? Different niche mechanisms share the common property that competition is stronger between species with more similar strategies. It thus seems paradoxical that clusters should emerge. However, it is precisely because species with similar niches compete more strongly that clusters appear [60]. While similar pairs compete more strongly, they experience similar competitive pressure (or relief) from the rest of the community. If a given niche strategy is favored because it minimizes competition with the rest of the community or capitalizes on greater resource supplies, then similar strategies are similarly favored. This hilly fitness landscape causes exclusion to be slower near the center of the niches than in the gaps between them, making it easy for immigration to permanently maintain the clusters [22].

Modern coexistence theory [7] splits coexistence-promoting processes into those that reduce competition among species relative to competition within species (stabilizing mechanisms, here referred to as niche mechanisms), and those that reduce differences in average fitness between species (equalizing mechanisms). It is thus tempting to interpret clusters as reflecting a harmonious combination of stabilizing and equalizing forces: species within a cluster are equalized, while those in different clusters are stabilized. However, this interpretation is problematic. The equalization-stabilization dichotomy is based on applying invasibility criteria to closed communities regulated by a small number of limiting factors [61], an approach which does not extend easily to multispecies communities under immigration and a continuum of resources. In diverse communities with complex competitive interactions, it is difficult to calculate equalizing and stabilizing terms and tie them to trait differences, let alone interpret how specific patterns such as clustering connect with equalization and stabilization forces [61]. While this difficulty can be overcome by assuming that all pairs of species compete with equal intensity [7, 62], this assumption is strongly violated in all models where clustering has been observed so far.

While competition is responsible for clusters' emergence, immigration is responsible for their persistence. Immigration joins other mechanisms that have been previously shown to sustain clusters, namely specialist enemies [18, 63] and environmental fluctuations [24, 25]. We found that clustering appears generally under different immigration regimes, especially if the number of species far exceeds the number of available niches. In studies of stochastic competitive preceding Scheffer and van Nes 2006, [36] and [37] found little impact of immigration on resulting trait pattern. However, successful immigration was highly infrequent in those models. In Tilman's model, resources made available through deaths were assumed to be redistributed broadly, so that resource supply remained low everywhere, making recruitment of new individuals highly unlikely. In [37], immigration took the form of a single immigrant seed being added to a large pool of local seeds competing for the site. As such, immigration rates were effectively much lower than ours. [38] tested both very high and very low immigration, and also saw no clusters (in no

small part because the authors were not looking for them!). This could be due to their use of a Gaussian competition kernel, which leads to weak niche sorting dynamics, easily overwhelmed under high immigration [59].

The fact that immigration maintains clusters seemingly defers the question of coexistence to the regional scale. The problem dissipates by considering mass effects in the metacommunity framework [64]. The regional pool is a combination of local communities, and species are selected for different traits at different sites due to their own local niche dynamics. Therefore, each community receives immigrants which may be dominant elsewhere despite being disfavored locally. However, this does not mean that pattern is expected regionally, as the sum total of heterogeneous communities, each with a different trait pattern, may result in no discernible pattern at a regional scale.

Clusters caused by partitioning of a niche axis are often distinct from clustering due to other processes. Environmental filters favoring a single best trait and one-sided competitive dynamics where a particular trait outperforms all others without a balancing tradeoff [27] will produce a single cluster as opposed to multiple clusters. Where evolutionary rates are commensurate with ecological dynamics, small mutations and sympatric speciation may also generate species or genotype clusters without niche differentiation [65, 66]. Ruling out these alternative sources of clustering could require sampling at a larger scale than applicable to the niche mechanism [5, 58], or directly verifying competitive effects and frequency dependence [4].

We presented and validated two nonparametric abundance-weighted tools for detecting and quantifying clusters. Our metrics successfully distinguished clustering in niche-differentiated communities from no clustering in neutral communities and a single cluster in communities under environmental filters without a niche mechanism. We note that none of these results would appear without considering species abundances, as presence-absence counts do not reveal clustering in our communities (results not shown). Also, our metrics did not require arbitrarily binning traits, nor fitting parameters to the data.

Although our study focused on one-dimensional trait axes, competitive interactions may often be mediated by variation in multiple traits [67]. Theoretical work on simple models indicates that multidimensional niche space leads to multidimensional clustering [20]. Our metrics can quantify clusters in any dimension, using generalized measures of niche separation. The main challenge is to connect multidimensional phenotypes to competitive relations, in order to define the correct measure of distance in high-dimensional trait space [60].

Clustering as a signature of coexistence under competition is an update to the still dominant paradigm that competing species will display greater-than-chance differences [3–5, 16, 17]. Our finding that clusters appear under various niche mechanisms and can be easily maintained by immigration, even when confounding forces are at play, suggests that clusters are a likely feature of nature beyond the instances where it is currently known to occur [18, 29, 30]. For example in tropical forests, where both competition and dispersal are recognized as key drivers of community assembly [68, 69], clustering could help explain the high diversity and seemingly continuous phenotypic variation.

Methods

Models and simulation design

We used a lottery model framework [70] to implement stochastic niche dynamics in a fixed-size community open to immigration from a regional pool. We start with a random draw of offspring from the regional pool, and then alternate death and recruitment events until species abundance distributions are stationary. A proportion m of deaths are replaced by immigrants, and the remainder by local offspring. This is analogous to Hubbell's neutral model [35], except here the niche mechanism sets the probabilities of birth and death across different species. Schematic illustrations of our niche models are shown in Fig. 1.

We used the 50-hectare plot of tropical forest on Barro Colorado Island, Panama, as a reference point for our community size (21,000 individuals $>$ 10 cm dbh [35]) and immigration rate (0.08 immigrant recruits per recruitment event [71]). Our regional pool is a fixed neutral metacommunity with biodiversity parameter $\theta = 50$ [35] and 150,000 individuals, leading to c. 400 species (bigger metacommunities did not change results).

All simulations and statistical analyses were done in R [72].

Lotka-Volterra

The probability that species i is selected for a death event is $\sum_j A_{ij} N_i N_j / \sum_k \sum_l A_{kl} N_k N_l$, where N_i is the abundance of species i and the competition coefficient A_{ij} quantifies the competition caused on species i by species j. The sums are over all species present in the local community, and the denominator normalizes the probabilities to 1. The probability that species i is chosen at the subsequent recruitment stage is $m\, q_i + (1-m)\, r_i N_i / \sum_j r_j N_j$, where q_i is the relative abundance of species i in the regional pool, and r_i is its intrinsic growth rate. (We are placing competitive interactions on mortality, but results are unchanged when we place them on births instead.)

We link competition to traits by setting $A_{ij} = \exp[-(|x_i - x_j|/w)^4]$, where x_i, $x_j \in [0,1]$ are the trait values of species i and j. Thus competition peaks at $x_i = x_j$ (intraspecific competition), and declines with increasing trait difference (Fig. 1A). This is a generic unimodal function commonly used in competition models [52]. Other unimodal shapes bring no qualitative changes (with a few exceptions, see [59]). The scaling parameter w sets the width of niche overlap, and therefore the number of niches available to species. We set $w = 0.063$, leading to c. 13 clusters. Where possible, we set parameters to keep the number of clusters consistent.

In the scenario with environmental filtering we set intrinsic growth rates $r_i = x_i(1 - x_i)$, such that species with intermediate traits grow faster than those with extreme traits. In all other Lotka-Volterra scenarios, we set all $r_i = 1$ identical across species.

We set default values $m = 0.08$ for immigration rate, and $S \simeq 400$ for regional diversity (via $\theta = 50$ in the regional pool). When examining the impact of immigration pressure, we used $m = 0.005, 0.01, 0.05, 0.08, 0.15$. When examining regional diversity, we used $S \simeq 50, 100, 180, 250, 400$ by setting $\theta = 5, 10, 20, 30, 50$.

Partitioning resources

Let X_a and N_i be the abundances of resource a and species i. We assume resource consumption affects species reproduction, and all species have equal mortality. As such, the probability that

species i is chosen for a death event is simply $N_i / \sum_j N_j$, and the probability that it is chosen for a recruitment event is $m\, q_i + (1-m) \sum_a C_{ai} X_a N_i / \sum_j \sum_b C_{bj} X_b N_j$, where C_{ai} quantifies the preference of species i for resource a.

Resources simultaneously follow their own stochastic birth-death process. Consumption lowers resource populations, which would otherwise grow logistically. Therefore, resource a is selected for a death event with probability $\sum_i C_{ai} X_a N_i / \sum_b \sum_j C_{bj} X_b N_j$, and for a birth event with probability $(1 - X_a/K) X_a / \sum_b (1 - X_b/K) X_b$, where K is the resource carrying capacity, set equal across all resources for simplicity. There is no resource immigration; if a resource population is depleted to zero, the resource is extirpated and cannot be replenished.

Resources are arranged on a line between 0 and 1. We define $C_{ai} = \exp[-(d_{ai}/\sigma)^2]$, where d_{ai} is the difference between resource a and the preferred resource of consumer i (Fig. 1B). The width parameter σ controls degree of resource overlap and thus number of niches. We set $\sigma = 0.03$ as it leads to about 13 clusters, for consistency with the Lokta-Volterra scenarios. We control resource depletion via the resource carrying capacity. When $K = 100$, most resources are never severely depleted by consumption, whereas when $K = 400$ resource depletion is severe and many resources are extirpated. Both scenarios start with 500 resources drawn from a standard uniform distribution.

Partitioning habitat

We used the model introduced by [37]. We assume a linear array of sites forming a habitat gradient. Different species are optimally adapted to different environments, and competition arises from overlap in environmental preference (Fig. 1C). Here, a death event is followed by a lottery in the local seed bank, which is formed via dispersal from all individuals in the community, plus a proportion of immigrants. A seed's probability to win the lottery is given by its tolerance to the local environment, which declines with differences from the optimal environment of that species.

All individuals have the same probability of dying. The probability that species i fills the vacancy is then $mq_i + (1-m) \sum_{b \neq a} T_{ai} D_{ab} / \sum_{j=1}^{S} \sum_{c \neq a} T_{aj} D_{ac}$, where D_{ab} is the probability of dispersal between an individual at site b and the vacated site a, and T_{ai} is species i's tolerance to site a. We set $T_{ai} = \exp[-(\Delta_{ai}/\sigma)^2]$, where Δ_{ai} is the difference between the local environment a and the niche optimum of species i. The width parameter σ controls niche overlap, and hence the number of niches. We set $\sigma = 0.07$ as it gives us c. 13 clusters.

We use a Gaussian dispersal kernel, $D_{ab} = \exp[-0.5(d_{ab}/d_{\text{disp}})^2]$, where d_{ab} is the distance between vacated site a and a parent individual at site b. We use 1,000 sites on a line, with adjacent sites one unit distance apart ($d_{a,a+1} = 1$), leading to a fixed community size of 1,000 individuals. The reduced size relative to 21,000 individuals in the other models was used to keep this agent-based model computationally expedient. We explore a scenario with local dispersal ($d_{\text{disp}} = 50$), and one global dispersal ($d_{\text{disp}} = 10^5$), i.e. where the probability of arrival is independent of distance to the parent.

Competition-colonization tradeoff

Under this mechanism, changes to a population result from gains through recruitment and losses via death and displacement. Collecting the density-dependent and density-independent factors separately, one can show (see Supplementary Information) that the deterministic dynamics fol-

low the equation

$$\frac{1}{N_i}\frac{\mathrm{d}N_i}{\mathrm{d}t} = (f_i - \mu) - \sum_{j=1}^{S}(f_i + f_j)\Gamma_{ij}N_j$$

where f_i is the fecundity of species i, μ is a constant, and Γ_{ij} is the ability of a propagule of species j to displace an individual of species i.

The equation above is in Lotka-Volterra form. Therefore in parallel with our Lotka-Volterra scenarios, we place the density dependence on mortality. We thus set the probability that species i is chosen to die as $\sum_j (f_i + f_j)\Gamma_{ij}N_iN_j / \sum_k \sum_l (f_k + f_l)\Gamma_{kl}N_kN_l$, while the probability it is chosen to recruit is $mf_iq_i / \sum_j f_jq_j + (1-m)(f_i - \mu)N_i / \sum_j (f_j - \mu)N_j$. Notice that high-fecundity species immigrate more frequently. The difference $f_i - \mu$ is an intrinsic growth rate, i.e. the net rate at which species grow in the absence of competition. Fecundities are log-uniform distributed between 1 and 6.2. We set $\mu = 1$.

The tradeoff between fecundity and ability to displace is encoded in the displacement matrix Γ_{ij}. If we arrange species in order of fecundity, Γ_{ij} will be approximately bottom-triangular, with $\Gamma_{ij} \approx 0$ if $i < j$ and $\Gamma_{ij} \approx 1$ if $i > j$. That is, species with higher fecundity rarely displace those with lower fecundity, while low-fecundity species commonly displace high-fecundity ones. In the limit of an absolute tradeoff, lower-ranking species never displace and are always displaced by higher-ranking species. This limit has been shown to drastically inflate coexistence under this mechanism, despite the fact that such absolute tradeoffs are unlikely to occur in nature [73–75]. As such, here we use a probabilistic tradeoff: $\Gamma_{ij} = 0.5\,(1 - \tanh[s(f_j - f_i)])$, which has a sigmoidal shape, is equal to 0.5 when $f_i = f_j$, and asymptotes to 1 and 0 when $f_i \gg f_j$ and $f_i \ll f_j$, respectively. Parameter s controls the steepness of the transition, and hence the number of clusters. We set $s = 0.15$, which in the absence of immigration and stochasticity leads to c. 13 coexisting species (Fig. S5).

Clustering metrics: the gap statistic

We apply a variation of the gap statistic method [48], using two different clustering measures: k-means dispersion [46] and Ripley's K function [76]. In general terms, the metric takes in the list of species traits and abundances, and returns the number of clusters (k-means) or average trait separation between them (Ripley's K), as well as a z-score and a p-value. In Box 1 we summarize the k-means version and give a step-by-step recipe for its implementation, and illustrate it at work on two example communities. Both versions are described in detail in the Supplementary Information. The code for the k-means version is available on GitHub [77].

We assess statistical significance and degree of clustering by comparing our communities against null communities where we randomly shuffle abundances across species. This allows us to test specifically for a nonrandom association between traits and abundances, as opposed to a nonrandom distribution of traits or abundances per se. One alternative null model is neutrality, whereby abundances follow a characteristic dispersal-limited multinomial distribution [78].

Because our models are stochastic, we run multiple replicates to account for variation within the same scenario. For each niche scenario we report the average z-score across 10 replicates, as well the percentage of replicates that were significantly clustered at level $p < 0.05$.

We check for false positives by testing our metric on neutral communities and communities where differences in species performance are due strictly to environmental filtering. From the former we expect no clustering, and from the latter we expect that species will cluster around the

favored trait. Thus we distinguish niche differentiation from neutrality and pure environmental filtering by the presence of multiple clusters as opposed to none or a single one.

Box 1: Detecting clusters with k-means and the gap statistic

Here we describe our metric based on coupling the k-means clustering algorithm [46] with the gap statistic method [48]. Conceptually, we assess the degree of clustering by measuring the total trait distance between species in the same cluster, summed over all clusters. Smaller values indicate tighter clustering. However, we cannot use this index directly unless we know the number of clusters to look for, as the index naturally declines with the number of clusters considered. To circumvent this, we compare this total distance to that found in null communities, and find the number of clusters that maximizes the difference from the null. This maximal difference is the statistic.

Stepwise, we proceed as follows. We start with a candidate number of clusters, k. Out of all possible ways to assign the observed species to k clusters, the k-means algorithm finds the one that minimizes within-cluster trait dispersion: $D_k = \sum_C \sum_{i,j \in C} n_i n_j d_{ij}$, where C refers to a cluster ($1 \leq C \leq k$), n_i is the abundance of species i, d_{ij} is the trait distance between species i and j. We define the clustering index $F_k = -\log(D_k)$, such that F_k is high for low dispersion D_k. We then repeat this step across a set of 5,000 null communities, and the difference in F_k between the observed community and the mean null value is the gap for k clusters, $G_k = F_k - \overline{F}_{k,\text{null}}$. We then test a range of numbers of clusters and find the one that maximizes the gap. This maximal value (the peak in the gap curve in Fig. B1B) is the gap statistic, $G = \max(G_k)$, and the value of k at which it occurs is the estimated number of clusters K. Note that the gap step is needed because F_k will always increase with k: as we increase the number of clusters, the average cluster becomes smaller and therefore within-cluster dispersion decreases. By comparing against null communities, the gap method selects the number of clusters that most increases the clustering index beyond those expectations.

We then perform the same routine on each of the null communities to obtain a null distribution of gap statistics, from which we extract significance (p-value) and standardized effect size (z-score). The z-score is $Z = (G - \mu)/\sigma$, where $\mu = \frac{1}{5000} \sum_{\text{nulls}} G_{\text{null}}$ is the mean of the null gap statistics and $\sigma^2 = \frac{1}{5000} \sum_{\text{nulls}} (G_{\text{null}} - \mu)^2$ is the variance. The p-value is the proportion of null communities with a higher gap statistic than the observed community, $P = \frac{1}{5000} \sum_{\text{nulls}} I(G_{\text{null}} > G)$, where the indicator function I is 1 if its argument is true, and zero otherwise.

Fig. B1 shows two test communities, and the respective results from our metric. The first community (Fig. B1A) is constructed to contain four clusters, whereas in the second community (Fig. B1C) abundances are unrelated to traits. The metric correctly diagnoses both (compare Fig. B1B, D).

Figure B1: **A**: Hypothetical species assemblage representing a community under niche differentiation + immigration, with visible clustering. **B**: Plotting the dispersion gap against the number of clusters, we see a clear peak at $K = 4$ clusters. The gap value at the peak, $G_4 = 1.66$, is the gap statistic. It far exceeds the 95th percentile of the gap statistic across the null communities (red line), leading to a very high z-score and a very low p-value. **C**: Neutral community with same size as A, but abundances unrelated to traits. **D**: Corresponding gap curve indicates that no number of clusters gives a gap beyond null expectations. Witness the low z-score and high p-value. [*Numerical details*: Community A: 200 species traits were uniformly drawn between 0 and 1. We then assigned a species to each of 1,000 individuals, with probability of choosing species i given by $P_i = k \sum_{j=1}^{4} \exp(-|x_i - x_j^*|^2/2\sigma^2)$, where x_i is the trait of species i, $\mathbf{x}^* = \{0.2, 0.3, 0.6, 0.9\}$ are the centers of the four clusters, $\sigma = 0.02$ is the cluster width, and k is a normalizing constant ensuring $\sum_{i=1}^{200} P_i = 1$. Finally, abundances are boosted by a random small amount to represent immigration. Community C: we drew species abundances from a log-series distribution with parameters $N = 1,000$ and $\alpha = 50$, and assigned them to randomly chosen trait values.]

Glossary

Cluster: a set of species with similar trait values and relatively high abundance, separated from other such sets by sparsely populated regions in trait space.

Competition-colonization tradeoff: a niche mechanisms commonly invoked for plant coexistence. Species may offset disadvantages in one-on-one contests for recruitment via superior ability to colonize freshly available habitat, or vice-versa [44].

Ecological drift: ecological analogue to genetic drift, whereby changes in the abundance of a species occur via random sampling from a finite population, independent of fitness or competitive standing related to species traits [35].

Limiting similarity: a classical theoretical concept that has taken on multiple meanings over time [5, 79], here interpreted as greater-than-chance trait differences between adjacent species on the niche or trait axis. While competing species were traditionally expected to display limiting similarity, the demonstrated robustness of clusters to demographic stochasticity, their persistence under immigration, and their appearance under disparate niche mechanisms suggest that clusters may be more common than limiting similarity.

Mass effects: phenomenon in source-sink dynamics whereby a locally disfavored population can be sustained via immigration from populations favored elsewhere [80].

Niche, niche strategy, niche differences: a concept in community ecology pertaining to idiosyncrasies in how a species responds to and in turn affects the environment, availability of resources, and/or other species [2]. Traditionally used to describe species differences in environmental preferences or in strategies for procuring resources, avoiding enemies, etc. In modern coexistence theory, 'niche differences' are operationally defined species differences that stabilize competitive coexistence, which requires that they make competition stronger within than between species[6, 7]. Stable competitive coexistence is when each species would grow from very low abundance in the presence of the others.

Niche axis: the axis of variation among species in how they respond to and in turn affect the environment, availability of resources, and/or other species. For example if species differ in their strategy of feeding on resources that can be arranged by size, one niche axis relevant to their interacton would be the resource axis. In contrast, in the competition-colonization tradeoff, the axis runs from competition specialist (high-quality low-number seeds) to colonization specialist (low-quality high-number seeds).

Niche mechanism: any set of circumstances of differences between species that, if large enough, could stabilize coexistence. In a niche mechanism, the nature of species interactions, plus sometimes the presence of tradeoffs, lead to opportunities for species to differ in their interaction with limiting factors [81]. Examples include specialization to different prey and/or different abiotic environments.

Optimal niche strategies: In a closed community, when the current number of species exceeds the number of species that can stably coexist, only those with what we here call 'optimal niche strategies' will persist indefinitely. These optimal niche strategies must be niche differentiated enough from one another (i.e. have enough space between them on the niche axis) to all stably coexist together. The particular set of mutually niche differentiated strategies that emerges from the competitive process as dominant can be determined in part by external circumstances such as which resources are in greater supply, or simply by the starting conditions or order of arrival of species to the community. Putting this phenomenon in mathematical terms, models of competition among species on niche axes typically have a pattern-forming instability [82], where the pattern formed is that of species abundances on the niche axis, with species with optimal strategies becoming abundant and other species decaying in abundance to zero.

Stabilization vs. equalization: a dichotomy in modern coexistence theory [7] whereby competitive exclusion can only be mitigated via mechanisms that either stabilize the community (niche mechanisms), or reduce average fitness differences between species (equalizing mechanisms). While the concept is a useful theoretical construct, it can be difficult or impossible to calculate the degree of stabilization and equalization in multispecies communities, or to assign trait differences to one or the other [61] (but see [83] for an update to the theory).

Acknowledgments

We thank David Ackerly for helpful discussions, and James O'Dwyer, Brian Sedio, Stephen Smith, and Lacey Knowles for comments on previous versions of the manuscript. This work was supported by the National Science Foundation under grant 1038678, "Niche versus neutral structure in populations and communities" funded by the Advancing Theory in Biology program, as well as by the Danish National Research Foundation through support of the Center of Macroecology, Evolution and Climate (grant DNRF96), and by the Miller Institute for Basic Research in Science at the University of California, Berkeley.

References

[1] Simberloff D, Boecklen W. Santa Rosalia Reconsidered: Size Ratios and Competition. Evolution. 1981;35(6):1206–1228.

[2] Chase JM, Leibold M. Ecological Niches: Linking Classical and Contemporary Approaches. The University of Chicago Press; 2003.

[3] Götzenberger L, de Bello F, Bråthen KA, Davison J, Dubuis A, Guisan A, et al. Ecological assembly rules in plant communities-approaches, patterns and prospects. Biological Reviews. 2012;87(1):111–127.

[4] HilleRisLambers J, Adler PB, Harpole WS, Levine JM, Mayfield MM. Rethinking Community Assembly through the Lens of Coexistence Theory. Annual Review of Ecology, Evolution, and Systematics. 2012 dec;43(1):227–248. Available from: `http://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-110411-160411`.

[5] D'Andrea R, Ostling AM. Challenges in linking trait patterns to niche differentiation. Oikos. 2016;125(10):1369–1385.

[6] Chesson P. A Need for Niches ? Trends in ecology & evolution. 1991;6(1):26–28.

[7] Chesson P. Mechanisms of maitenance of species diversity. Annual Review of Ecology and Systematics. 2000;31(2000):343–358.

[8] MacArthur RH, Levins R. The Limiting Similarity, Convergence, and Divergence of Coexisting Species. The American Naturalist. 1967;101(921):377–385.

[9] Grant PR. Colonization of islands by ecologically dissimilar species of birds bill length feeding. Canadian Journal of Zoology. 1969;47(1):41–43.

[10] Botta-Dukát Z. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. Journal of Vegetation Science. 2005;16(5):533–540. Available from: `http://doi.wiley.com/10.1111/j.1654-1103.2005.tb02393.x`.

[11] Laliberté E, Legendre P. A distance-based framework for measuring functional diversity from multiple traits. Ecology. 2010 jan;91(1):299–305. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/20380219`.

[12] de Bello F. The quest for trait convergence and divergence in community assembly: are null-models the magic wand? Global Ecology and Biogeography. 2011 jun;p. no–no. Available from: `http://doi.wiley.com/10.1111/j.1466-8238.2011.00682.x`.

[13] Stubbs WJ, Wilson JB. Evidence for limiting similarity in a sand dune community. Journal of Ecology. 2004 aug;92(4):557–567. Available from: `http://doi.wiley.com/10.1111/j.0022-0477.2004.00898.x`.

[14] Cornwell WK, Schwilk DW, Ackerly DD. A trait-based test for habitat filtering: Convex hull volume. Ecology. 2006;87(6):1465–1471.

[15] Ingram T, Shurin JB. Trait-based assembly and phylogenetic structure in northeast Pacific rockfish assemblages. Ecology. 2009 sep;90(9):2444–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19769123.

[16] Astor T, Strengbom J, Berg MP, Lenoir L, Marteinsdottir B, Bengtsson J. Underdispersion and overdispersion of traits in terrestrial snail communities on islands. Ecology and Evolution. 2014;4(11):2090–2102.

[17] Pontarp M, Petchey OL. Community trait overdispersion due to trophic interactions : concerns for assembly process inference. Proc Roy Soc Lond B. 2016;283:20161729.

[18] Scheffer M, van Nes EH. Self-organized similarity, the evolutionary emergence of groups of similar species. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(16):6230–6235.

[19] Fort H, Scheffer M, van Nes EH. The paradox of the clumps mathematically explained. Theoretical Ecology. 2009;2(3):171–176.

[20] Fort H, Scheffer M, van Nes EH. The clumping transition in niche competition: a robust critical phenomenon. Journal of Statistical Mechanics: Theory and Experiment. 2010 may;2010(05):P05005. Available from: http://stacks.iop.org/1742-5468/2010/i=05/a=P05005?key=crossref.863eb86a4d76b0cdc7127e9b4a89999f.

[21] Scranton K, Vasseur DA. Coexistence and emergent neutrality generate synchrony among competitors in fluctuating environments. Theoretical Ecology. 2016;9(3):353–363. Available from: http://dx.doi.org/10.1007/s12080-016-0294-z.

[22] D'Andrea R, Ostling A. Biodiversity maintenance may be lower under partial niche differentiation than under neutrality. Ecology. 2017;98(12):3211–3218.

[23] Scheffer M, van Nes EH, Vergnon R. Toward a unifying theory of biodiversity. Proceedings of the National Academy of Sciences. 2018;115(4):201721114. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1721114115.

[24] Ernebjerg M, Kishony R. Dynamic phenotypic clustering in noisy ecosystems. PLoS computational biology. 2011 mar;7(3):e1002017. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3060162{\&}tool=pmcentrez{\&}rendertype=abstract.

[25] Sakavara A, Tsirtsis G, Roelke DL, Mancy R, Spatharis S. Lumpy species coexistence arises robustly in fluctuating resource environments. Proceedings of the National Academy of Sciences. 2018;Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1705944115.

[26] Keddy PA. Assembly and response rules: two goals for predictive community ecology. Journal of Vegetation Science. 1992;3(2):157–164. Available from: http://apps.webofknowledge.com/full{_}record.do?product=UA{\&}search{_}mode=GeneralSearch{\&}qid=2{\&}SID=W1pPLzAZjdAHkE5rN3O{\&}page=1{\&}doc=1.

[27] Mayfield MM, Levine JM. Opposing effects of competitive exclusion on the phylogenetic structure of communities. Ecology letters. 2010 sep;13(9):1085–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20576030.

[28] Vergnon R, Dulvy NK, Freckleton RP. Niches versus neutrality: uncovering the drivers of diversity in a species-rich community. Ecology letters. 2009 oct;12(10):1079–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19747181.

[29] Segura AM, Calliari D, Kruk C, Conde D, Bonilla S, Fort H. Emergent neutrality drives phytoplankton species coexistence. Proceedings Biological sciences / The Royal Society. 2011 aug;278(1716):2355–61. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3119015{\&}tool=pmcentrez{\&}rendertype=abstract.

[30] Segura AM, Kruk C, Calliari D, García-Rodriguez F, Conde D, Widdicombe CE, et al. Competition Drives Clumpy Species Coexistence in Estuarine Phytoplankton. Scientific Reports. 2013;3(1):1037. Available from: http://www.nature.com/articles/srep01037.

[31] Holt RD. Emergent neutrality. Trends in Ecology & Evolution. 2006 oct;21(10):531–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16901580.

[32] Gaston KJ, Chown SL. Neutrality and the niche. Functional Ecology. 2005;19(1):1–6.

[33] Vellend M. Conceptual synthesis in community ecology. The Quarterly review of biology. 2010 jun;85(2):183–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20565040.

[34] Rosindell J, Hunter MO, Etienne RS. The unified neutral theory of biodiversity and biogeography at age ten. Trends in ecology & evolution. 2011 jul;26(7):340–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21561679.

[35] Hubbell SP. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton, NJ: Princeton University Press; 2001.

[36] Tilman D. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. Proceedings of the National Academy of Sciences of the United States of America. 2004 jul;101(30):10854–61. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=503710{\&}tool=pmcentrez{\&}rendertype=abstract.

[37] Schwilk DW, Ackerly DD. Limiting similarity and functional diversity along environmental gradients. Ecology Letters. 2005 jan;8(3):272–281. Available from: http://doi.wiley.com/10.1111/j.1461-0248.2004.00720.x.

[38] Gravel D, Canham CD, Beaudet M, Messier C. Reconciling niche and neutrality: the continuum hypothesis. Ecology Letters. 2006 apr;9(4):399–409. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16623725.

[39] Roelke DL, Eldridge PM. Mixing of supersaturated assemblages and the precipitous loss of species. The American Naturalist. 2008 feb;171(2):162–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18197769.

[40] O'Dwyer JP. Whence Lotka-Volterra? Theoretical Ecology. 2018;p. 1–12. Available from: http://link.springer.com/10.1007/s12080-018-0377-0.

[41] MacArthur RH. Geographical ecology; patterns in the distribution of species. Harper & Row; 1972.

[42] Abrams P. Limiting Similarity and the Form of the Competition Coefficient. Theoretical population biology. 1975;8:356–375.

[43] Abrams P, Rueffler C, Dinnage R. Competition-similarity relationships and the nonlinearity of competitive effects in consumer-resource systems. The American Naturalist. 2008 oct;172(4):463–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18717636.

[44] Levins R, Culver D. Regional Coexistence of Species and Competition between Rare Species. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(6):1246–1248.

[45] Muller-Landau HC. The tolerance-fecundity trade-off and the maintenance of diversity in seed size. Proceedings of the National Academy of Sciences of the United States of America. 2010 mar;107(9):4242–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2840174{\&}tool=pmcentrez{\&}rendertype=abstract.

[46] MacQueen JB. Some Methods for classification and Analysis of Multivariate Observations. 5th Berkeley Symposium on Mathematical Statistics and Probability 1967. 1967;1(233):281–297. Available from: http://projecteuclid.org/euclid.bsmsp/1200512992.

[47] Ripley BD. The Second-Order Analysis of Stationary Point Processes. Journal of Applied Probability. 1976;13(2):255–266.

[48] Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001;63:411–423. Available from: http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00293/abstract.

[49] Yang Z, Powell J, Zhang C, Du G. The effect of environmental and phylogenetic drivers on community assembly in an alpine meadow community. Ecology. 2012;93(11):2321–2328.

[50] Rosenzweig ML, MacArthur RH. Graphical Representation and Stability Conditions of Predator-Prey Interactions. American Naturalist. 1963;97(895):209–223.

[51] Abrams P, Rueffler C. Coexistence and limiting similarity of consumer species competing for a linear array of resources. Ecology. 2009 mar;90(3):812–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19341150.

[52] Leimar O, Sasaki A, Doebeli M, Dieckmann U. Limiting similarity, species packing, and the shape of competition kernels. Journal of Theoretical Biology. 2013;339:3–13.

[53] Harner RF, Harper KT. The Role of Area , Heterogeneity , and Favorability in Plant Species Diversity of Pinyon- Juniper Ecosystems. Ecology. 1976;57:1254–1263.

[54] Grubb PJ. the Maintenance of Species-Richness in Plant Communities: the Importance of the Regeneration Niche. Biological Reviews. 1977;52(1):107–145. Available from: http://doi.wiley.com/10.1111/j.1469-185X.1977.tb01347.x.

[55] Comins HN, Noble IR. Dispersal, Variability, and Transient Niches: Species Coexistence in a Uniformly Variable Environment. The American Naturalist. 1985;126(5):617–640.

[56] Amarasekare P. Competitive coexistence in spatially structured environments: A synthesis. Ecology Letters. 2003;6(12):1109–1122.

[57] Tilman D. Resource competition and community structure. Monographs ed. Princeton University Press; 1982.

[58] Kraft NJB, Ackerly DD. Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. Ecological Monographs. 2010 aug;80(3):401–422. Available from: http://www.esajournals.org/doi/abs/10.1890/09-1672.1.

[59] Hernández-García E, López C, Pigolotti S, Andersen KH. Species competition: coexistence, exclusion and clustering. Philosophical Transactions of the Royal Society A. 2009 aug;367(1901):3183–95. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3263774{\&}tool=pmcentrez{\&}rendertype=abstract.

[60] D'Andrea R, Ostling A, O'Dwyer JP. Translucent windows: How uncertainty in competitive interactions impacts detection of community pattern. Ecology Letters. 2018;p. 826–835.

[61] Barabás G, D'Andrea R, Stump SM. Chesson's coexistence theory. Ecological Monographs. 2018 apr;Available from: http://doi.wiley.com/10.1002/ecm.1302.

[62] Chesson P. Multispecies Competition in Variable Environments. Theoretical Population Biology. 1994;45(3):227–276. Available from: http://www.sciencedirect.com/science/article/pii/S0040580984710136.

[63] Barabás G, D'Andrea R, Rael R, Meszéna G, Ostling AM. Emergent Neutrality or Hidden Niches? Oikos. 2013;122:1565–1572.

[64] Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, et al. The metacommunity concept: A framework for multi-scale community ecology. Ecology Letters. 2004;7(7):601–613.

[65] D'Andrea R, Ostling AM. Can Clustering in Genotype Space Reveal "Niches"? The American Naturalist. 2016;187(1):130–135. Available from: http://www.jstor.org/stable/info/10.1086/684116{\%}5Cnhttp://www.journals.uchicago.edu/doi/10.1086/684116.

[66] Houchmandzadeh B. Neutral aggregation in finite-length genotype space. Physical Review E. 2017;95(1):1–12.

[67] Clark JS, Dietze M, Chakraborty S, Agarwal PK, Ibanez I, LaDeau S, et al. Resolving the biodiversity paradox. Ecology letters. 2007 aug;10(8):647–59; discussion 659–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17594418.

[68] Leigh EG, Davidar P, Dick CW, Puyravaud JP, Terborgh J, ter Steege H, et al. Why do some tropical forests have so many species of trees? Biotropica. 2004;36(March 2003):447–473. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1744-7429.2004.tb00342.x/abstract.

[69] Seidler TG, Plotkin JB. Seed dispersal and spatial pattern in tropical trees. PLoS Biology. 2006;4(11):2132–2137.

[70] Chesson P, Warner RR. Environmental Variability Promotes Coexistence in Lottery Competitive Systems. The American Naturalist. 1981;117(6):923–943.

[71] Chisholm RA, Lichstein JW. Linking dispersal, immigration and scale in the neutral theory of biodiversity. Ecology Letters. 2009 dec;12(12):1385–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19807774.

[72] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018. Available from: https://www.r-project.org/.

[73] Barabás G, D'Andrea R, Ostling AM. Species packing in nonsmooth competition models. Theoretical Ecology. 2013 jan;6(1):1–19. Available from: http://www.springerlink.com/index/10.1007/s12080-011-0151-z.

[74] D'Andrea R, Barabás G, Ostling AM. Revising the tolerance-fecundity trade-off; or, on the consequences of discontinuous resource use for limiting similarity, species diversity, and trait dispersion. The American Naturalist. 2013;181(4):E91–101. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23535625.

[75] Haegeman B, Sari T, Etienne RS. Predicting coexistence of plants subject to a tolerance-competition trade-off. Journal of Mathematical Biology. 2014;68(7):1815–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23728210.

[76] Dixon PM. Ripley's K function. In: El-shaarawi AH, Piegorsch WW, editors. Encyclopedia of Environmetrics. vol. 3. Chichester: John Wiley & Sons; 2002. p. 1796–1803.

[77] D'Andrea R. Code for clustering metrics; 2018. Available from: https://github.com/rafaeldandrea/Clustering-metric.

[78] Etienne RS, Alonso D, McKane AJ. The zero-sum assumption in neutral biodiversity theory. Journal of theoretical biology. 2007 oct;248(3):522–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17640675.

[79] Abrams P. The Theory of Limiting Similarity. Annual Review of Ecology and Systematics. 1983 nov;14(1):359–376.

[80] Shmida A, Mark V Wilson. Biological Determinants of Species Diversity. Journal of Biogeography. 1985;12(1):1–20.

[81] Levin S. Community equilibria and stability, and an extension of the competitive exclusion principle. The American Naturalist. 1970;104(939):413–423.

[82] Pigolotti S, López C, Hernández-García E. Species Clustering in Competitive Lotka-Volterra Models. Physical Review Letters. 2007 jun;98(25):1–4. Available from: http://link.aps.org/doi/10.1103/PhysRevLett.98.258101.

[83] Chesson P. Updates on mechanisms of maintenance of species diversity. Journal of Ecology. 2018;106:1773–1794. Available from: https://besjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/1365-2745.13035.

[84] Kinzig aP, Levin S, Dushoff J, Pacala S. Limiting Similarity, Species Packing, and System Stability for Hierarchical Competition-Colonization Models. The American Naturalist. 1999 apr;153(4):371–383. Available from: http://www.jstor.org/stable/10.1086/303182.

[85] Yan M, Ye K. Determining the number of clusters using the weighted gap statistic. Biometrics. 2007;63(4):1031–1037.

Supplementary Information

Supplementary Figures

Figure S1: **Lotka-Volterra niche scenarios with environmental filtering**. **A**: We implemented environmental filters via a modal relationship between intrinsic growth rate r and species traits x_i as follows: $r = 0.5$ (no filtering, black); $r_i = x_i(1 - x_i)$ (red); $r_i = \exp(-(x_i - 0.5)^2/\sigma^2)$ with $\sigma = 0.5$ (green), 0.2 (blue), 0.1 (magenta). **B**: Under no filtering, the community shows no overarching abundance trend. **C-F**: Under increasing filter intensity, communities show increasingly steeper abundance trends. Our metrics detected multiple clusters in 10/10 replicates of B-D, but only 2/10 in E and 1/10 in F, with the remaining having a single cluster.

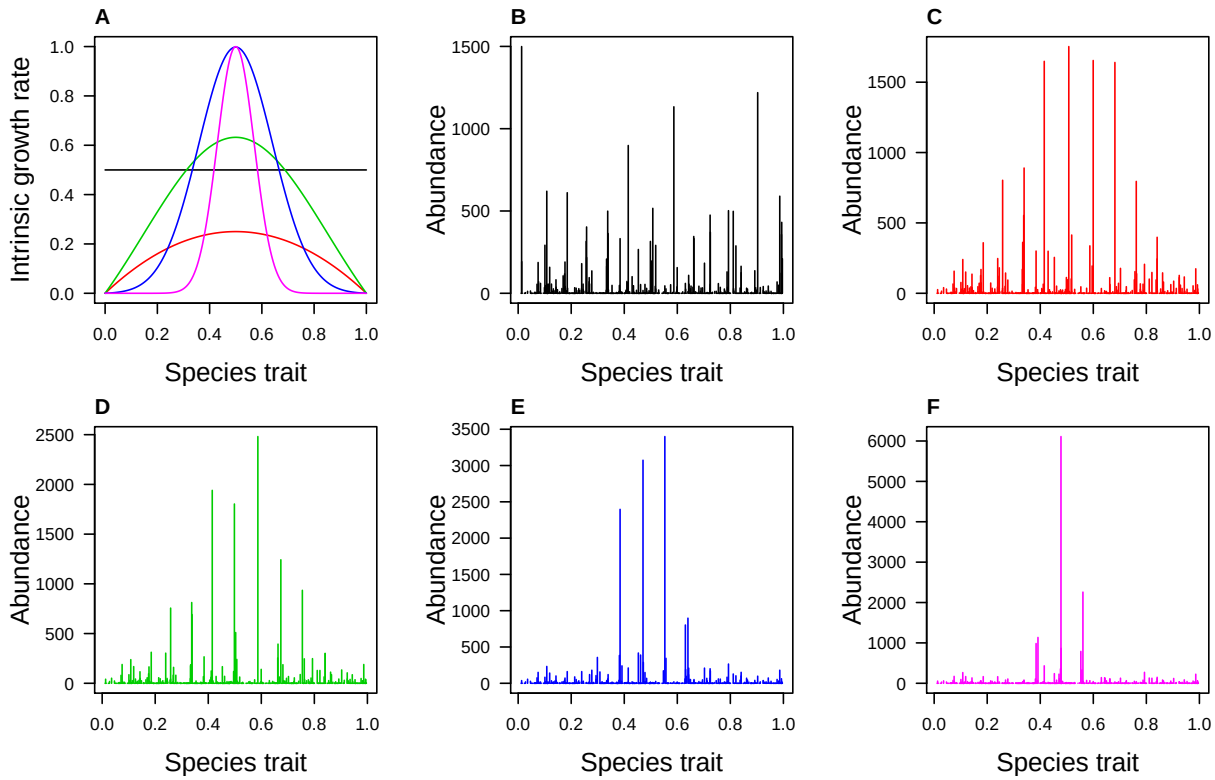


Figure S2: **Communities under lower immigration**, $m = 0.01$. Compared with $m = 0.08$ (Fig. 2), communities show a similar number of clusters but fewer species per cluster, and gaps between clusters are more rarified.

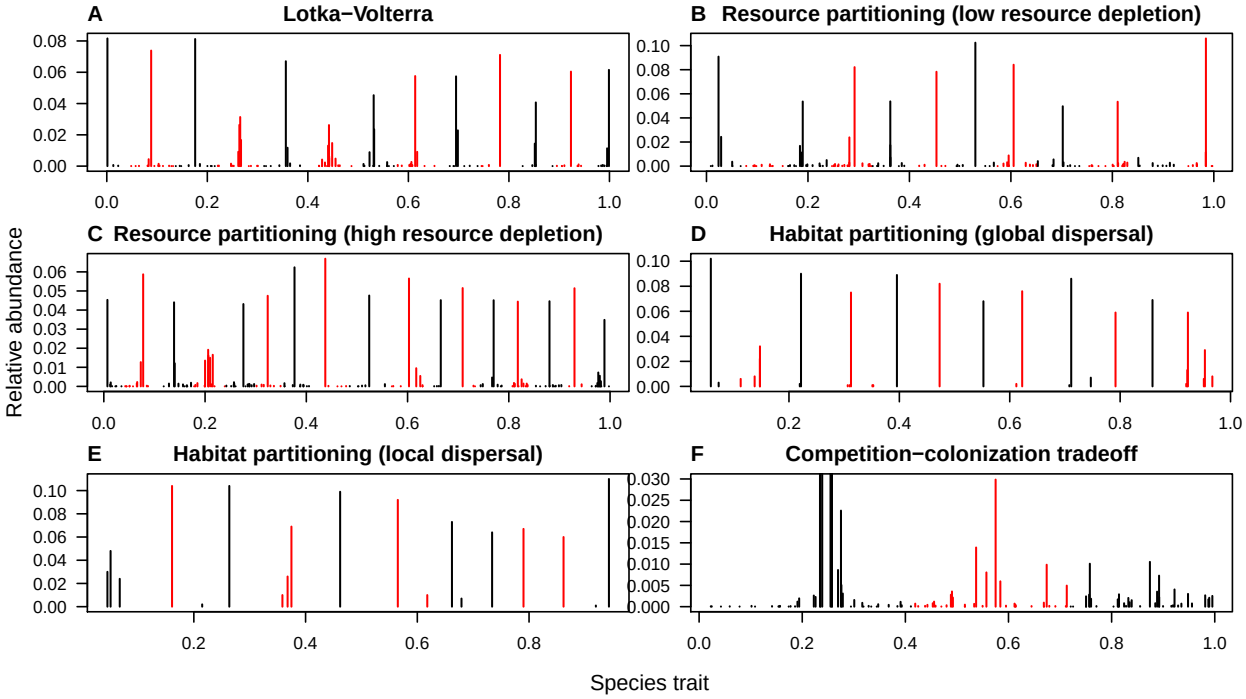


Figure S3: **Clustering results under lower immigration**, $m = 0.01$. Results are similar to $m = 0.08$ (compare with Fig. 3), although z-scores and significance are often higher, particularly for resource-consumer communities under low resource depletion. This indicates that higher immigration is drowning the pattern.

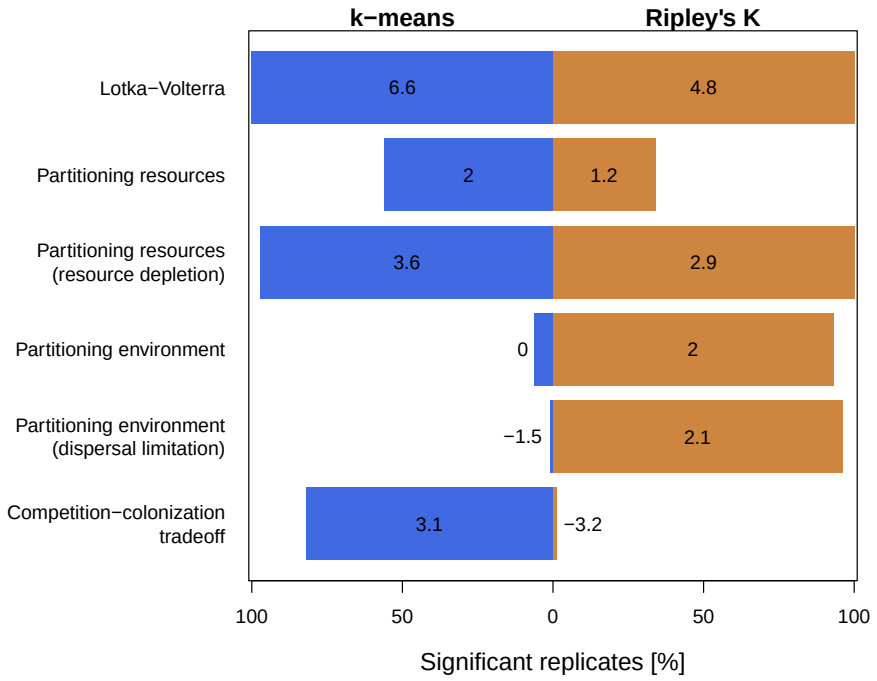


Figure S4: **Effects of resource depletion.** Example communities where consumers partition resources, under low and high resource depletion (left and right columns, respectively). Resources are shown on top, consumers at the bottom. Under high resource depletion, gaps left by resource extirpation cause corresponding gaps among consumer species. These gaps potentially explain why clustering was stronger in this scenario than under low depletion. ($m = 0.08$, c. 400 regional species.)

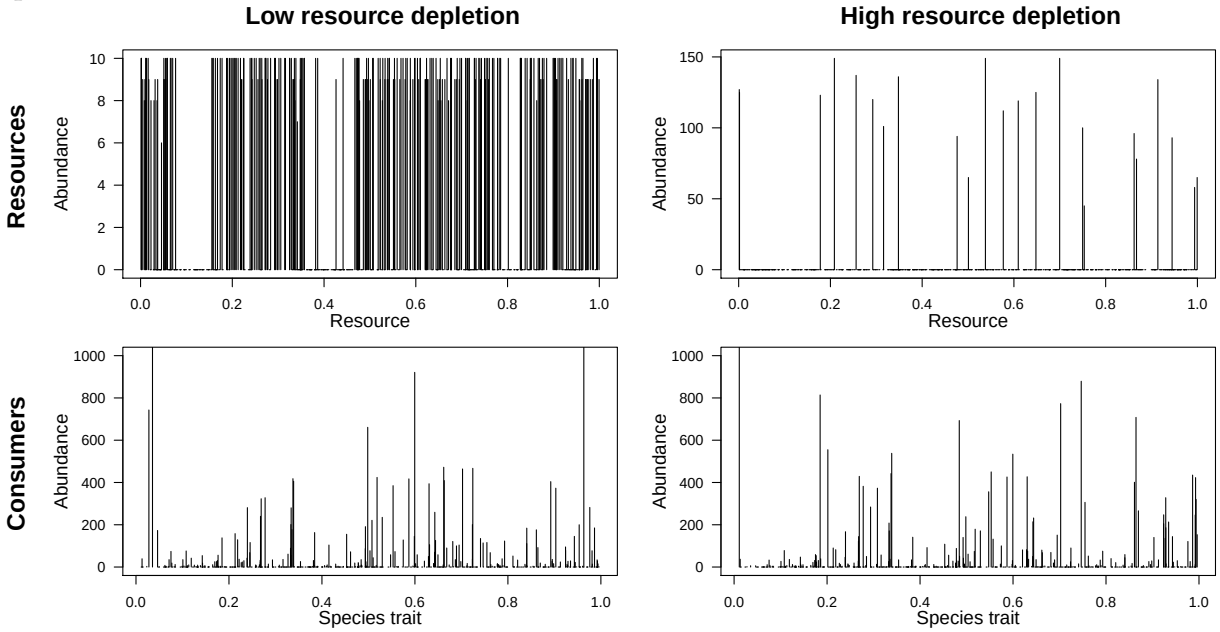


Figure S5: **Competition-colonization tradeoff, deterministic implementation without immigration.** **A**. Transient state shows visible clustering. **B**. Equilibrium state, showing abundances of the coexisting species, which are those that dominate their respective clusters in the transient state (marked with red dots in A). Legend shows maximum eigenvalue of the Jacobian of the equilibrium, indicating dynamical stability of the equilibrium. **C**. Competition kernel shows the strongest competitors on focal species, whose traits are shown in legend, are species with similar traits, despite the competitive hierarchy. This contributes to stabilization of the community. **D**. For each species (traits plotted on x-axis), the position of the peak of the competition kernel (i.e. the species with the strongest net competitive impact on it) is plotted on the y-axis. The proximity of the curve to the one-to-one line (dotted line) throughout the trait axis shows that the competition-colonization tradeoff stabilizes the community, thus acting as a niche mechanism. The kernel maxima plotted here also explain the absence of coexisting species below trait $\simeq 0.2$, as the strongest competitors of those species have higher fecundity than themselves (solid curve is above dotted line), thus being both more competitive and more fecund. The wide gap between the first cluster and the other clusters is also reflected in the relatively larger distance from the one-to-one curve in that region of the trait axis.

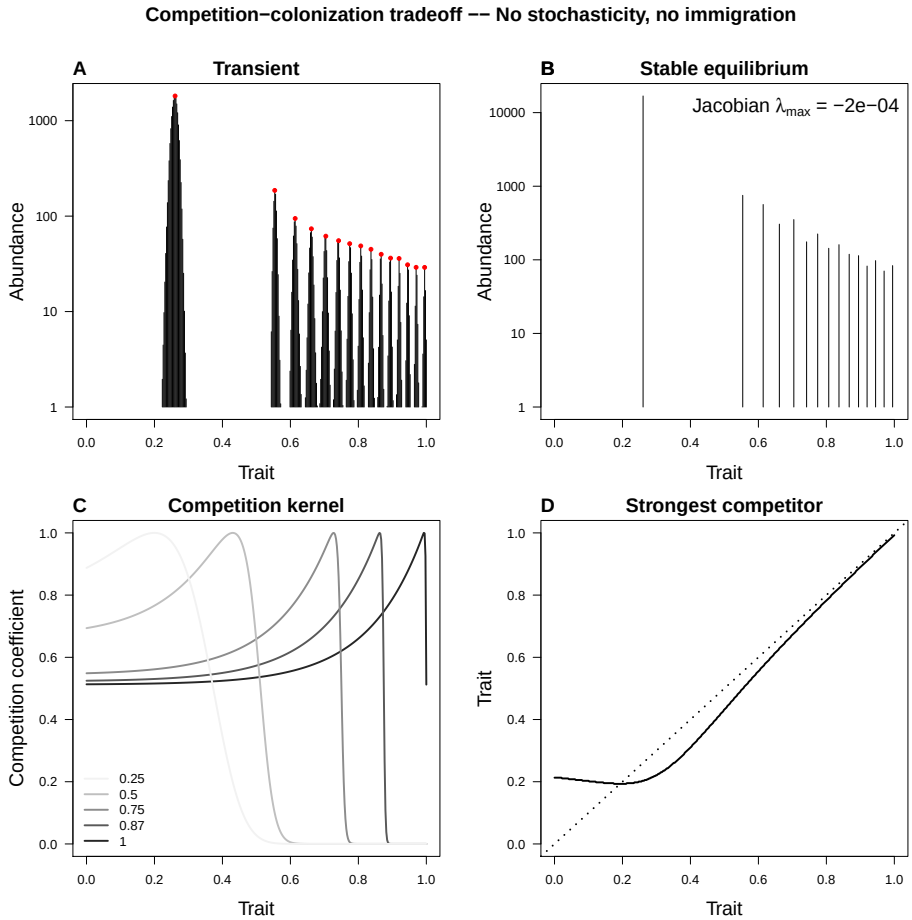
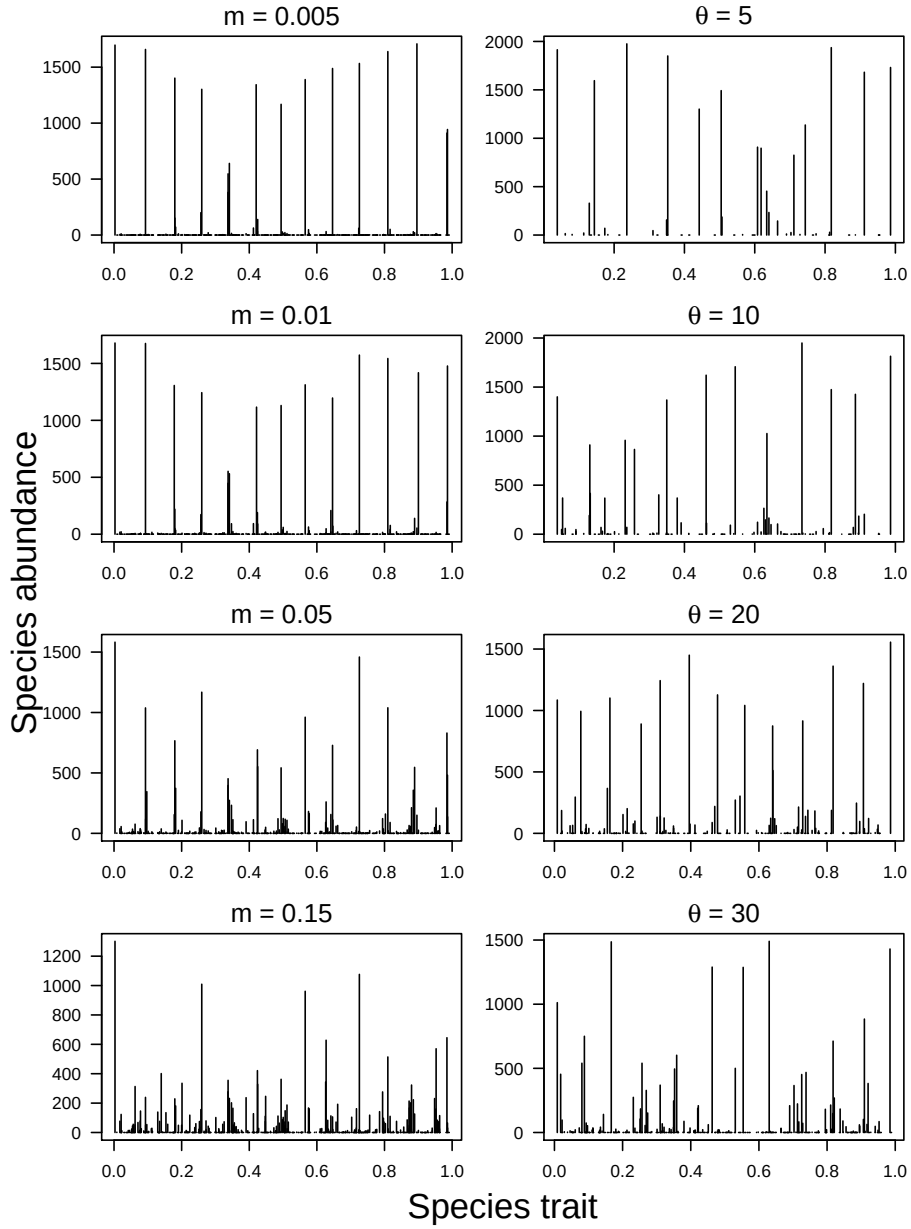


Figure S6: **Effects of immigration rates and regional diversity.** Example Lokta-Volterra communities with increasing immigration rate m (left) and regional diversity parameter θ (right).



Additional Methods

Competition-colonization tradeoff model

This model was originally formulated as follows [84]

$$\frac{1}{p_i}\frac{\mathrm{d}p_i}{\mathrm{d}t} = f_i(1 - \sum_{j=1}^{i} p_j) - \sum_{j=1}^{i-1} f_j p_j - \mu, \tag{S1}$$

where p_i is species i's relative abundance, and f_i is its fecundity (i.e. number of propagules per individual per unit time). The first term represents recruitment in available sites—i.e. all sites not currently occupied by species i or stronger competitors; the second term represents displacement by stronger competitors; the last term is intrinsic mortality μ, here assumed identical for all species.

We can gather the density-independent and the density-dependent terms separately:

$$\frac{1}{p_i}\frac{\mathrm{d}p_i}{\mathrm{d}t} = (f_i - \mu) - \sum_{j=1}^{S} \Theta_{ij}(f_i + f_j)p_j, \tag{S2}$$

The step function Θ_{ij} (equal to 1 if $i > j$, 0 if $i < j$, and 0.5 if $i = j$) encodes the strict competitive hierarchy. It has been shown [74] that the strict hierarchy is unrealistic (because arbitrarily similar species will have large differences in competitive ability) and drastically inflates coexistence. Here we use instead a probabilistic, gradual hierarchy: the higher the difference in rank, the higher the likelihood of displacement by the better competitor. We do this by replacing Θ_{ij} with the continuous function $\Gamma_{ij} = 0.5\,(1 - \tanh[s(f_j - f_i)])$, which is equal to 0.5 when $f_i = f_j$ and asymptotes to 1 and 0 when $f_i \gg f_j$ and $f_i \ll f_j$, respectively. The parameter s controls the steepness of the hierarchy, and hence the degree of coexistence. (We recover the step function when $s \to \infty$.)

Notice that Equation (S2) is in the Lotka-Volterra shape, $\frac{1}{N_i}\frac{\mathrm{d}N_i}{\mathrm{d}t} = r_i - \sum_j A_{ij}N_j$, where the effective intrinsic growth rate $r_i = b_i - d_i$ is the difference between intrinsic fecundity and mortality. We use that correspondence in our stochastic formulation, where we place the density dependence in the deaths.

Detailed description of clustering metrics

Our method proceeds in the following general steps:

1. Given the data and a certain parameter value φ within a provided range $[\varphi_{\min}, \varphi_{\max}]$, calculate a clustering measure $F(\varphi)$.

2. Repeat step 1 in each of n null communities, $\widetilde{F}_i(\varphi)$, $i = 1, 2, \ldots, n$, and take the mean, $\overline{\widetilde{F}}(\varphi) = \frac{1}{n}\sum_i \widetilde{F}_i(\varphi)$.

3. The difference $G(\varphi) = F(\varphi) - \overline{\widetilde{F}}(\varphi)$ is the *gap* at that parameter value.

4. Repeat steps 1-3 for all parameter values within $[\varphi_{\min}, \varphi_{\max}]$.

5. The gap statistic is the extremum (maximum or minimum) of the gap function, $G = \underset{\varphi}{\text{extr}}\, G(\varphi)$.

6. Obtain a z-score and a p-value by repeating steps 1-5 on each of the null communities, which provides a null distribution for G. The z-score is then $Z = (G - \mu_{\widetilde{G}}) / \sigma_{\widetilde{G}}$ and the p-value is $P = \frac{1}{n}\sum_i I(\widetilde{G}_i \geq G)$, where $\widetilde{G}_i$ is the gap statistic obtained for the i-th null assemblage, $\mu_{\widetilde{G}}$ and $\sigma_{\widetilde{G}}$ are the mean and standard deviation of those values, and index function I is 1 if its argument is true and 0 otherwise.

Note that this recipe works with different clustering measures, F. Here we use two: k-means dispersion [46] and Ripley's K function [76]. In the k-means version, the parameter φ is the number of clusters, and the value $\hat{\varphi}$ at which the gap is maximal is the estimated number of clusters in the community. In the Ripley's K version, φ is the trait distance between pairs of individuals, and the value at which the gap is minimal is the average distance between clusters.

Note also that we can use any number of traits to describe our species, as long as we can define a "distance" between species (e.g. Euclidean distance in a high-dimensional trait space, or simple trait differences on a single trait axis).

Gap statistic via k-means

Here the parameter φ is the candidate number of clusters k, and we use the k-means clustering algorithm [46]. For a given k, the algorithm finds the partition of individuals into k groups that minimizes the within-group dispersion. Let D_k be the total pairwise squared distance between members of each group, i.e., $D_k = \sum_{C=1}^{k}\sum_{i,j\in C} n_i n_j d_{ij}$, where C refers to a cluster, n_i is the abundance of species i, d_{ij} is the trait distance between species i and j. The k-means algorithm assigns each species to one of the k groups so that D_k is as low as possible[2,3]. We then set $F(\varphi) = F_k = \log(1/D_k)$ as the clustering measure[4]. See Box 1 for a step-by-step recipe for this metric.

Gap statistic via Ripley's K

The approach with Ripley's K is to count the number of pairs of individuals within each distance in trait space and determine whether or not it deviates significantly from what we would expect if the same set of abundances were randomly distributed among the present species. In particular, we look for significant open spaces between clusters, i.e. distances at which Ripley's K is surprisingly low. The parameter φ here is therefore the trait distance d.

[2]Dispersion can be defined in different ways, depending on how one wishes to weigh the number of members per group. While [48] choose to scale the squared distances by the number of individuals in the group, [85] argue that scaling it by the number of pairs in the group is a a better choice in some cases. In our study we chose not to scale the square distances, as that choice proved more robust to environmental filtering.

[3]In our model, all conspecific individuals have the same trait value, and therefore necessarily belong in the same cluster. Species are therefore trivial clusters of individuals. Ecologists are of course interested in clustering beyond the obvious grouping of individuals into species.

[4]Note that it does not make sense to compare F_k directly across different k's because F_k necessarily increases with k, as the average within-cluster distance is always lower for higher numbers of clusters. By comparing against null assemblages, the gap method finds the biggest increase in goodness of fit *beyond* what is expected from the increase in k. The reason we use $\log(1/D_k)$ rather than simply $1/D_k$ is that the expected increments in $1/D_k$ with increasing k are multiplicative rather than additive (see [48] for more details).

Ripley's K function at distance d is defined as $K(d) = \frac{\sum_{i \neq j} I(d_{ij} < d) N_i N_j}{\sum_{i \neq j} N_i N_j}$, where N_i is the abundance of species i, d_{ij} is the distance in trait space between species i and j, and function $I(.)$ is the indicator function, equal to 1 if its argument is true and 0 otherwise. We then calculate Ripley's K in each of our null assemblages, and define the Ripley's K clustering measure as the standardized K function: $F(d) = K(d)/\sigma_{\widetilde{K}(d)}$, where $\sigma_{\widetilde{K}(d)}$ is the standard deviation of the null values. The standardization compensates for the natural increase in the variance of $K(d)$ at large d (see Fig. S7). The gap function is then $G(d) = K(d)/\sigma_{\widetilde{K}(d)} - \mu_{\widetilde{K}(d)}/\sigma_{\widetilde{K}(d)}$.

Because we are interested in significantly low counts, we define our Ripley gap statistic as the minimum value of $G(d)$ across candidate distances, $G = \min_d G(d)$. This is analogous to our implementation with the k-means algorithm (see Table S1 for a direct comparison between the two implementations).

Fig. S7 shows the metric at work on an example of a Lotka-Volterra community, a community under neutral competition and environmental filtering, and a purely neutral community. When the community has significant open spaces (Fig. S7A), i.e. a lower G than the null communities, it suggests multiple clusters in trait space of high abundance species, with a dearth of species in between them. This is what we would expect from a community shaped by niche differentiation. When instead it is significantly lacking in gaps (Fig. S7B), that indicates a single clump of abundant species, as could be expected when the environment filters for a single trait value.

Note that while k-means and Ripley's K can characterize a community in terms of clustering structure, the gap statistic tells us whether that structure is significant compared to a null model. Furthermore, while clustering measures will typically depend on a parameter (number of clusters in the case of k-means, distance between clusters in the case of Ripley's K), the gap statistic removes the parameter by comparing the data to the null model across the parameter's range.

Table S1: **Comparison of our two implementations of the gap method**. Both apply the gap method using different measures of dispersion. $W(k)$ is the k-means dispersion for k clusters; $K(d)$ is Ripley's K at trait distance d; $F(k)$ ($F(d)$) is the dispersion for k clusters (distance d); I is the indicator function, equal to 1 if its argument is true and zero otherwise; N_i is the abundance of species i; $G(k)$ ($G(d)$) is the gap function for k clusters (distance d); $\mu_{\widetilde{X}}$ and $\sigma_{\widetilde{X}}$ are the mean and standard deviation of quantity X taken across the null communities; $\hat{d}$ is the distance at which $G(d) = G$. n is the number of null communities.

	k-means	**Ripley's K**
Clustering measure	$F_k = \log(1/W(k))$	$F(d) = K(d)/\widetilde{\sigma}_K(d)$
Gap function	$G_k = F_k - \mu_{\widetilde{F}_k}$	$G(d) = F(d) - \mu_{\widetilde{F}(d)}$
Gap statistic	$G = \max_k G_k$	$G = \min_d G(d)$
Estimated number of clusters	k such that $G_k = G$	Number of abundance-peaks within distance $\hat{d}$ of each other
Z-score	$Z = (G - \mu_{\widetilde{G}}) \,/\, \sigma_{\widetilde{G}}$	$Z = (G - \mu_{\widetilde{G}}) \,/\, \sigma_{\widetilde{G}}$
P-value	$P = \frac{1}{n}\sum_i I(\widetilde{G}_i \geq G)$	$P = \frac{1}{n}\sum_i I(\widetilde{G}_i \geq G)$

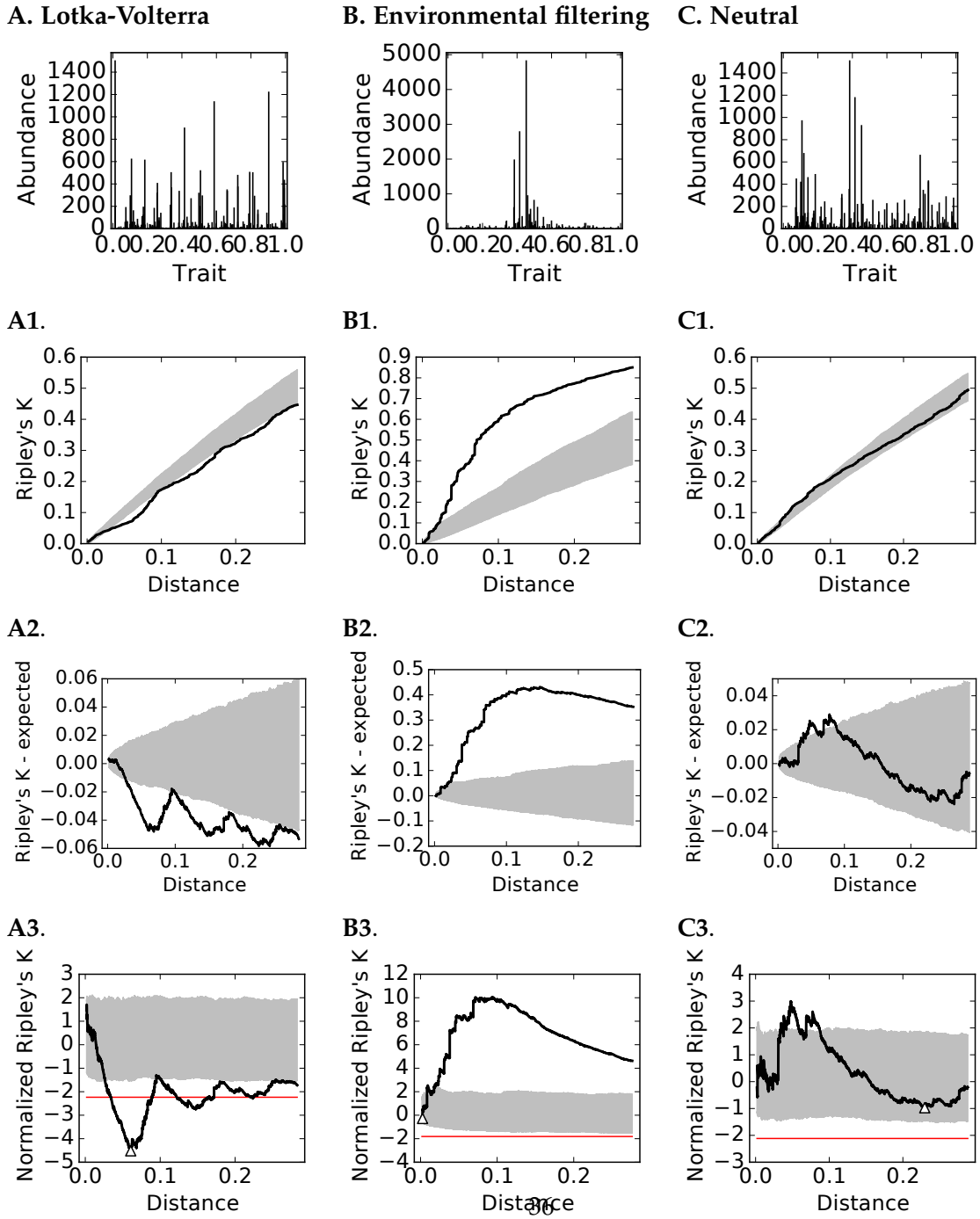



Figure S7: **Top**: Abundances plotted against trait values in a sample replicate of (**A**) the Lotka-Volterra niche model, (**B**) environmental filtering without a niche mechanism, and (**C**) the neutral model. **A1-C1**: Ripley's K vs. trait distance in the same replicates. Values for the observed community are plotted in black. The %95 confidence interval of $K(d)$ among the null communities (obtained by reshuffling abundances across species) is shaded in gray. **A2-C2**: same results with mean null value at each distance subtracted out, $K(d) - \widetilde{\mu}_K(d)$. **A3-C3**: rescaled by the variance at each distance $G(d) = K(d)/\widetilde{\sigma}_K(d) - \widetilde{\mu}_K(d)/\widetilde{\sigma}_K(d)$. The red line is the threshold for a significantly low gap statistic ($p < 0.05$), and the triangle indicates the distance with the most surprisingly low density of pairs, i.e. the distance d where the Ripley gap statistic G is achieved, $G(d) = G$.