

# Observed Ecological Communities Are Formed by Species Combinations That Are among the Most Likely to Persist under Changing Environments

Lucas P. Medeiros,<sup>1</sup> Karina Boege,<sup>2</sup> Ek del-Val,<sup>3</sup> Alejandro Zaldívar-Riverón,<sup>4</sup> and Serguei Saavedra<sup>1,\*</sup>

1. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; 2. Instituto de Ecología, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Ciudad Universitaria, 04510 Mexico City, Mexico; 3. Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701, 58190 Morelia, Michoacán, Mexico; 4. Instituto de Biología, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Ciudad Universitaria, 04510 Mexico City, Mexico

Submitted February 3, 2020; Accepted July 21, 2020; Electronically published November 18, 2020

Online enhancement: appendix.

**ABSTRACT:** Despite the rich biodiversity found in nature, it is unclear to what extent some combinations of interacting species, while conceivable in a given place and time, may never be realized. Yet solving this problem is important for understanding the role of randomness and predictability in the assembly of ecological communities. Here we show that the specific combinations of interacting species that emerge from the ecological dynamics within regional species pools are not all equally likely to be seen; rather, they are among the most likely to persist under changing environments. First, we use niche-based competition matrices and Lotka-Volterra models to demonstrate that realized combinations of interacting species are more likely to persist under random parameter perturbations than the majority of potential combinations with the same number of species that could have been formed from the regional pool. We then corroborate our theoretical results using a 10-year observational study, recording 88 plant-herbivore communities across three different forest successional stages. By inferring and validating plant-mediated communities of competing herbivore species, we find that observed combinations of herbivores have an expected probability of species persistence higher than half of all potential combinations. Our findings open up the opportunity to establish a formal probabilistic and predictive understanding of the composition of ecological communities.

**Keywords:** community composition, environmental change, niche framework, persistence, probability, species pool, structural stability.

\* Corresponding author; email: sersaa@mit.edu.

**ORCIDs:** Medeiros, <https://orcid.org/0000-0002-0320-5058>; Boege, <https://orcid.org/0000-0003-2971-080X>; del-Val, <https://orcid.org/0000-0003-3862-1024>; Zaldívar-Riverón, <https://orcid.org/0000-0001-5837-1929>; Saavedra, <https://orcid.org/0000-0003-1768-363X>.

Am. Nat. 2021. Vol. 197, pp. E000–E000. © 2020 by The University of Chicago. 0003-0147/2021/19701-5975\$15.00. All rights reserved.

DOI: 10.1086/711663

## Introduction

The Earth contains millions of species that can form ecological communities, but only a small fraction of all possible combinations of species are observed in nature (Diamond 1975). Understanding and anticipating which combinations of interacting species from a regional pool should be observed at a given locality is a long-standing problem in ecology (Fukami 2015; Vellend 2016). Numerous studies have revealed that the specific combination of species observed in a given place and time is the result of a complex interplay between many ecological and evolutionary processes, including species interactions (Chase and Leibold 2003), environmental conditions (Kraft et al. 2015), dispersal (Moore et al. 2008), priority effects (Fukami 2015), phylogenetic relationships (Maherali and Klironomos 2007; Song et al. 2018a), and stochasticity (Rosindell et al. 2011). Although each of these processes can have important effects at different stages of community assembly, we still lack a simple framework that would allow us to anticipate which specific combinations of interacting species are more likely to be observed in nature.

Typically, the composition of a local community results from an assembly (or disassembly) process that starts from a regional pool and gives rise to a subset of coexisting species (Odum 1969; Holyoak et al. 2005). Essentially, this subset emerges from the compatibility of each species with local environmental conditions (which we define as abiotic filters) and with each other (i.e., biotic interactions; Chase and Leibold 2003; Kraft et al. 2015). However, abiotic and biotic factors are in constant change, and their relative importance in shaping communities is also

expected to change over time (Odum 1969; Fukami 2015). Therefore, there is a great degree of uncertainty about how environmental filtering, species arrival order, population dynamics, and other factors will impact the final composition of a community (Chase 2010; Fukami 2015; Cadotte and Tucker 2017). This accumulated uncertainty from measuring multiple ecological factors indicates that to understand which combinations of species should emerge from a regional pool, we need to work under a probabilistic framework (Cazelles et al. 2016; Serván et al. 2018; Song et al. 2020).

In this line, previous work has already established the question of what the probability is that a species can invade a feasible community with a given structure (typically characterized by an interaction matrix; Case 1990). In the case of unstructured communities (e.g., random matrices built from a probability distribution centered around zero), it has been shown that this probability is about 0.5 (Rossberg and Barabás 2019). Instead, when one considers structured communities, the probability can be different from an unbiased coin toss (Case 1990). Similarly, one can treat the invasion problem as the equilibrium between extinct residents and new colonizers, leading to the problem of the expected fraction of persistent species from a regional pool. It has also been shown that the fraction of persistent species follows a distribution with central tendency, and the expected fraction can be either half the pool or different from half depending on the community structure (Goh and Jennings 1977; Serván et al. 2018). Nevertheless, while the expected fraction of persistent species from a regional pool may be possible to anticipate, knowing which of the potentially many combinations with the same fraction of species should be realized is still an open challenge in community ecology (Fukami 2015; Friedman et al. 2017; Song et al. 2018a; Maynard et al. 2020). Yet answering this question could aid in the development of strategies for maintaining biodiversity in the face of environmental change (Walther 2010; Dirzo et al. 2014).

Across many areas of biology, the structuralist view has provided a systematic and probabilistic platform of understanding the diversity (or lack of diversity) that we observe in nature (Alberch 1989; Solé and Valverde 2004; Valverde et al. 2017). In particular, in community ecology, the structural approach posits that communities with a wider range of environmental conditions compatible with their persistence should have a higher probability of being observed under changing environments (Cenci et al. 2018b). While traditional approaches to community dynamics typically analyze environmental filtering and ecological dynamics separately (Kraft et al. 2015; Cadotte and Tucker 2017), the structural approach phenomenologically integrates external perturbations (e.g., changes in abiotic factors) and internal constraints (e.g., species interactions) to estimate

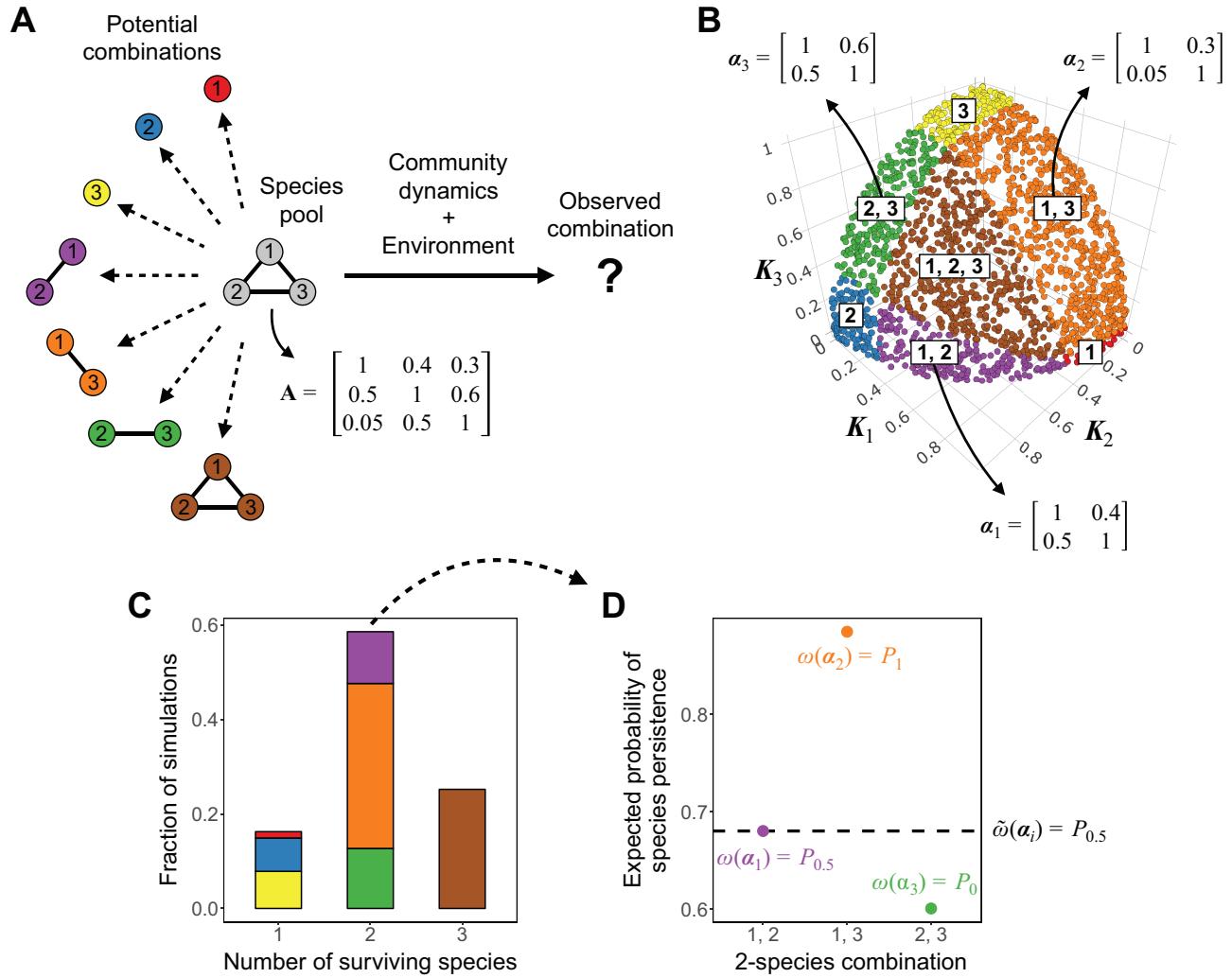
the probability of persistence of a given community (Song et al. 2018b, 2020). In particular, the structural approach uses tractable population dynamics models, such as the Lotka-Volterra model, to establish the internal rules of design of a community. Then, under these models, the structural approach investigates the range of parameter values compatible with the persistence of a community to determine the set of external perturbations under which a community may be observed (Solé and Valls 1992; Bastolla et al. 2005; Rohr et al. 2014; Saavedra et al. 2014, 2017b). In this sense, the structural approach in community ecology allows us to hypothesize which communities we are more likely to observe given the often unobserved abiotic variations at a given locality (Song et al. 2017).

Here we follow a structural approach based on feasibility and stability conditions to investigate whether combinations of species that result from Lotka-Volterra competition dynamics within a structured regional pool have a less, equal, or greater chance to persist under changing environmental conditions than potential combinations that could have been formed from the regional pool (fig. 1). More specifically, we test the hypothesis that while the fraction of persistent species from a regional pool can be the result of an unbiased die (i.e., a distribution with central tendency), the observed combination of species is the result of a biased die regardless of the fraction of persistent species (i.e., many of the combinations are almost never observed because of the constraints imposed by biotic interactions). This structural hypothesis is based on the notion that under changing environments (i.e., under abiotic variation), observed combinations of species should be among the most probable to persist, where this expected probability is given by their community structure outside of the regional pool (fig. 1). We first illustrate our hypothesis using competition communities generated with a niche framework and performing simulations of Lotka-Volterra dynamics. Then, we test our hypothesis using a 10-year observational study, recording a total of 88 annual plant-herbivore communities that comprise three different forest successional stages.

## Methods

### *Expected Probability of Species Persistence under Changing Environments*

Following recent work (Saavedra et al. 2017b; Cenci and Saavedra 2018; Song et al. 2020), we use a structural approach to study the expected probability of species persistence in a given community under changing environmental conditions. In what follows, we describe our framework, which relies on the following motivation and assumptions:



**Figure 1:** Illustration of our structural hypothesis stating that observed combinations of interacting species should have a higher expected probability of species persistence than potential combinations that could be formed from a regional species pool. *A*, Illustration of a regional pool of three species (*center*) from which seven combinations of interacting species can be formed (*left*). Links represent the direct effects between species, and the matrix  $A$  contains the interaction coefficients of the Lotka-Volterra dynamics (eq. [1]). After community dynamics proceed in a given environment, which combination of competing species is more likely to be observed (*right*)? *B*, Resulting species combinations after 2,000 simulations of the Lotka-Volterra dynamics (each with a different direction of the carrying capacity vector  $K$ , where  $\|K\| = 1$  and  $K_i > 0 \forall i$ ) starting from the regional pool shown in *A*. Species combinations are represented using the same colors as in *A*, and the three interaction submatrices ( $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ ) with two species are shown. *C*, Fraction of simulations in which one, two, or three species persisted. Each bar is partitioned into the fraction of simulations that resulted in each species combination (colors as in *A*). Note that the number of persistent species follows a distribution with central tendency. *D*, Illustration of our structural hypothesis, showing that from the expected number of persistent species, the two-species combination that is most likely to be observed under changing environments is the one with the expected probability of species persistence ( $\omega(\alpha)$ ) higher than the median across all potential two-species combinations ( $\tilde{\omega}(\alpha_i) = P_{0.5}$ ), which is shown as a dashed line. The  $r$ th percentile is represented as  $P_r$ .

(i) we consider Lotka-Volterra competition models to leverage on their mathematical tractability; (ii) we assume constant and direct pairwise interactions based on the time-invariant assumption of direct interactions under Lotka-Volterra dynamics; (iii) we focus on niche-based species interactions, which produce ecologically motivated and globally stable interaction matrices; and (iv) we use no prior

information on how environmental perturbations will impact ecological dynamics and therefore assume that all directions of environment-related parameters are equally likely.

We assume that community dynamics can be described by any competition model topologically equivalent to the classic Lotka-Volterra model (Vandermeer 1975):

$$\frac{dN_i}{dt} = N_i \frac{r_i}{K_i} \left( K_i - \sum_{j=1}^S a_{ij} N_j \right), \quad (1)$$

where  $N_i$  represents the abundance (or biomass) of species  $i$ ,  $r_i > 0$  represents the intrinsic growth rate of species  $i$ ,  $K_i > 0$  represents the carrying capacity of species  $i$ , and  $S$  denotes the number of species, whereas  $a_{ij} > 0$  and  $a_{ii} = 1$  are elements in the unitless interaction matrix  $\mathbf{A}$ , representing the per capita effect of species  $j$  on species  $i$  relative to its own self-regulation (Case 2000). The classic Lotka-Volterra model in equation (1) is a phenomenological competition model written in a  $K$ -formalism, implying that carrying capacities ( $K_i$ ) represent the summary effect between mortality rate and resource intake on the growth rate of species  $i$  and thus are linked to environmental conditions (Case 2000). We opted to write the classic Lotka-Volterra model in the  $K$ -formalism to conform with our assumption of niche-based community structures (MacArthur and Levins 1967). Note that under the niche framework, interaction matrices  $\mathbf{A}$  are by construction unitless and Volterra dissipative, which guarantees global stability of the feasible fixed point (Goh 1977; app. sec. A1; the appendix is available online). Also note that a population dynamics model is considered to be topologically equivalent to the classic Lotka-Volterra model if the phase portrait of the candidate model can be mapped into the phase portrait of equation (1) (Cenci and Saavedra 2018; for details, see app. sec. A5).

To investigate the expected probability of species persistence under changing environments in a community defined by a niche-based interaction matrix  $\mathbf{A}$  and governed by Lotka-Volterra competition dynamics, we focus on the necessary (i.e., feasible solution) and sufficient (i.e., global stability) conditions for species persistence (Strobecks 1973). Because niche-based matrices are globally stable by construction (MacArthur and Levins 1967), it is necessary only to satisfy the feasibility condition to guarantee persistence. Note that a feasible solution represents the necessary condition for persistence and permanence (Hofbauer and Sigmund 1998). The feasible equilibrium of equation (1) is given by  $\mathbf{N}^* = \mathbf{A}^{-1}\mathbf{K}$ , where  $\mathbf{N}^* = [N_1^*, \dots, N_S^*]^T$  ( $N_i^* > 0$ ) is the vector of abundances at equilibrium,  $\mathbf{K} = [K_1, \dots, K_S]^T$  is the vector of carrying capacities,  $\mathbf{A}^{-1}$  is the inverse of the interaction matrix, and the superscript  $T$  is the transpose operator. Importantly, for a given interaction matrix  $\mathbf{A}$ , feasibility in equation (1) will be satisfied as long as  $\mathbf{K}$  falls inside the feasibility domain  $D_F(\mathbf{A}) = \{\mathbf{K} | \mathbf{K} = N_1^* \mathbf{v}_1 + \dots + N_S^* \mathbf{v}_S, \text{ with } N_1^*, \dots, N_S^* > 0\}$ , where  $\mathbf{v}_i$  is the  $i$ th column vector of the interaction matrix  $\mathbf{A}$  (Song et al. 2018b). Note that if  $\mathbf{K}$  falls inside the feasibility domain, any vector  $c\mathbf{K}$  ( $c > 0$ ) has the same direction as  $\mathbf{K}$  and therefore also leads to a feasible solution.

Following Song et al. (2020), we calculate the expected probability of species persistence under changing environmental conditions within a community  $\mathbf{A}$  with  $S$  competing species as

$$\omega(\mathbf{A}) = \left( \frac{2^S \text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)} \right)^{1/S}, \quad (2)$$

where  $\text{vol}(\mathbb{B}^S)$  represents the volume of the  $S$ -dimensional unit ball,  $2^S$  normalizes the unit ball to the positive orthant, and  $\text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)$  corresponds to the volume of the intersection of the feasibility domain with the unit ball (Song et al. 2018b). Given that the intersection of the feasibility domain with the unit ball is effectively a solid angle spanned by the column vectors of  $\mathbf{A}$ ,  $\omega(\mathbf{A})$  is a measure of this solid angle normalized by the unit ball. Because we are interested only in the direction of positive  $\mathbf{K}$ -vectors, we can consider only vectors  $\mathbf{K}$  for which  $\|\mathbf{K}\| \leq 1$  and normalize the size of the feasibility domain using the positive orthant of the unit ball (i.e.,  $\mathbb{B}^S \cap \mathbb{R}_{\geq 0}^S$ ). That is, the volume of the unit ball must be scaled by  $2^{-S}$ . Note that we are fixing the magnitude of the  $\mathbf{K}$ -vectors to one under the Euclidean norm (i.e.,  $\|\mathbf{K}\| = 1$ ); however, the analysis can be done using any norm without altering the conclusions (Rohr et al. 2016).

Thus,  $\omega(\mathbf{A}) \in [0, 1]$  is a probabilistic measure and can be efficiently computed for even relatively large communities (app. sec. A1; Song et al. 2018b). Ecologically,  $\omega(\mathbf{A})$  can be interpreted as either the probability of persistence of a randomly chosen species or the expected fraction of persistent species within community  $\mathbf{A}$  under changing environmental conditions (Song et al. 2018b)—that is, the expected probability of species persistence under changing conditions. Note that Song et al. (2020) have used  $\omega(\mathbf{A})$  to study the expected switches of interaction classes within a feasible community. Here we use this methodology to study the properties of realized combinations of interacting species formed from a regional pool under changing environments. Recall that by changing the direction of the  $\mathbf{K}$ -vector, we can phenomenologically represent changes in resource availability associated with the abiotic environment.

### Theoretical Analysis

To investigate our structural hypothesis theoretically, we follow a niche framework to generate ecologically motivated interaction matrices (MacArthur and Levins 1967; Rohr et al. 2016; app. sec. A2). Specifically, we randomly generate  $S \times S$ -dimensional matrices  $\mathbf{A}$  (regional species pools), whose elements are constructed as

$$a_{ij} = \exp\left(-\frac{\|\mu_i - \mu_j\|}{4\sigma^2}\right), \quad (3)$$

where  $\mu_i$  (drawn randomly from a uniform distribution in  $[0, 1]$ ) gives the position of species  $i$  in a one-dimensional niche space and  $\sigma$  corresponds to the niche width (Svirezhev and Logofet 1983). We assume the same  $\sigma$  for all species. However, confirming previous work (Rohr et al. 2016), relaxing this constraint does not affect our conclusions (app. sec. A2). To obtain a desired mean competition strength  $\rho(\mathbf{A})$  in the regional pool, we tune  $\sigma$  so that

$$\frac{\sum_{i \neq j} a_{ij}}{(S-1)S} = \rho(\mathbf{A})$$

on average. Note that as long as  $\rho(\mathbf{A})$  is small enough such that matrices remain diagonally dominant (i.e., the sum of its off-diagonal elements must be less than its diagonal element for each row,  $\sum_{j \neq i} a_{ij} < a_{ii} \forall i$ ), interaction matrices generated via this niche framework are Volterra dissipative and guaranteed to be globally stable (Goh 1977; Lu 1998). Additionally, we explore other niche-based methodologies to generate structured interaction matrices and obtain the same qualitative results (app. sec. A2).

We start our simulations by generating 500 globally stable matrices  $\mathbf{A}$  (as described above), representing regional pools of size  $S = 20$  for three different values of mean competition strength ( $\rho(\mathbf{A}) \in \{0.025, 0.05, 0.1\}$ ; total = 1,500 matrices; fig. 2A). Then we compute  $\omega(\mathbf{A})$  for each of the regional pools  $\mathbf{A}$  to get an estimate of the expected fraction of species that would be present (i.e., persist) in the realized combinations after running Lotka-Volterra competition dynamics (eq. [1]). Recall that  $\omega(\mathbf{A})$  can be interpreted as the expected fraction of persistent species within community  $\mathbf{A}$  under changing environmental conditions. Then, for 30 matrices (10 for each value of  $\rho(\mathbf{A})$ ), we randomly sample 100 directions of  $\mathbf{K}$ -vectors uniformly on the positive orthant of the  $S$ -dimensional unit ball. Therefore, we ran 1,000 simulations per  $\rho(\mathbf{A})$  value. Initial species abundances ( $\mathbf{N} = [N_1, \dots, N_S]^T$ ) and intrinsic growth rates ( $\mathbf{r} = [r_1, \dots, r_S]^T$ ) were chosen to be 1 for every species (i.e.,  $N_i = 1$  and  $r_i = 1 \forall i$ ) and do not affect our results because of the global stability condition (Takeuchi 1996).

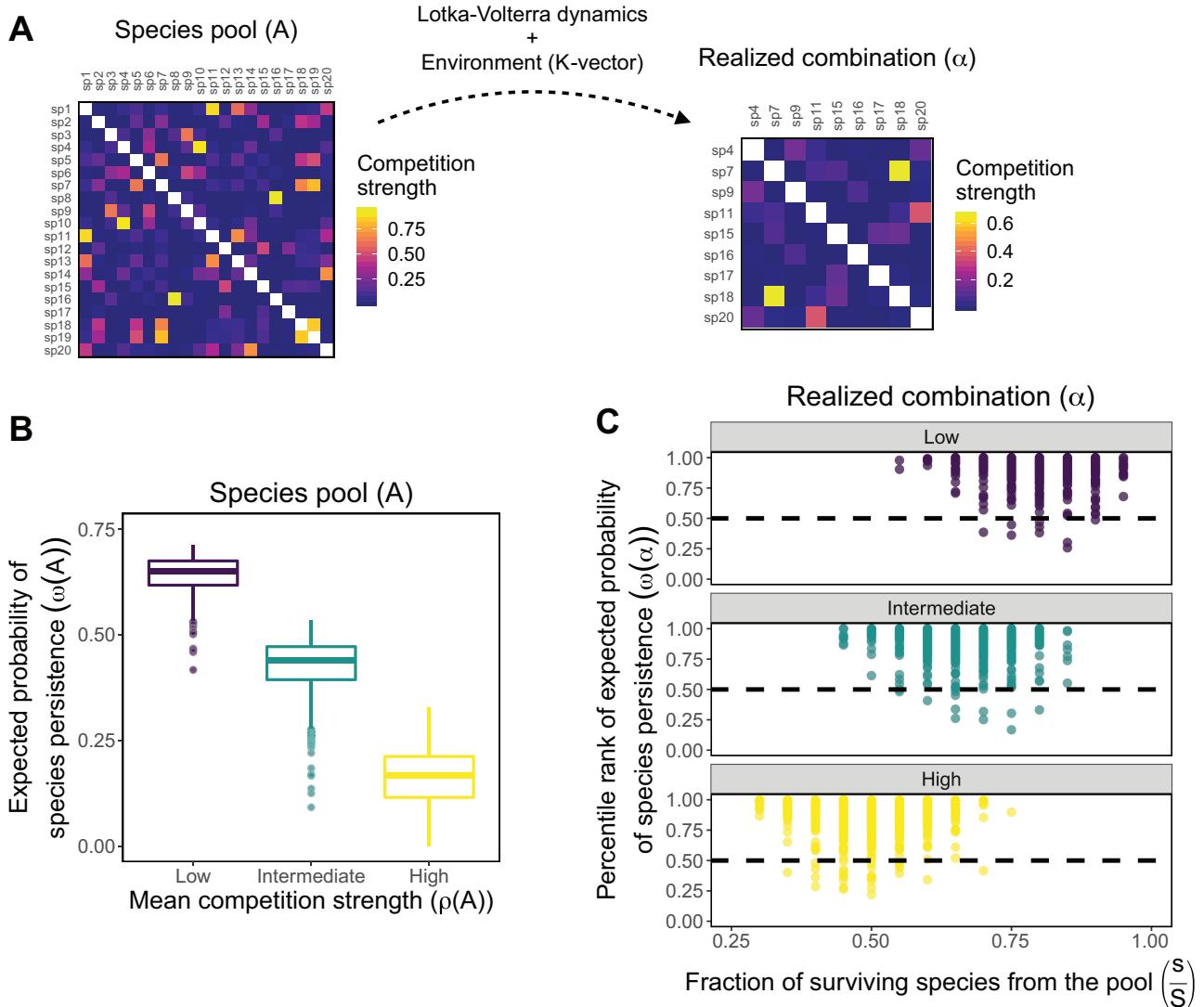
Next, for each level of competition strength  $\rho(\mathbf{A})$  and regional pool  $\mathbf{A}$ , we reach the global equilibrium point  $\mathbf{N}^*$  after numerical integration of equation (1). We record the realized combination formed by the subset of  $s \leq S$  persistent species (i.e., the subset  $\mathbf{N}^* = [N_1^*, \dots, N_s^*]^T$ , with  $N_i^* > 0 \forall i$ ). For each realized combination, we define its interaction matrix  $\alpha$  as the submatrix of  $\mathbf{A}$ , where indexes of rows and columns belong to the  $s$  persistent species. This generated 100 submatrices  $\alpha$  of different sizes  $s$  (i.e., one matrix for each sampled direction of the  $\mathbf{K}$ -vector) from each regional pool  $\mathbf{A}$ . Therefore, we generated 1,000 submatrices per  $\rho(\mathbf{A})$  value. Note that global stabil-

ity is also guaranteed for every  $\alpha$  that is a principal submatrix of a globally stable matrix  $\mathbf{A}$  (Cross 1978; Takeuchi 1996). Thus, for each realized combination defined by submatrix  $\alpha$ , we compute its expected probability of species persistence under changing conditions  $\omega(\alpha)$  following equation (2). Finally, we define the set of all potential combinations with the same number of species as  $\alpha$  as  $\Psi = \{\alpha_1, \dots, \alpha_k\}$ , where  $|\Psi| = k = \binom{S}{s}$  denotes the total number of such combinations and  $\alpha_1, \dots, \alpha_k$  denotes all the possible  $s \times s$ -dimensional submatrices that can be formed from the regional pool  $\mathbf{A}$ . For each realized combination  $\alpha$ , we sampled 500 potential combinations  $\alpha_i$  from  $\Psi$  and computed  $\omega(\alpha_i)$  for each of the potential combinations. Note that our random sampling from the species pool allows the same potential combination to be sampled twice; however, this resampling effect can be considered negligible with  $S = 20$  and 500 potential combinations. Next, using the empirical cumulative distribution function of  $\omega(\alpha_i)$ , we computed the percentage of  $\omega(\alpha_i)$  values below  $\omega(\alpha)$ . That is, we computed the percentile rank  $r$  ( $0 \leq r \leq 1$ ) of the  $r$ th percentile  $P_r = \omega(\alpha)$  of the distribution of  $\omega(\alpha_i)$ .

Following a structuralist view, we hypothesize that while the number of species forming the realized combinations has a distribution with central tendency linked to  $\omega(\mathbf{A})$ , the expected probability of species persistence within the realized combination  $\omega(\alpha)$  should be biased toward top percentile rank values regardless of the size of the realized combination. Specifically, we computed the percentage of all percentile rank values (1,000 values per  $\rho(\mathbf{A})$  value) that were above 0.5, which corresponds to the median  $P_{0.5} = \tilde{\omega}(\alpha_i)$  of  $\omega(\alpha_i)$ . We use a one-sided binomial test (unbiased binomial trial with probability of success = 0.5) to test whether the percentile rank of  $\omega(\alpha)$  is above 0.5 (i.e., the percentile rank of the median  $\tilde{\omega}(\alpha_i)$ ) for the majority of realized combinations. Note that our simulation settings were chosen simply for illustration purposes, and preliminary analyses showed that changing the number of species pools ( $\mathbf{A}$ ) built, the number of carrying capacity vectors ( $\mathbf{K}$ ) sampled, or the number of potential combinations ( $\alpha_i$ ) sampled does not affect our conclusions.

### Empirical Analysis

To test our structural hypothesis, we use a data set of 88 annual local communities formed by 22 herbivore species and 15 plant species on average from a total of 521 herbivore species feeding on 138 plant species. The study site is located in the tropical dry forest at the Chamela-Cuixmala Biosphere Reserve ( $19^{\circ}22' - 19^{\circ}39'N$ ,  $104^{\circ}56' - 105^{\circ}10'W$ ) in Jalisco, Mexico (Boege et al. 2019). Specifically, the data comprise monthly records during the rainy seasons (from



**Figure 2:** Theoretical results of our structural hypothesis using competition communities generated with a niche framework and performing simulations of Lotka-Volterra dynamics. *A, left*, illustrative example of a regional species pool (matrix A) with  $S = 20$  competing species generated following a niche framework with a high mean competition strength ( $\rho(A) = 0.1$ ). *Right*, illustrative example of a realized combination (submatrix  $\alpha$ ) formed by  $s = 9$  competing species obtained by simulating Lotka-Volterra dynamics on the regional pool with a given direction of the K-vector. Colors represent the level of competition strength  $\alpha_{ij}$ . Diagonal elements ( $\alpha_{ii}$ ) are all set to one (see “Methods”) and are not colored to improve visualization. *B*, Expected probability of species persistence under changing environments of 500 model-generated regional pools ( $\omega(A)$ ) for each value of mean competition strength (low:  $\rho(A) = 0.025$ ; intermediate:  $\rho(A) = 0.05$ ; high:  $\rho(A) = 0.1$ ). Boxplots denote the median and interquartile range. *C*, Percentile rank values ( $r$ ) of the expected probability of species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $s/S$ ) and level of mean competition strength. Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of K-vectors) per level of mean competition strength. Within each level of mean competition strength, more than 95% of the points are above the percentile rank of the median (dashed line). Furthermore, the fraction of persistent species decreases as mean competition strength increases, as anticipated by the expected probability of species persistence within the regional pool ( $\omega(A)$ ) reported in *B*. In *B* and *C*, colors indicate different levels of mean competition strength: the lighter the color, the higher the level.

July to October) from 2007 to 2017. The data collection was carried out for three stages of secondary succession: initial stage (approximately 6 years after being excluded and protected from cattle ranching use), middle stage (approximately 20 years after being excluded and protected), and late stage (more than 50 years without anthropogenic perturbations). We had three independent sampling plots of 20 m × 50 m for each successional stage (nine plots in total per year), with a minimum distance of 3 km between them. For each plot, monthly records were aggregated to build annual communities. The sites were not sampled in 2015 because of extreme bad weather conditions caused by Hurricane Patricia, and there was no access to one site in 2016 and 2017, providing a total of 88 annual plant-herbivore communities (for further details, see table A1, available online).

It is known that herbivores can negatively affect each other through a variety of mechanisms, such as induction of plant defenses or attraction of parasitoids and predators (Pallini et al. 1998; Denno et al. 2000; Redman and Scriber 2000; Ohgushi 2005). Therefore, our data set allows us to test the hypothesis that observed combinations of plant-mediated communities of competing herbivores are among the most likely to persist under the within- and across-year climatic changes that occurred across the 10-year period, including the impact of two hurricanes. Furthermore, negative effects among herbivores are expected to increase along forest succession because of an increase in host plant phylogenetic diversity and hence greater opportunities for niche differentiation, particularly for specialist herbivore species that are constrained to feed on particular plant taxa (Saavedra et al. 2017a). Thus, we leverage on these expected biotic changes across ecological succession and investigate our hypothesis separately for each forest successional stage.

For each of the 88 observed combinations (local communities) of herbivore species, we inferred their corresponding regional pool **A**. For this purpose, we first reconstructed a binary regional pool matrix **B** ( $b_{ij} = 1$  if plant species *i* is consumed by herbivore species *j*, and  $b_{ij} = 0$  otherwise) by merging all herbivore species from the same year and successional stage that fed on plant species present at a given plot. For example, to reconstruct the regional pool for plot 5 in 2011 in the middle successional stage, we merged all herbivore species from plots 4, 5, and 6 (i.e., the other plots from 2011 and the middle successional) that fed on plants from plot 5. The rationale behind this approach is that the regional pool for a given plot must contain all herbivores that could potentially feed on the plant species present at this plot. Thus, each regional pool contained only plant-herbivore interactions observed in our data set, and therefore we do not assume that herbivores could feed on any arbitrary plant species. Note that

herbivore species, plant species, and their interactions changed across years and successional stages. We also performed our analyses using different methods to reconstruct the regional pools (merging herbivores only by year or only by stage) and obtained the same qualitative results (app. sec. A3).

Next, because we have no information about the level of interaction strength among herbivore species, we follow a niche framework to infer the herbivore competition matrix **A** by computing the normalized monopartite projection of the binary bipartite matrix **B** (Cenci et al. 2018a). Specifically, we project the binary matrix on the herbivore layer as  $\mathbf{A} = \mathbf{B}^T \mathbf{B}$ , and then we normalize the columns of matrix **A** to sum one and set its diagonal elements to one afterward. By doing so, we eliminate tuning parameters that could be used to infer interaction coefficients (Saavedra et al. 2014). Thus, the matrix element  $a_{ij}$  after normalization can be interpreted as an approximation to the negative indirect effect of herbivore *j* on herbivore *i* and is proportional to the number of shared plant species between *j* and *i* while taking into account the total number of plant species consumed by *j*. Note that the effect of species *j* on *i* is not necessarily the same as the effect of species *i* on *j*. This niche-based methodology generates diagonally dominant and consequently globally stable regional pool matrices **A** (app. sec. A4).

To estimate the expected fraction of species that could be observed from these inferred pools after ecological dynamics, we compute  $\omega(\mathbf{A})$  for each of the 88 regional pool matrices **A**. We also perform prediction analyses to show that the Lotka-Volterra dynamics together with the inferred interaction coefficients can be used to perform accurate out-of-sample predictions of temporal changes in species composition and therefore support our inference methodology as a reasonable approximation of interaction matrices (app. sec. A4). In addition, we obtain the same qualitative results by performing our analyses using different niche-based methods to infer interaction coefficients (app. sec. A4).

Finally, we infer the observed interaction matrices **α** for each of the 88 observed plant-mediated communities of competing herbivores. We parameterize these matrices following the niche-based methodology mentioned above by computing the normalized monopartite projection of its binary bipartite matrix **β**, in which  $\beta_{ij} = 1$  if plant species *i* is consumed by herbivore species *j* and  $\beta_{ij} = 0$  otherwise. We then use these inferred matrices **α** to calculate the expected probability of species persistence of observed combinations under changing environments ( $\omega(\alpha)$ ). Then, to deal with the trade-off between computational constraints and statistical support, we sample a population **Ψ** of 1,000 potential combinations **α**, with the same size as **α** from the regional pool **A**. We normalize each of these potential combinations as we did for observed

combinations. Then, similarly to our theoretical analyses, we compute the percentile rank  $r$  of the  $r$ th percentile  $P_r = \omega(\alpha)$  of the distribution of  $\omega(\alpha)$ . For each successional stage, we perform a one-sided binomial test (unbiased binomial trial with probability of success = 0.5) to test whether the percentile rank of  $\omega(\alpha)$  is above 0.5 (i.e., the percentile rank of the median  $\tilde{\omega}(\alpha_i)$ ) for the majority of observed combinations.

## Results

### *Theoretical Results*

We first illustrate our structural hypothesis by performing Lotka-Volterra simulations with a toy three-species competition community. Figure 1A shows this three-species regional species pool (matrix A; *center*) and all potential combinations of interacting species (*left*) that could result from community dynamics under a given environment (i.e., a given direction of the K-vector). Figure 1B shows the realized community composition after simulating Lotka-Volterra dynamics (eq. [1]) under many different environments. Each point on the positive orthant of the unit ball represents one of 2,000 random directions of K-vectors (i.e., environmental conditions) used to simulate the dynamics. The figure shows that the feasibility domain  $D_F(A)$  (i.e., the direction of K-vectors for which the three species have positive abundances at equilibrium) is much smaller than the domain of infeasibility. In particular, different species combinations (i.e., different colors in fig. 1B) have different probabilities of being observed under random environmental conditions. Figure 1C confirms that the fraction of persistent species follows a distribution with central tendency. In this case, the expected number of persistent species is two, with three possible combinations (i.e., {1, 2}, {1, 3}, {2, 3}). The question we are addressing in this study is which of these combinations we are more likely to observe in nature. Figure 1D shows the expected probability of species persistence under changing environments ( $\omega(\alpha)$ ) for each of these two-species combinations. Our hypothesis is that the two-species combination  $\alpha_2$  containing species 1 and 3 is more likely to be observed, as it has an expected probability of species persistence higher than half of all possible combinations.

To show the generality of our hypothesis, we present our results for niche-based regional pools A with  $S = 20$  competing species under three different levels of mean competition strength (see “Methods”; fig. 2A). We found that Lotka-Volterra dynamics together with changing environmental conditions (phenomenologically represented by equally probable directions of K-vectors) lead to realized combinations of interacting species that are among the most probable to persist. In particular, figure 2B shows

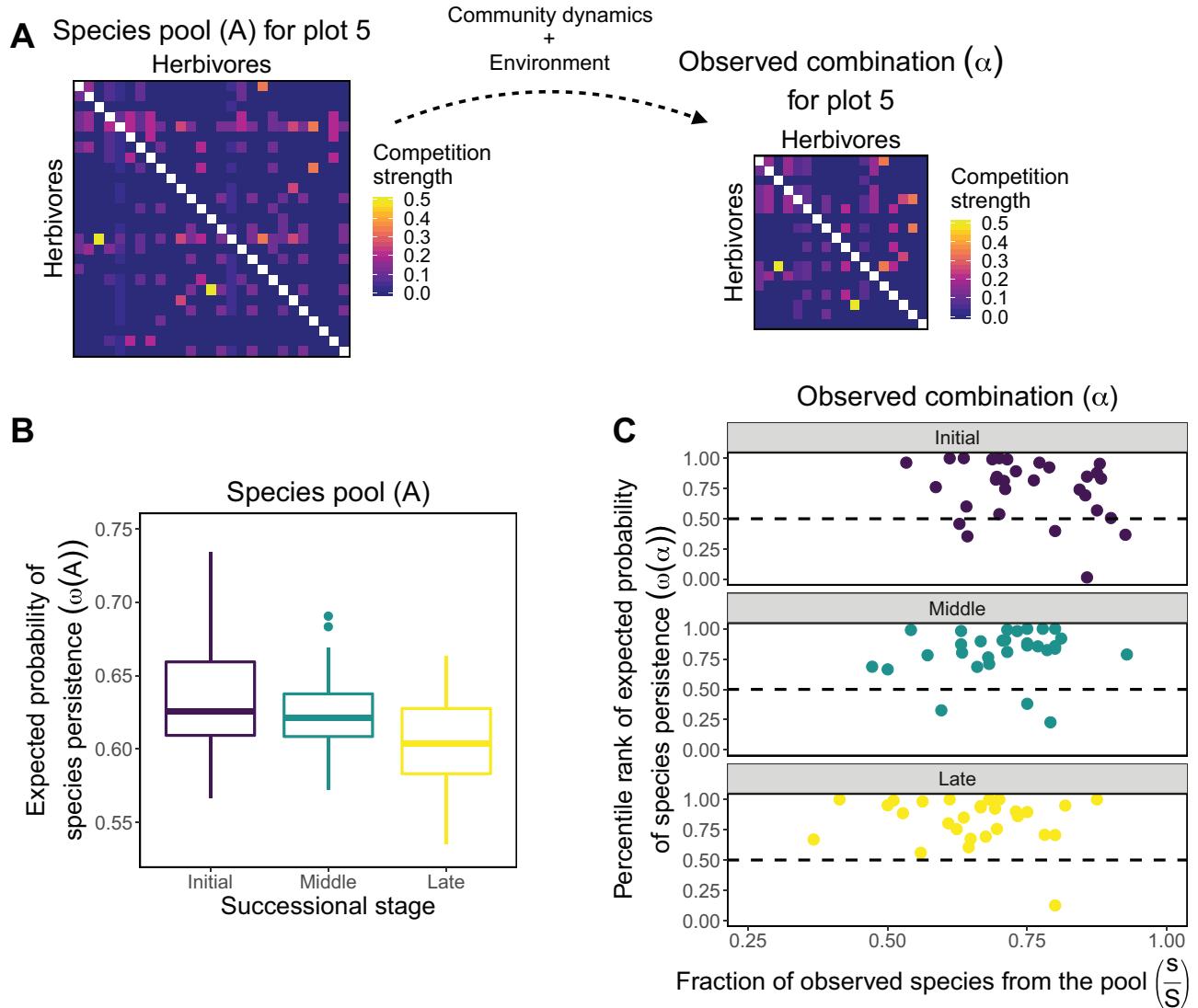
that the expected probability of species persistence within the regional pools ( $\omega(A)$ ) was higher for communities with a lower mean competition strength (low:  $0.641 \pm 0.045$ ; intermediate:  $0.422 \pm 0.069$ ; high:  $0.162 \pm 0.065$ ; mean  $\pm$  SD). As a consequence, figure 2C shows that the expected fraction of persistent species ( $s/S$ ) was also higher for communities with a lower mean competition strength (low:  $0.784 \pm 0.068$ ; intermediate:  $0.655 \pm 0.073$ ; high:  $0.488 \pm 0.075$ ). These results confirm that the expected fraction of persistent species (forming realized combinations of competing species) follows a distribution with central tendency linked to  $\omega(A)$ .

Importantly, figure 2C shows that the expected probability of species persistence of realized combinations ( $\omega(\alpha)$ ) exhibits a bias toward the top percentile rank values of the distribution of potential combinations ( $\omega(\alpha_i)$ ). Specifically, for the vast majority of realized combinations,  $\omega(\alpha)$  was above the median  $\tilde{\omega}(\alpha_i)$  of potential combinations within all three levels of mean competition strength (percentage of values above the median: low, 99.1%; intermediate, 98.8%; high, 97.3%;  $P < .0001$  for all levels of competition strength for one-sided unbiased binomial tests). We confirmed these results by building regional pools according to different frameworks (figs. A1–A4; figs. A1–A10 are available online) and by performing simulations with modified Lotka-Volterra models (figs. A9, A10).

### *Empirical Results*

We corroborated our structural hypothesis using an empirical data set of 88 plant-herbivore communities (table A1). These communities were sampled across a 10-year observational period, allowing us to test our hypothesis across three different successional stages encompassing different abiotic and biotic conditions (i.e., environmental conditions as well as herbivore and plant species change over time). As mentioned in “Methods,” we first reconstructed regional pools by merging herbivore species from the same year and successional stage. Then, we inferred herbivore interaction matrices following a niche framework with no tuning parameters by computing normalized monopartite projections for the regional pools A, observed combinations  $\alpha$ , and potential combinations  $\alpha_i$  (fig. 3A). In line with our theoretical results, we found that regardless of the number of observed interacting species, the majority of observed species combinations have an expected probability of species persistence higher than half of all the potential combinations that could have been formed with the same number of species from the reconstructed regional pools.

Figure 3B shows that the expected probability of species persistence of reconstructed regional pools ( $\omega(A)$ ) is



**Figure 3:** Empirical test of our structural hypothesis using a 10-year data set recording 88 plant-herbivore communities across three different forest successional stages. *A, left*, illustrative example of a reconstructed regional species pool (matrix  $A$ ) with  $S = 27$  competing herbivores mediated by plant resources from plot 5 (2011, middle successional stage). *Right*, observed combination (submatrix  $\alpha$ ) with the  $s = 18$  competing herbivores from plot 5. Colors represent the interaction strength  $\alpha_{ij}$ , which is proportional to the number of shared plant species. Diagonal elements ( $\alpha_{ii}$ ) are all set to one (see “Methods”) and are not colored to improve visualization. *B*, Expected probability of species persistence under changing environments of the 88 reconstructed regional pools ( $\omega(A)$ ) separated by successional stage. Boxplots denote the median and interquartile range. We removed one outlier point from the late stage to improve visualization. *C*, Percentile rank values ( $r$ ) of the expected probability of species persistence of observed combinations ( $\omega(\alpha)$ ) as a function of the fraction of observed species ( $s/S$ ) and successional stage. Note that each percentile rank is computed by comparing a given observed combination within the population formed by 1,000 sampled potential combinations with the same number of species. Each point corresponds to one of 88 different empirical communities. Within each successional stage, more than 80% of the points are above the percentile rank of the median (dashed line). Furthermore, the fraction of observed species decreases along forest succession, as suggested by the expected probability of species persistence within the regional pool ( $\omega(A)$ ) reported in *B*. In *B* and *C*, colors indicate different successional stages: the lighter the color, the later the stage.

slightly higher in initial successional stages (initial:  $0.637 \pm 0.045$ ; middle:  $0.625 \pm 0.030$ ; late:  $0.615 \pm 0.051$ ), although the difference between stages is not statistically significant (one-way ANOVA:  $F(2, 85) = 1.917$ ,  $P = .153$ ). Figure 3C shows that the expected fraction of ob-

served species (i.e., the number of observed species  $s$  divided by the number of species in the regional pool  $S$ ) is higher in initial successional stages (initial:  $0.750 \pm 0.107$ ; middle:  $0.704 \pm 0.101$ ; late:  $0.653 \pm 0.120$ ; one-way ANOVA:  $F(2, 85) = 5.722$ ,  $P = .005$ ). This result

suggests a link between  $\omega(A)$  and the expected number of species forming observed combinations for our empirical data.

Furthermore, figure 3C shows that observed combinations exhibit a bias toward top percentile rank values of the expected probability of species persistence ( $\omega(\alpha)$ ). Specifically, for the majority of observed combinations,  $\omega(\alpha)$  was above the median  $\tilde{\omega}(\alpha_i)$  of potential combinations within all three successional stages (initial: 83.3%; middle: 90.0%; late: 96.4%;  $P < .001$  for all successional stages for one-sided unbiased binomial tests). These results hold for different methods of reconstructing the regional pool (figs. A5–A7).

## Discussion

Several studies have elucidated key mechanisms that are likely to drive the assembly of ecological communities (Odum 1969; Maherli and Kliromos 2007; Fukami 2015; Kraft et al. 2015). For example, it has been shown how mechanisms such as environmental filtering (Kraft et al. 2015), phylogenetic constraints (Maherli and Kliromos 2007; Song et al. 2018a), population dynamics (Serván et al. 2018), and ecological succession (Odum 1969; Fukami 2015) can structure the composition of such communities. Nevertheless, we still lack a framework to understand and anticipate the actual combination of species that should be expected to be observed in nature. Yet answering this question requires obtaining extensive amounts of data, knowing the governing community dynamics, and understanding how environmental conditions change and affect biological populations within each community. Such complexity reveals that this answer may be addressed only under a probabilistic framework (Cazelles et al. 2016; Serván et al. 2018; Song et al. 2020).

In this line, here we have adopted a structuralist and probabilistic approach to enhance our understanding about what is likely to be seen in nature and what is not. Indeed, a fundamental question about biological systems in general is whether the forms and structures that we observe in nature represent the whole spectrum of all possible systems or only a small proportion of all potential diversity (Diamond 1975; Alberch 1989; Solé and Valverde 2004). Following this reasoning, we have combined a theoretical framework with a comprehensive data set to investigate whether the combinations of interacting species that we observe (or ought to observe) in a given place and time represent the full spectrum of possibilities or only a small fraction of them. We have postulated a testable hypothesis grounded on the feasibility and stability of population dynamics models (Saavedra et al. 2017b; Song et al. 2020). Specifically, we tested whether observed com-

binations of interacting species forming ecological communities are among the ones more likely to persist under changing environmental conditions within all the possible combinations that could have been formed with the same number of species from a regional pool of species.

Corroborating our hypothesis, we have found that the observed species combinations are not all equally likely or unlikely to be seen; rather, they are among the most likely to persist under changing environments. Specifically, we have found that half of all potential combinations of interacting species (i.e., those with expected probabilities of species persistence below the median) were rarely observed theoretically and empirically. This result suggests that the interplay between internal rules of design (i.e., community dynamics) and external forces (e.g., environmental perturbations) constrains the set of combinations of interacting species that can be realized in nature (Alberch 1989; Solé and Valverde 2004; Song et al. 2017). We have first demonstrated the validity of our hypothesis using model-generated communities and simulations of Lotka–Volterra competition dynamics. Because a model is only an approximation of the real world, we have corroborated our hypothesis using an extensive longitudinal data set comprising 88 different combinations of interacting herbivore species mediated by plant resources subject to within- and across-year variation of abiotic factors.

It is worth mentioning that our methodology is phenomenological by construction, meaning that no causative knowledge should be directly derived. Moreover, our proposed methodology is built on four important assumptions. First, community dynamics are governed by population dynamics models topologically equivalent to the classic Lotka–Volterra model (eq. [1]). Therefore, extensions of density feedback, such as Lotka–Volterra models with type II functional response, can be integrated into our framework. We show a confirmation of this in appendix section A5. Models with higher-order interactions are not integrated into our framework. However, it has been shown that these higher-order polynomial models can produce mathematical artifacts, limiting our capacity to understand ecological dynamics (AlAdwani and Saavedra 2019). Second, species interactions are defined as constant direct pairwise interactions. That is, regardless of whether two species interact within a regional pool or within a subset of such a pool, their interactions remain the same. Furthermore, we also assume that species interactions are not affected by evolutionary processes in the short timescales that we are studying, but our framework can be used to explore adaptive processes (Saavedra et al. 2017a; Cenci et al. 2018a).

Third, we assume that communities are structured following a niche framework. While this is ecologically motivated and minimizes the choice of free-tuning parameters,

this framework imposes the property of global stability for persistence that may not be necessarily met in nature. While persistence may also be attained in the absence of a stable equilibrium through stochastic processes or switching dynamics (Hofbauer and Sigmund 1998; Schreiber et al. 2019), a careful analysis of these nonequilibrium processes is outside of the scope of our work. Nevertheless, previous work (Cenci and Saavedra 2018) has shown that the structural and probabilistic approach presented here can be used for Lotka–Volterra models with demographic stochasticity, as these models are topologically equivalent to the classic Lotka–Volterra model (eq. [1]). We show a confirmation of this in appendix section A5. Fourth, we assume that all environmental conditions are equally probable. While this imposes a uniform distribution as a prior, this allows us to separate the role of interactions from the environment in shaping the realized combinations of interacting species given that any combination from the regional pool is initially possible. That being said, previous work (Cenci et al. 2018b; Song et al. 2020) has shown how this assumption can be modified if researchers have clear information about the directionality of environmental changes.

Finally, we stress that our results should be taken only as a first-order approximation of the probabilities of a randomly chosen species. Similarly, the matrices inferred from our empirical data should be taken only as an approximation of species interactions mediated by plant resources and natural enemies present at each plot. Nevertheless, the structural and probabilistic thinking that we have established here can serve as a basis for future theoretical and experimental work aiming to answer more specific questions in different ecological systems under changing environments. For example, our results have shown that while the realized combinations of interacting species appear to be the result of a biased die (i.e., many potential combinations are very unlikely to be observed), these realized combinations are not necessarily the ones maximizing the expected probability of species persistence under changing environments. Indeed, previous work has shown that community structures may be responding to environmental pressures (Song et al. 2017). Thus, the expected probability of species persistence can be higher in places where environmental variability is greater (Cenci and Saavedra 2019). This implies that it is essential to understand and anticipate how directional and nondirectional environmental stressors can change the bias of the dice again in ecological communities.

### Acknowledgments

Funding was provided by National Science Foundation grant DEB-2024349 (S.S.), Universidad Nacional Autónoma de México–Programa de Apoyo a Proyectos de Investiga-

ción e Innovación Tecnológica grant IN211916, and Secretaría de Educación Pública–Consejo Nacional de Ciencia y Tecnología grant 2015-255544 (E.d.-V. and K.B.). We thank M. AlAdwani, S. Cenci, C. Song, and P. Zu for numerous insightful discussions about this work. We also thank R. Pérez-Ishiwara and M. Baltazar for help with logistic and technological support. Finally, we thank the editors and three reviewers for suggestions that improved the manuscript. We declare no competing financial interests.

### Statement of Authorship

L.P.M. and S.S. designed the study; K.B. and E.d.-V. designed and carried out the fieldwork to obtain empirical data; A.Z.-R. carried out the molecular analysis for species identification; L.P.M. performed the analyses; S.S. supervised the study; L.P.M. and S.S. wrote a first version of the manuscript; and all authors contributed with substantial revisions.

### Data and Code Availability

The R code and data supporting the results are archived on GitHub ([https://github.com/MITEcology/AmNat\\_Medeiros\\_etal\\_2020](https://github.com/MITEcology/AmNat_Medeiros_etal_2020)) and Zenodo (<https://doi.org/10.5281/zenodo.4026319>).

### Literature Cited

- AlAdwani, M., and S. Saavedra. 2019. Is the addition of higher-order interactions in ecological models increasing the understanding of ecological dynamics? *Mathematical Biosciences* 315:108222.
- Alberch, P. 1989. The logic of monsters: evidence for internal constraint in development and evolution. *Geobios* 22:21–57.
- Bastolla, U., M. Lassig, S. Manrubia, and A. Valleriani. 2005. Biodiversity in model ecosystems. I. Coexistence conditions for competing species. *Journal of Theoretical Biology* 235:521–530.
- Boege, K., E. Villa-Galaviz, A. López-Carretero, R. Pérez-Ishiwara, A. Zaldivar-Riverón, A. Ibarra, and E. del Val. 2019. Temporal variation in the influence of forest succession on caterpillar communities: a long-term study in a tropical dry forest. *Biotropica* 51:529–537.
- Cadotte, M. W., and C. M. Tucker. 2017. Should environmental filtering be abandoned? *Trends in Ecology and Evolution* 32:429–437.
- Case, T. J. 1990. Invasion resistance arises in strongly interacting species-rich model competition communities. *Proceedings of the National Academy of Sciences of the USA* 87:9610–9614.
- . 2000. An illustrated guide to theoretical ecology. Oxford University Press, Oxford.
- Cazelles, K., N. Mouquet, D. Mouillot, and D. Gravel. 2016. On the integration of biotic interaction and environmental constraints at the biogeographical scale. *Ecography* 39:921–931.
- Cenci, S., A. Montero-Castaño, and S. Saavedra. 2018a. Estimating the effect of the reorganization of interactions on the adaptability of species to changing environments. *Journal of Theoretical Biology* 437:115–125.

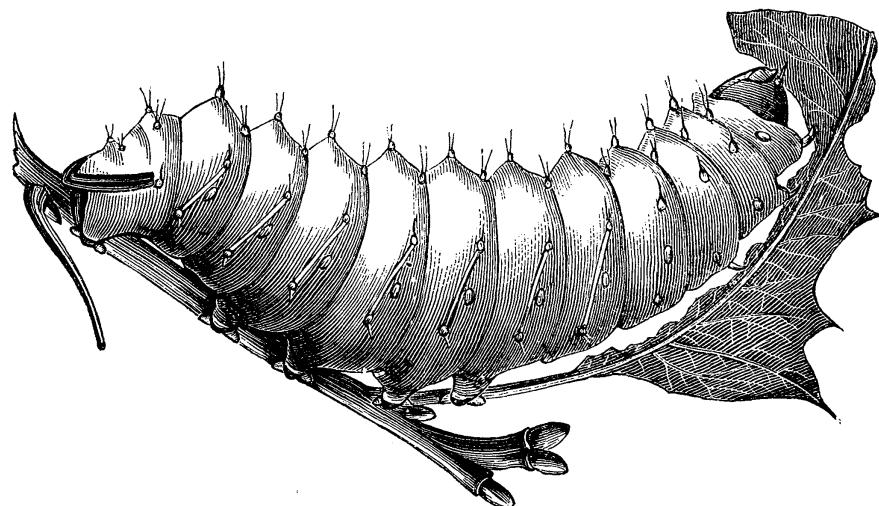
- Cenci, S., and S. Saavedra. 2018. Structural stability of nonlinear population dynamics. *Physical Review E* 97:012401.
- . 2019. Non-parametric estimation of the structural stability of non-equilibrium community dynamics. *Nature Ecology and Evolution* 3:912.
- Cenci, S., C. Song, and S. Saavedra. 2018b. Rethinking the importance of the structure of ecological networks under an environment-dependent framework. *Ecology and Evolution* 8:6852–6859.
- Chase, J. M. 2010. Stochastic community assembly causes higher biodiversity in more productive environments. *Science* 328:1388–1391.
- Chase, J. M., and M. A. Leibold. 2003. Ecological niches: linking classical and contemporary approaches. University of Chicago Press, Chicago.
- Cross, G. 1978. Three types of matrix stability. *Linear Algebra and Its Applications* 20:253–263.
- Denno, R., M. Peterson, C. Gratton, J. Cheng, G. Langellotto, A. Huberty, and D. Finke. 2000. Feeding-induced changes in plant quality mediate interspecific competition between sap-feeding herbivores. *Ecology* 81:1814–1827.
- Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 in M. L. Cody and J. M. Diamond, eds. *Ecology and evolution of communities*. Harvard University Press, Cambridge, MA.
- Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. *Science* 345:401–406.
- Friedman, J., L. M. Higgins, and J. Gore. 2017. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology and Evolution* 1:0109.
- Fukami, T. 2015. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics* 46:1–23.
- Goh, B. S. 1977. Global stability in many-species systems. *American Naturalist* 111:135–143.
- Goh, B. S., and L. S. Jennings. 1977. Feasibility and stability in randomly assembled Lotka-Volterra models. *Ecological Modelling* 3:63–71.
- Hofbauer, J., and K. Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge University Press, Cambridge.
- Holyoak, M., M. A. Leibold, and R. D. Holt. 2005. Metacommunities: spatial dynamics and ecological communities. University of Chicago Press, Chicago.
- Kraft, N. J., P. B. Adler, O. Godoy, E. C. James, S. Fuller, and J. M. Levine. 2015. Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology* 29:592–599.
- Lu, Z. 1998. Global stability for a Lotka-Volterra system with a weakly diagonally dominant matrix. *Applied Mathematics Letters* 11:81–84.
- MacArthur, R., and R. Levins. 1967. The limiting similarity, convergence, and divergence of coexisting species. *American Naturalist* 101:377–385.
- Maherali, H., and J. N. Klironomos. 2007. Influence of phylogeny on fungal community assembly and ecosystem functioning. *Science* 316:1746–1748.
- Maynard, D. S., Z. R. Miller, and S. Allesina. 2020. Predicting coexistence in experimental ecological communities. *Nature Ecology and Evolution* 4:91–100.
- Moore, R., W. Robinson, I. Lovette, and T. Robinson. 2008. Experimental evidence for extreme dispersal limitation in tropical forest birds. *Ecology Letters* 11:960–968.
- Odum, E. P. 1969. The strategy of ecosystem development. *Science* 164:262–270.
- Ohgushi, T. 2005. Indirect interaction webs: herbivore-induced effects through trait change in plants. *Annual Review of Ecology, Evolution, and Systematics* 36:81–105.
- Pallini, A., A. Janssen, and M. W. Sabelis. 1998. Predators induce interspecific herbivore competition for food in refuge space. *Ecology Letters* 1:171–177.
- Redman, A. M., and J. M. Scriber. 2000. Competition between the gypsy moth, *Lymantria dispar*, and the northern tiger swallowtail, *Papilio canadensis*: interactions mediated by host plant chemistry, pathogens, and parasitoids. *Oecologia* 125:218–228.
- Rohr, R. P., S. Saavedra, and J. Bascompte. 2014. On the structural stability of mutualistic systems. *Science* 345:1253497.
- Rohr, R. P., S. Saavedra, G. Peralta, C. M. Frost, L.-F. Bersier, J. Bascompte, and J. M. Tylianakis. 2016. Persist or produce: a community trade-off tuned by species evenness. *American Naturalist* 188:411–422.
- Rosindell, J., S. P. Hubbell, and R. S. Etienne. 2011. The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology and Evolution* 26:340–348.
- Rossberg, A. G., and G. Barabás. 2019. How carefully executed network theory informs invasion ecology. *Trends in Ecology and Evolution* 34:385–386.
- Saavedra, S., S. Cenci, E. del Val, K. Boege, and R. P. Rohr. 2017a. Reorganization of interaction networks modulates the persistence of species in late successional stages. *Journal of Animal Ecology* 86:1136–1146.
- Saavedra, S., R. P. Rohr, J. Bascompte, O. Godoy, N. J. Kraft, and J. M. Levine. 2017b. A structural approach for understanding multispecies coexistence. *Ecological Monographs* 87:470–486.
- Saavedra, S., R. P. Rohr, L. J. Gilarranz, and J. Bascompte. 2014. How structurally stable are global socioeconomic systems? *Journal of the Royal Society Interface* 11:20140693.
- Schreiber, S. J., M. Yamamichi, and S. Y. Strauss. 2019. When rarity has costs: coexistence under positive frequency-dependence and environmental stochasticity. *Ecology* 100:e02664.
- Serván, C. A., J. A. Capitán, J. Grilli, K. E. Morrison, and S. Allesina. 2018. Coexistence of many species in random ecosystems. *Nature Ecology and Evolution* 2:1237.
- Solé, R. V., and J. Valls. 1992. On structural stability and chaos in biological systems. *Journal of Theoretical Biology* 155:87–102.
- Solé, R. V., and S. Valverde. 2004. Information theory of complex networks: on evolution and architectural constraints. Pages 189–207 in E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, eds. *Complex networks*. Springer, Berlin.
- Song, C., S. V. Ahn, R. P. Rohr, and S. Saavedra. 2020. Towards a probabilistic understanding about the context-dependency of species interactions. *Trends in Ecology and Evolution* 35:384–396.
- Song, C., F. Altermatt, I. Pearse, and S. Saavedra. 2018a. Structural changes within trophic levels are constrained by within-family assembly rules at lower trophic levels. *Ecology Letters* 21:1221–1228.
- Song, C., R. P. Rohr, and S. Saavedra. 2017. Why are some plant-pollinator networks more nested than others? *Journal of Animal Ecology* 86:1417–1424.
- . 2018b. A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *Journal of Theoretical Biology* 450:30–36.
- Strobeck, C. 1973. *N* species competition. *Ecology* 54:650–654.

- Svirezhev, Y. M., and D. O. Logofet. 1983. Stability of biological communities. Mir, Moscow.
- Takeuchi, Y. 1996. Global dynamical properties of Lotka-Volterra systems. World Scientific, Singapore.
- Valverde, S., J. M. Montoya, L. Joppa, and R. Solé. 2017. The architecture of mutualistic networks as an evolutionary spandrel. *Nature Ecology and Evolution* 2:94–99.
- Vandermeer, J. H. 1975. Interspecific competition: a new approach to the classical theory. *Science* 188:253–255.
- Vellend, M. 2016. The theory of ecological communities. Monographs in Population Biology 57. Princeton University Press, Princeton, NJ.
- Walther, G. R. 2010. Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B* 365:2019–2024.
- Genz, A., and F. Bretz. 2009. Computation of multivariate normal and  $t$  probabilities. Lecture Notes in Statistics 195. Springer, Dordrecht.
- Harris, D. J. 2016. Inferring species interactions from co-occurrence data with Markov networks. *Ecology* 97:3308–3314.
- Hastings, A., and T. Powell. 1991. Chaos in a three-species food chain. *Ecology* 72:896–903.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. *PLoS Biology* 2:e312.
- Jordano, P. 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *American Naturalist* 129:657–677.
- Logofet, D. O. 1993. Matrices and graphs: stability problems in mathematical ecology. CRC, Boca Raton, FL.
- McKane, A. J., T. Biancalani, and T. Rogers. 2014. Stochastic pattern formation and spontaneous polarisation: the linear noise approximation and beyond. *Bulletin of Mathematical Biology* 76:895–921.
- Morueta-Holme, N., B. Blonder, B. Sandel, B. J. McGill, R. K. Peet, J. E. Ott, C. Violette, B. J. Enquist, P. M. Jørgensen, and J.-C. Svenning. 2016. A network approach for inferring species associations from co-occurrence data. *Ecography* 39:1139–1150.
- Ribando, J. M. 2006. Measuring solid angles beyond dimension three. *Discrete and Computational Geometry* 36:479–487.

### References Cited Only in the Online Enhancements

- Aggarwal, C. C. 2015. Data mining: the textbook. Springer, Cham.
- Allesina, S., and S. Tang. 2015. The stability-complexity relationship at age 40: a random matrix perspective. *Population Ecology* 57:63–75.
- Barner, A. K., K. Coblenz, S. Hacker, and B. Menge. 2018. Fundamental contradictions among common estimates of non-trophic species interactions. *Journal of the Royal Society Interface* 99:557–566.
- Carrara, F., A. Giometto, M. Seymour, A. Rinaldo, and F. Altermatt. 2015. Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity. *Methods in Ecology and Evolution* 6:895–906.

Associate Editor: Chuan Yan  
Editor: Jennifer A. Lau



"The cocoons of *Platysamia Cecropia* may be rendered of some commercial value, as the silk can be carded, but the chief objection as stated above, is the difficulty of raising the larva. The *Polyphemus* worm spins a strong, dense, oval cocoon, which is closed at each end, while the silk has a very strong and glossy fibre." From "The American Silk Worm" by L. Trouvelot (*The American Naturalist*, 1867, 1:30–38).

## Online Appendix

Observed ecological communities are formed by species combinations that are among the most likely to persist under changing environments

Lucas P. Medeiros<sup>1</sup>, Karina Boege<sup>2</sup>, Ek del-Val<sup>3</sup>,  
Alejandro Zaldivar-Riverón<sup>4</sup>, Serguei Saavedra<sup>1</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, MIT,  
77 Massachusetts Av., 02139 Cambridge, MA, USA

<sup>2</sup>Instituto de Ecología, Universidad Nacional Autónoma de México  
Av. Universidad 3000, Ciudad Universitaria  
04510 Ciudad de México, México

<sup>3</sup>Instituto de Investigaciones en Ecosistemas y Sustentabilidad,  
Universidad Nacional Autónoma de México  
Antigua Carretera a Pátzcuaro 8701, 58190 Morelia, Michoacán, México

<sup>4</sup>Instituto de Biología, Universidad Nacional Autónoma de México  
Av. Universidad 3000, Ciudad Universitaria  
04510 Ciudad de México, México

## Contents

<b>A1 Feasibility domain in the Lotka-Volterra model</b>	<b>3</b>
<b>A2 Frameworks for generating species pools</b>	<b>6</b>
A2.1 Niche overlap in resource space . . . . .	6
A2.2 Resource overlap in bipartite matrices . . . . .	8
A2.3 Random interaction matrices . . . . .	10
A2.4 Niche overlap with different niche widths for species . . . . .	12
<b>A3 Empirical data</b>	<b>14</b>
A3.1 Data collection . . . . .	14
A3.2 Reconstruction of regional pools . . . . .	19
<b>A4 Parameterization of empirical interaction strengths</b>	<b>22</b>
A4.1 Normalizing by columns . . . . .	22
A4.2 Using a tuning parameter . . . . .	23
A4.3 Prediction analyzes . . . . .	25
<b>A5 Modifications of the classic Lotka-Volterra model</b>	<b>28</b>

## A1 Feasibility domain in the Lotka-Volterra model

In this study, we focus on the classic Lotka-Volterra (LV) model (Case, 2000, Logofet, 1993) in order to estimate the expected probability of species persistence of combinations of interacting species and to derive our hypothesis. It is important to note, however, that our hypothesis also applies to any model topologically equivalent to the LV model (Cenci and Saavedra, 2018) (Section A5). The LV model written in a  $K$ -formalism is given by (Vandermeer, 1975):

$$\frac{dN_i}{dt} = N_i \frac{r_i}{K_i} \left( K_i - \sum_{j=1}^S a_{ij} N_j \right), \quad (\text{A1})$$

where  $N_i$  is the abundance (or biomass) of species  $i$ ,  $r_i$  is the intrinsic growth rate of species  $i$ ,  $K_i$  is the carrying capacity of species  $i$ , and  $a_{ij}$  is an element in the interaction matrix  $\mathbf{A} \in \mathbb{R}^{S \times S}$  representing the per capita effect of species  $j$  on species  $i$  (Case, 2000). Although most of our results are for the competition LV system (i.e.,  $r_i > 0$ ,  $K_i > 0$ , and  $a_{ij} > 0$ ), note that the LV model can also be expressed in the  $r$ -formalism to denote other types of ecological systems, such as food webs (Case, 2000). We use a  $K$ -formalism here in order to be consistent with the niche framework we used to generate theoretical matrices (MacArthur and Levins, 1967, Rohr et al., 2016).

A non-trivial equilibrium for the LV model is given by:

$$K_i = \sum_{j=1}^S a_{ij} N_j^*, \quad i = 1, \dots, S \quad (\text{A2})$$

In matrix form, this equilibrium can be written as  $\mathbf{K} = \mathbf{A}\mathbf{N}^* \iff \mathbf{N}^* = \mathbf{A}^{-1}\mathbf{K}$ , where  $\mathbf{N}^* = [N_1^*, \dots, N_S^*]^\top$  is a vector with equilibrium abundances,  $\mathbf{K} = [K_1, \dots, K_S]^\top$  is a vector of carrying capacities, and the superindex  $^\top$  represents the transpose operator. A feasible equilibrium of the LV model satisfies these equations as well as  $N_i > 0 \forall i$ . If a feasible equilibrium exists, then this equilibrium is guaranteed to be globally stable under certain specific conditions. Global stability implies that any trajectory of this dynamical system converges to the equilibrium point regardless of the initial condition (Goh, 1977). If  $\mathbf{A} + \mathbf{A}^\top$  is positive definite, then  $\mathbf{A}$  is Volterra-dissipative (note that the opposite is not necessarily true) and, therefore, globally stable (Goh, 1977). In this study we assume feasibility of equilibrium points as well as Volterra-dissipative matrices in order to guarantee persistence. Note that all matrices generated under a niche framework that are diagonally dominant are also Volterra-dissipative and, therefore, globally stable (Goh, 1977),

MacArthur and Levins, 1967). In addition, note that global stability is also guaranteed for every  $\alpha$  that is a principal submatrix of a globally stable matrix  $\mathbf{A}$  (Cross, 1978, Takeuchi, 1996).

In order to study the expected probability of species persistence of a given community  $\mathbf{A}$  under changing environments, we have to estimate the range of different environments that gives us feasible equilibrium points. With the LV model presented here this can be done by assuming that interaction terms ( $a_{ij}$ ) are constant and the environment affects carrying capacities ( $K_i$ ). Under the  $K$ -formalism, we can use the fact that  $\mathbf{K} = \mathbf{A}\mathbf{N}^*$  to represent the feasibility domain of the community as the set of all  $\mathbf{K}$  vectors that give a feasible equilibrium:

$$D_F(\mathbf{A}) = \{\mathbf{K} = N_1^*\mathbf{v}_1 + \dots + N_S^*\mathbf{v}_S, \text{with } N_i^* > 0 \forall i\} \quad (\text{A3})$$

where the vector  $\mathbf{v}_j$  corresponds to the  $j$ -th column of matrix  $\mathbf{A}$ . Because the feasibility domain is described by the positive linear combinations of the columns of matrix  $\mathbf{A}$ , it corresponds to an algebraic cone in  $\mathbb{R}^S$  (Saavedra et al., 2017b, Song et al., 2018). This geometric view of the feasibility domain leads to a probabilistic interpretation. The size of the feasibility domain is equivalent to the probability of randomly sampling a direction of the  $\mathbf{K}$ -vector that falls inside the algebraic cone defined by  $D_F(\mathbf{A})$ .

To compute the size of the feasibility domain, we first define the unit ball in  $\mathbb{R}^S$  as  $\mathbb{B}^S = \{\mathbf{u} \in \mathbb{R}^S, \|\mathbf{u}\| \leq 1\}$ . Then, for a competition LV system, the size of the feasibility domain (i.e., probability of persistence under changing environments) can be computed as the volume of the algebraic cone that is inside the unit ball divided by the volume of the unit ball inside the positive quadrant  $\mathbb{R}_{>0}^S$ :

$$\Omega(\mathbf{A}) = \frac{2^S \text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)}, \quad (\text{A4})$$

where  $2^S$  normalizes the feasibility domain to positive parameter values (i.e.,  $K_i > 0 \forall i$ ). Note, however, that this probability of persistence can be computed for ecological systems in which  $K_i$  and  $a_{ij}$  can be negative by dropping the  $2^S$  term in the equation above (see Section A2.3). For competitive interactions,  $\Omega(\mathbf{A}) = P(\mathbf{K} \in D_F(\mathbf{A})) \in [0, 1]$ , where the upper bound represents a scenario with no interspecific interactions. Assuming that any direction of the  $\mathbf{K}$ -vector is equally likely, the size of the feasibility domain can be numerically estimated by solving the following integral (Ribando, 2006):

$$\Omega(\mathbf{A}) = 2^S \frac{\text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)} \quad (\text{A5})$$

$$= 2^S \frac{\det(\mathbf{A})}{\sqrt{\pi^S}} \int_{\mathbb{R}_{\geq 0}^S} e^{-\frac{1}{2}\mathbf{N}^{*\top} \mathbf{A}^\top \mathbf{A} \mathbf{N}^*} d\mathbf{N}^* \quad (\text{A6})$$

$$= 2^S \frac{1}{\sqrt{(2\pi)^S \det(\Sigma)}} \int_{\mathbb{R}_{\geq 0}^S} e^{-\frac{1}{2}\mathbf{N}^{*\top} \Sigma^{-1} \mathbf{N}^*} d\mathbf{N}^* \quad (\text{A7})$$

where we set  $\mathbf{A}^\top \mathbf{A} = \Sigma^{-1}$  in order to obtain the cumulative distribution function of a multivariate normal distribution with mean  $\mu = 0$  and covariance matrix  $\Sigma$ . Equation (A5) can be computed efficiently for even relatively large communities by solving numerically the integration using a quasi-Monte Carlo method (Genz and Bretz, 2009, Song et al., 2018). Note that the integration is done over the entire positive orthant of the  $S$ -dimensional state space.

Following Song et al. (2020), by assuming that the sampling of the directions of carrying capacities is independent and identically distributed for every species, we can also compute the average probability that a randomly chosen species  $i$  is feasible (i.e.,  $N_i^* > 0$ ) as:

$$\omega(\mathbf{A}) = \Omega(\mathbf{A})^{\frac{1}{S}} \quad (\text{A8})$$

$$= \left( 2^S \frac{\text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)} \right)^{\frac{1}{S}} \quad (\text{A9})$$

$$= 2 \left( \frac{\text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)} \right)^{\frac{1}{S}} \quad (\text{A10})$$

This species-level measure is bounded as  $\omega(\mathbf{A}) \in [0, 1]$  and can be interpreted either as the probability of persistence of a randomly chosen species or as the expected fraction of persistent species within community  $\mathbf{A}$  under changing environmental environments (Song et al., 2018)—i.e., the expected probability of species persistence under changing environments.

## A2 Frameworks for generating species pools

In this section we describe the three niche-based frameworks we used to generate the regional pools for our theoretical analyzes.

### A2.1 Niche overlap in resource space

In the main text, we report the results for species pools generated according to a niche framework (MacArthur and Levins, 1967, Rohr et al., 2016). To generate a  $S \times S$  competition matrix  $\mathbf{A}$  under this framework, we first sampled  $\mu_i$  randomly from a uniform distribution in  $[0, 1]$ , which gives the position of species  $i$  in a 1-dimensional niche space. Then, we computed the Euclidean distances  $\|\mu_i - \mu_j\|$  and interaction coefficients  $a_{ij} = e^{-\frac{\|\mu_i - \mu_j\|}{4\sigma^2}}$  for each pair of species. The parameter  $\sigma$ , which measures the niche breadth of each species, was tuned in order to set a desired mean competition strength in the matrix as  $\frac{\sum_{i \neq j} a_{ij}}{(S-1)S} = \rho(\mathbf{A})$ . For illustration purposes, we used  $S = 20$  and three values of  $\rho(\mathbf{A})$ : 0.025, 0.05, and 0.1. Importantly, matrices generated via this niche framework are guaranteed to be globally stable (MacArthur and Levins, 1967, Rohr et al., 2016).

In the main text, we report the simulation results for species pools  $\mathbf{A}$  generated as described above. Here, we also report the results for the expected probability of species persistence of the regional pool ( $\omega(\mathbf{A})$ ) and the percentile rank of realized combinations ( $\omega(\boldsymbol{\alpha})$ ) for regional pool matrices  $\mathbf{A}$  with normalized columns (Figure A1). These matrices were generated as described above and then we set  $a_{ij} = \frac{a_{ij}}{\sum_{i=1}^S a_{ij}}$  and  $a_{ii} = 1$  afterwards. We performed this analysis in order to check if the normalization procedure, which we also applied to the empirical data, could change our results. Figure A1 below shows that our results for regional pools normalized by column sum are qualitatively the same as the results in Figure 2 in the main text.

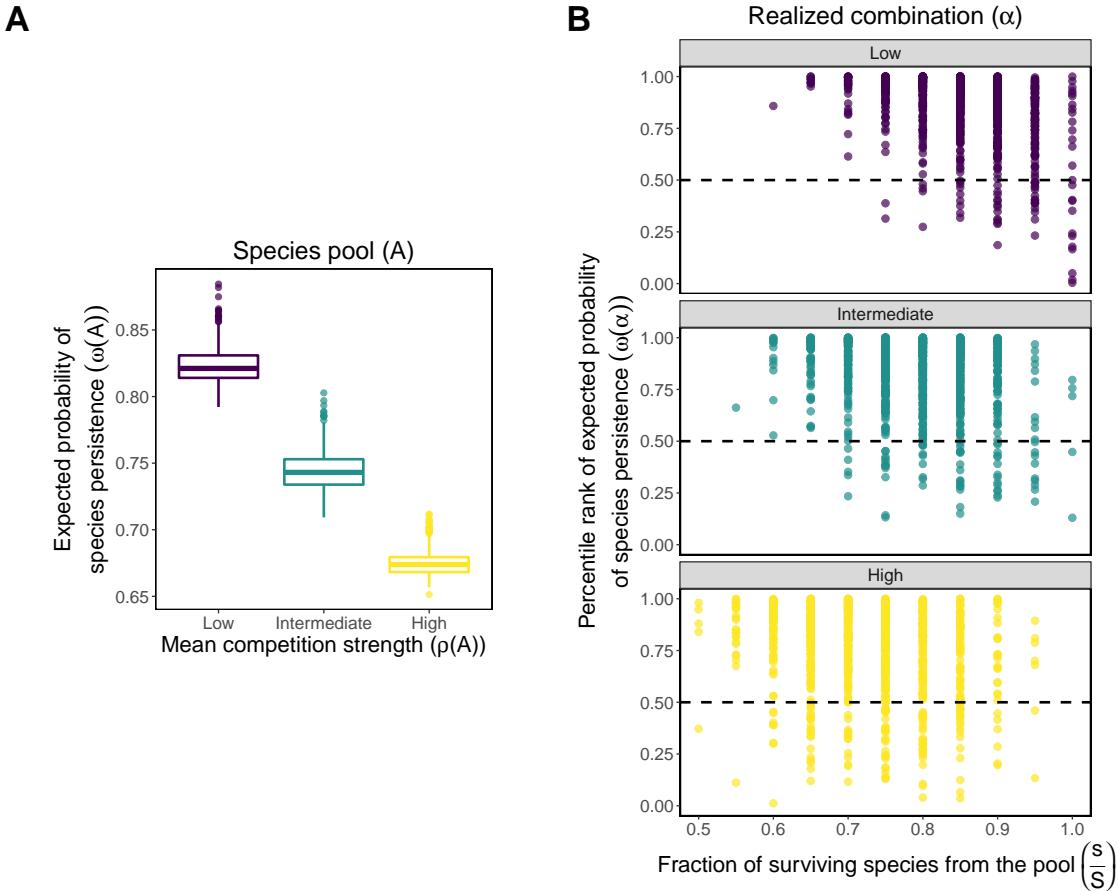


Figure A1: Theoretical results obtained by building the regional pool with the niche framework and normalizing competition matrices by column sum. **(A)** The expected probability of species persistence under changing environments of 500 model-generated regional pools ( $\omega(\mathbf{A})$ ) with  $S = 20$  for each value of mean competition strength (low:  $\rho(\mathbf{A}) = 0.025$ , intermediate:  $\rho(\mathbf{A}) = 0.05$ , and high:  $\rho(\mathbf{A}) = 0.1$ ). Boxplots denote the median and interquartile range. **(B)** Percentile rank values ( $r$ ) of the expected probability of species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and level of mean competition strength. Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of  $\mathbf{K}$ -vectors) per level of mean competition strength. The percentage of points above the median (dashed line) for each value of mean competition strength is: low: 94.3%, intermediate: 93.2%, and high: 86.4% ( $p < 0.0001$  for all levels of mean competition strength for one-sided unbiased binomial tests). In **(A)** and **(B)** colors indicate different levels of mean competition strength: the lighter the higher.

## A2.2 Resource overlap in bipartite matrices

In addition to the niche framework described above, we used an alternative niche-based framework built upon bipartite matrices which is more similar to our empirical analyzes. To do so, we first generated an  $S_1 \times S_2$  Erdős–Rényi bipartite matrix  $\mathbf{B}$  containing two groups of interacting species. In our empirical analyzes,  $S_1$  could correspond to the number of herbivore species and  $S_2$  could correspond to the number of plant species. We set each interaction term  $b_{ij}$  to 1 with probability  $p_c$  and to 0 with probability  $1 - p_c$ , in which  $p_c$  is the expected connectance in the matrix (Jordano, 1987). We only used three values of  $p_c$  for illustration purposes: 0.1, 0.2, and 0.3. Importantly, these values are close to the connectance values of our empirical plant-herbivore communities, which are computed as  $C = \frac{\sum_{i=1}^{S_1} \sum_{j=1}^{S_2} b_{ij}}{S_1 S_2}$  and range from 0.055 to 0.341. Then, we computed the monopartite projection of matrix  $\mathbf{B}$  as  $\mathbf{A} = \mathbf{B}^\top \mathbf{B}$  and normalized the columns of matrix  $\mathbf{A}$  to sum 1 (i.e.,  $a_{ij} = \frac{a_{ij}}{\sum_{i=1}^{S_1} a_{ij}}$ ) and set its diagonal elements to 1 (i.e.,  $a_{ii} = 1$ ) afterwards.

After generating the species pools from bipartite matrices, we performed the simulations of LV dynamics described in the main text and computed the expected probability of species persistence of the regional pool ( $\omega(\mathbf{A})$ ) and the percentile rank of the realized combinations ( $\omega(\boldsymbol{\alpha})$ ). Figure A2 below shows that our results for regional pools built from bipartite matrices are qualitatively the same as the results in Figure 2 in the main text.

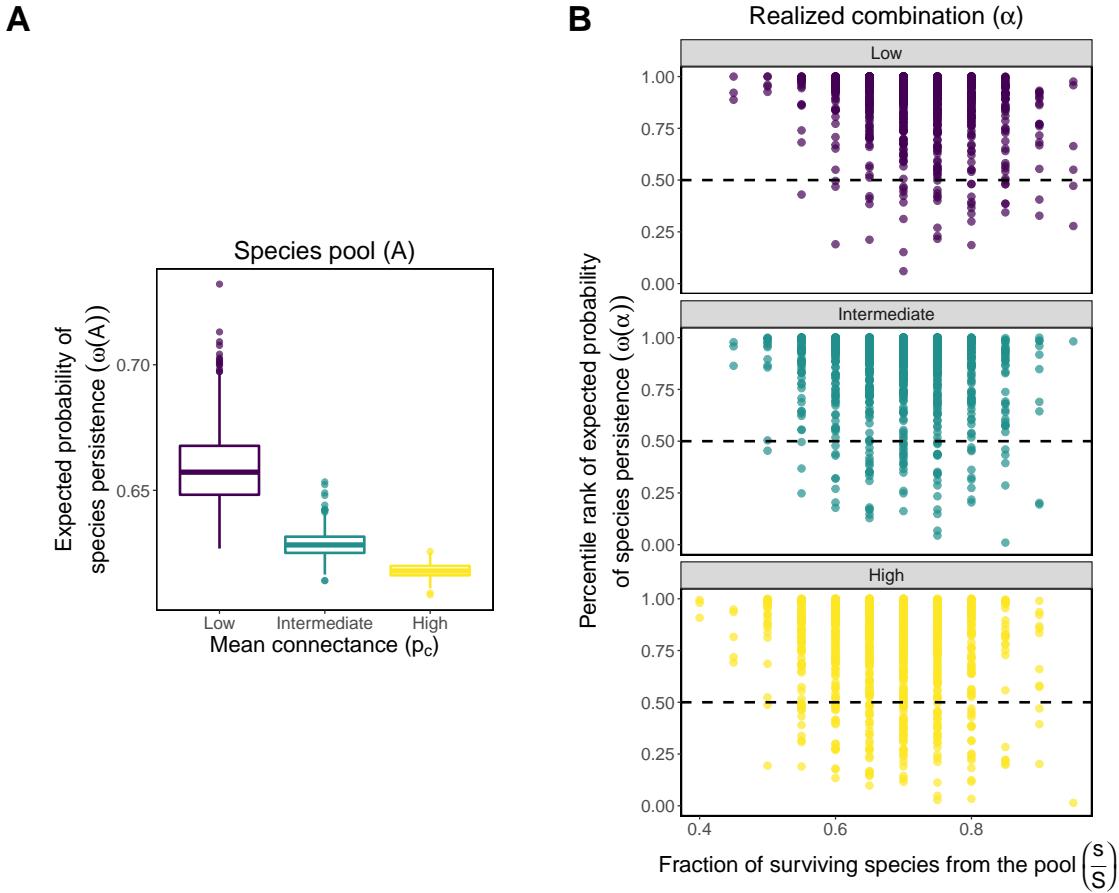


Figure A2: Theoretical results obtained by building the regional pool with bipartite networks summarizing resource use. **(A)** The expected probability of species persistence of 500 model-generated regional pools ( $\omega(A)$ ) with  $S = 20$  for each value of mean connectance (low:  $p_c = 0.1$ , intermediate:  $p_c = 0.2$ , and high:  $p_c = 0.3$ ). Boxplots denote the median and interquartile range. **(B)** Percentile rank values ( $r$ ) of the expected probability of species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and level of mean connectance. Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of  $\mathbf{K}$ -vectors) per level of mean connectance. The percentage of points above the median (dashed line) for each value of mean connectance is: low: 95.6%, intermediate: 91.7%, and high: 85.0% ( $p < 0.0001$  for all levels of mean connectance for one-sided unbiased binomial tests). In **(A)** and **(B)** colors indicate different levels of mean connectance: the lighter the higher.

### A2.3 Random interaction matrices

Our third framework for generating regional species pools assumes random interactions that can take positive or negative values (Allesina and Tang, 2015, Serván et al., 2018). We built random matrices  $\mathbf{A}$  by first creating a matrix  $\mathbf{B}$  with each element  $b_{ij}$  sampled independently from a standard normal distribution  $\mathcal{N}(\mu = 0, \sigma^2 = 1)$ . To guarantee diagonal stability and, therefore, global stability, we summed a value slightly larger than the largest eigenvalue of  $\mathbf{B} + \mathbf{B}^\top$  to each diagonal element  $b_{ii}$  of  $\mathbf{B}$  (Serván et al., 2018). That is, our final matrix was given by  $\mathbf{A} = \mathbf{B} + \mathbf{D}$ , where  $d_{ii} = \max(\Re(\lambda(\mathbf{B} + \mathbf{B}^\top))) + 0.0001$  and  $d_{ij} = 0$ . Adding  $d_{ii}$  to each diagonal element  $b_{ii}$  is sufficient to make the matrix  $\mathbf{A} + \mathbf{A}^\top$  positive definite. Note that  $d_{ij}$  corresponds to the elements of matrix  $\mathbf{D}$ . Following the previous frameworks to generate species pools, we created 10 matrices  $\mathbf{A}$  with  $S = 20$  species each and randomly sampled 100 directions of  $\mathbf{K}$ -vectors on the  $S$ -dimensional unit ball (i.e.,  $\|\mathbf{K}\| = 1$ ) for each matrix. Then, we integrated Equation (A1) for each species pool and  $\mathbf{K}$ -vector (total: 1,000 simulations).

Importantly, because interactions  $a_{ij}$  can take positive and negative values, we also assumed that each  $K_i$  can take positive and negative values. Because we are now considering the full parameter space of carrying capacities and not just the positive orthant, we drop the  $2^S$  term from the calculation of the expected probability of species persistence:

$$\omega(\mathbf{A}) = \left( \frac{\text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)} \right)^{\frac{1}{S}} \quad (\text{A11})$$

Despite this modification, the computational method and interpretation of  $\omega(\mathbf{A})$  remains the same (Song et al., 2018). Figure A3 below shows the percentile rank in the expected probability of species persistence under changing environments for realized combinations ( $\omega(\boldsymbol{\alpha})$ ) computed after performing LV simulations with the random regional pools  $\mathbf{A}$  and  $\mathbf{K}$ -vectors. This result is qualitatively the same as the result in Figure 2 in the main text.

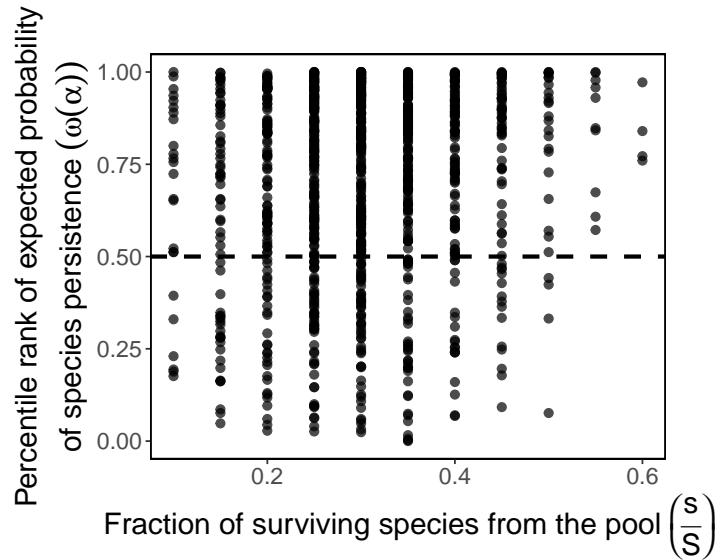


Figure A3: Theoretical results obtained by building the regional species pool with random interactions (i.e., both positive and negative interactions). Each point shows the percentile rank in the expected probability species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ). Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of  $\mathbf{K}$ -vectors). The percentage of points above the median (dashed line) is: 76.5% ( $p < 0.0001$  for one-sided unbiased binomial test).

## A2.4 Niche overlap with different niche widths for species

Here, we tested whether our main results presented in Figure 2 (main text) hold when we relax the condition that all species have the same niche width  $\sigma$  (Section A2.1). Specifically, we replicated our analysis by building regional species pools **A** with  $S = 20$  species by first sampling  $\mu_i$  randomly from a uniform distribution in  $[0, 1]$  and  $\sigma_i$  randomly from a uniform distribution in  $[0.005, 0.009]$ , where  $\sigma_i$  is the niche width of species  $i$  (MacArthur and Levins, 1967, Rohr et al., 2016). We used this uniform distribution for  $\sigma_i$  in order to obtain globally stable matrices, which we checked by confirming that the matrix  $\mathbf{A} + \mathbf{A}^\top$  was positive definite. Then, we computed the Euclidean distances  $\|\mu_i - \mu_j\|$  and interaction coefficients  $a_{ij} = e^{-\frac{\|\mu_i - \mu_j\|}{2(\sigma_i^2 + \sigma_j^2)}}$  for each pair of species (Rohr et al., 2016). The mean competition strength ( $\rho(\mathbf{A})$ ) for matrices built according to this procedure was approximately 0.05.

Figure A4 below shows the percentile rank in the expected probability of species persistence under changing environments for realized combinations ( $\omega(\boldsymbol{\alpha})$ ) computed after performing LV simulations with the regional pools **A** with variable niche widths and **K**-vectors. This result is qualitatively the same as the result in Figure 2 in the main text.

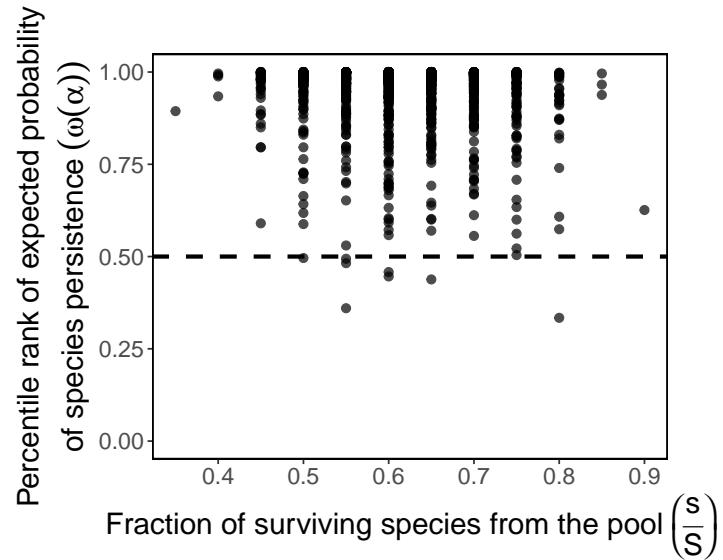


Figure A4: Theoretical results obtained by building the regional species pool with our niche framework considering that species have different niche widths  $\sigma_i$ . Each point shows the percentile rank in the expected probability species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ). Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of  $\mathbf{K}$ -vectors). The percentage of points above the median (dashed line) is: 99.2% ( $p < 0.0001$  for one-sided unbiased binomial test).

## A3 Empirical data

### A3.1 Data collection

To test our structural hypothesis, we used a 10-year data set of 88 annual plant-herbivore communities that comprise three different successional stages of a tropical dry forest. Here, we describe in detail the data collection procedures.

The study site is located in the Chamela-Cuixmala Biosphere Reserve ( $19^{\circ} 22' - 19^{\circ} 39'N$ ,  $104^{\circ} 56' - 105^{\circ} 10'W$ ) in Jalisco, Mexico (Boege et al., 2019, Saavedra et al., 2017a). Specifically, the data comprise monthly records during the rainy seasons (from July to October) from 2007 to 2017, with a total of 471 Lepidopteran caterpillar species feeding on 140 plants species. Annual average precipitation during this period was quite variable ranging from 523 to 1172 mm. The data collection was carried out for three stages of secondary succession: initial stage (approximately 6 years after being excluded and protected from cattle ranching use), middle stage (approximately 20 years after being excluded and protected), and late stage (more than 50 years without anthropogenic perturbations). We had three independent sampling plots of  $20 \times 50m$  for each successional stage ( $n = 9$  plots), with a minimum distance of 3 km between them. The sites were not sampled in 2015 due to extreme bad weather conditions caused by hurricane Patricia and there was no access to one site (plot 8) in 2016 and 2017.

Within each plot, four transects of  $2 \times 20m$  were established every 10m, in which we marked all woody plants  $\geq 1cm$  in diameter and  $\geq 50cm$  trees, excluding lianas. We monthly sampled plants looking for Lepidopteran larvae in leaves and stems 4-5 times each year during the rainy season. We reared larvae in the laboratory to confirm their trophic interactions with host plants, which were taxonomically identified. Lepidopterans were also identified using traditional taxonomy complemented with molecular identification of operational taxonomic units, following standard techniques (Hebert et al., 2004), especially for those cases in which adults failed to emerge. We edited sequences with Sequencer version 4.0.5 (Gene Codes) and aligned manually based on their translated amino acids. All caterpillars and their host plants registered within a year in each plot were considered as a single community (i.e., we merged monthly samples for each year), hence we had a total of 88 plant-herbivore communities (Table A1). Specific details on the sampling design, molecular identification, and species identities can be found in Boege et al. 2019.

Table A1: Empirical communities (i.e., observed combinations) of competing herbivore species mediated by plant resources used in this study. Stage: successional stage of the community. Plot: identification number of the sampling plot. Year: year that the data was collected.  $s$ : number of herbivore species in the observed combination.  $S$ : number of herbivore species in the reconstructed species pool.  $\omega(\mathbf{A})$ : expected probability of species persistence within regional pool.  $\omega(\alpha)$ : expected probability of species persistence within the observed combination.  $r$ : percentile rank of  $\omega(\alpha)$  within the population of potential combinations with the same number of species sampled from the regional pool.

Community	Stage	Plot	Year	$s$	$S$	$\omega(\mathbf{A})$	$\omega(\alpha)$	$r$
1	Initial	1	2007	16	21	0.679	0.725	0.817
2	Initial	1	2008	51	72	0.598	0.629	0.812
3	Initial	1	2009	17	29	0.619	0.661	0.761
4	Initial	1	2010	25	39	0.624	0.638	0.601
5	Initial	1	2011	22	35	0.624	0.656	0.458
6	Initial	1	2012	10	14	0.626	0.709	0.990
7	Initial	1	2013	9	14	0.597	0.616	0.355
8	Initial	1	2014	28	32	0.598	0.603	0.569
9	Initial	1	2016	15	17	0.689	0.707	0.832
10	Initial	1	2017	27	37	0.579	0.611	0.892
11	Initial	2	2007	25	27	0.637	0.642	0.368
12	Initial	2	2008	44	57	0.630	0.658	0.963
13	Initial	2	2009	27	38	0.609	0.641	0.745
14	Initial	2	2010	34	49	0.641	0.666	0.823
15	Initial	2	2011	16	23	0.621	0.664	0.846
16	Initial	2	2012	18	21	0.665	0.694	0.848
17	Initial	2	2013	18	20	0.613	0.617	0.506
18	Initial	2	2014	24	28	0.610	0.608	0.016
19	Initial	2	2016	14	16	0.595	0.602	0.878
20	Initial	2	2017	27	32	0.602	0.616	0.741
21	Initial	3	2007	22	25	0.687	0.726	0.954
22	Initial	3	2008	35	50	0.609	0.640	0.538
23	Initial	3	2009	22	36	0.626	0.729	0.999
24	Initial	3	2010	35	50	0.638	0.678	1.000

Community	Stage	Plot	Year	<i>s</i>	<i>S</i>	$\omega(\mathbf{A})$	$\omega(\boldsymbol{\alpha})$	<i>r</i>
25	Initial	3	2011	22	32	0.627	0.707	0.991
26	Initial	3	2012	8	15	0.715	0.879	0.963
27	Initial	3	2013	15	19	0.734	0.808	0.924
28	Initial	3	2014	14	22	0.720	0.846	1.000
29	Initial	3	2016	4	5	0.734	0.764	0.399
30	Initial	3	2017	41	48	0.566	0.573	0.693
31	Middle	4	2007	30	40	0.608	0.658	1.000
32	Middle	4	2008	41	56	0.637	0.682	0.981
33	Middle	4	2009	35	53	0.572	0.624	0.686
34	Middle	4	2010	27	38	0.631	0.680	0.907
35	Middle	4	2011	30	42	0.609	0.686	0.992
36	Middle	4	2012	11	14	0.683	0.761	0.824
37	Middle	4	2013	12	15	0.665	0.753	1.000
38	Middle	4	2014	20	35	0.608	0.690	0.782
39	Middle	4	2016	13	14	0.669	0.681	0.788
40	Middle	4	2017	31	52	0.575	0.594	0.327
41	Middle	5	2007	13	24	0.669	0.792	0.990
42	Middle	5	2008	24	38	0.650	0.704	0.872
43	Middle	5	2009	30	44	0.587	0.630	0.710
44	Middle	5	2010	24	38	0.621	0.667	0.982
45	Middle	5	2011	18	27	0.621	0.682	0.897
46	Middle	5	2012	12	17	0.638	0.725	0.903
47	Middle	5	2013	12	16	0.691	0.746	0.862
48	Middle	5	2014	15	20	0.583	0.601	0.381
49	Middle	5	2016	5	10	0.617	0.663	0.665
50	Middle	5	2017	17	36	0.600	0.683	0.687
51	Middle	6	2007	17	25	0.625	0.656	0.764
52	Middle	6	2008	30	37	0.634	0.676	0.920
53	Middle	6	2009	19	30	0.622	0.669	0.803
54	Middle	6	2010	15	20	0.625	0.657	0.880
55	Middle	6	2011	19	24	0.614	0.626	0.228

Community	Stage	Plot	Year	<i>s</i>	<i>S</i>	$\omega(\mathbf{A})$	$\omega(\boldsymbol{\alpha})$	<i>r</i>
56	Middle	6	2012	10	14	0.624	0.654	0.809
57	Middle	6	2013	7	9	0.648	0.684	1.000
58	Middle	6	2014	16	20	0.617	0.631	0.836
59	Middle	6	2016	16	20	0.614	0.627	0.857
60	Middle	6	2017	30	39	0.591	0.610	0.855
61	Late	7	2007	27	37	0.640	0.699	0.901
62	Late	7	2008	58	93	0.582	0.618	0.755
63	Late	7	2009	44	66	0.592	0.636	0.934
64	Late	7	2010	33	59	0.570	0.609	0.560
65	Late	7	2011	32	46	0.567	0.609	0.756
66	Late	7	2012	11	15	0.663	0.718	0.862
67	Late	7	2013	32	40	0.625	0.641	0.706
68	Late	7	2014	20	31	0.626	0.657	0.605
69	Late	7	2016	7	10	0.655	0.735	1.000
70	Late	7	2017	25	37	0.571	0.612	0.693
71	Late	8	2007	22	33	0.651	0.714	0.945
72	Late	8	2008	42	66	0.617	0.659	0.850
73	Late	8	2009	28	46	0.581	0.629	0.803
74	Late	8	2010	22	36	0.590	0.705	1.000
75	Late	8	2011	30	44	0.608	0.702	1.000
76	Late	8	2012	9	11	0.823	0.904	0.949
77	Late	8	2013	21	28	0.643	0.717	0.896
78	Late	8	2014	11	22	0.622	0.709	0.952
79	Late	9	2007	18	26	0.629	0.670	0.924
80	Late	9	2008	45	80	0.574	0.625	0.983
81	Late	9	2009	24	47	0.585	0.648	0.991
82	Late	9	2010	29	55	0.602	0.653	0.886
83	Late	9	2011	18	49	0.590	0.655	0.669
84	Late	9	2012	14	16	0.657	0.712	1.000
85	Late	9	2013	12	29	0.619	0.796	1.000
86	Late	9	2014	25	32	0.604	0.620	0.708

Community	Stage	Plot	Year	$s$	$S$	$\omega(\mathbf{A})$	$\omega(\boldsymbol{\alpha})$	$r$
87	Late	9	2016	12	15	0.603	0.608	0.125
88	Late	9	2017	24	37	0.535	0.584	0.675

### A3.2 Reconstruction of regional pools

Here we describe in detail how we reconstructed the different types of regional pools (**A**) for the each of the 88 empirical combinations of herbivore species. Because we do not have access to the true empirical species pool, the idea for these analyzes was to test the robustness of our results to different reconstruction methods of the species pool. We first describe the species pool reconstruction method used in the main text, which takes into account both the year and successional stage of the combination of herbivores. With this method, we first built a binary species pool matrix **B** for a given plot ( $b_{ij} = 1$  if plant species  $i$  is consumed by herbivore species  $j$  and  $b_{ij} = 0$ , otherwise) by merging all herbivore species from the same year and successional stage that fed on plant species from that given plot. For example, for plot 5 (2011, middle successional stage; Table A1) we merged all herbivore species present in both 2011 and middle successional stage and then created matrix **B** using these herbivores and only plants from plot 5 (Figure 3A). The rationale behind this approach is that the species pool for a given plot must contain all herbivores that could potentially feed on the resources (i.e., plant species) present at this plot. Then, we inferred the herbivore interaction matrix **A** that is mediated by their feeding interactions with plants by computing the monopartite projection of the binary bipartite matrix **B** and then scaling the off-diagonal elements ( $a_{ij}, i \neq j$ ) according to one of the methods described in Section A4. Our assumption is that herbivores can exert negative effects on each other through a variety of mechanisms, including induction of plant defenses or attraction of parasitoids and predators (Denno et al., 2000, Ohgushi, 2005, Pallini et al., 1998, Redman and Scriber, 2000).

In addition to the species pool reconstruction described above, we also built species pools by merging herbivore species from the same year or same successional stage. For the former approach, we built the binary species pool matrix **B** for a given plot by merging all herbivore species from the same year that fed on plant species from that given plot. For example, for plot 5 we merged all herbivore species present in 2011 and created matrix **B** using these herbivores and only plants from plot 5 (Figure A5). For the latter approach, we built the binary species pool matrix **B** for a given plot by merging all herbivore species from the same successional stage that fed on plant species from that given plot. For example, for plot 5 we merged all herbivore species present in the middle successional stage and created matrix **B** using these herbivores and only plants from plot 5 (Figure A6). Our results in Figures A5 and A6 are qualitatively the same as the result in Figure 3 in the main text.

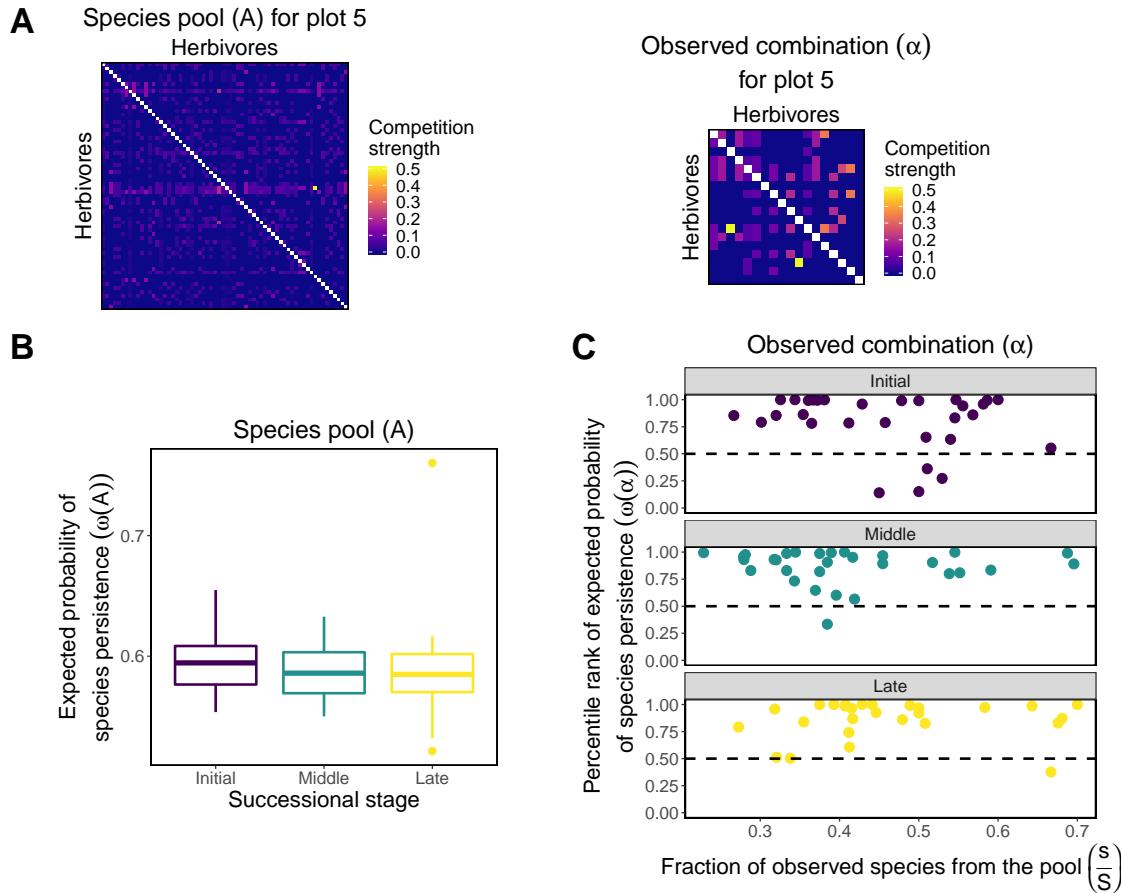


Figure A5: Empirical results building the regional pool by merging herbivores from the same year. **(A)** Left: an illustrative example of a reconstructed regional pool (matrix  $\mathbf{A}$ ) with  $S = 64$  competing herbivore species mediated by plant resources from plot 5 (2011, middle successional stage). Right: the observed combination (matrix  $\alpha$ ) with the  $s = 18$  competing herbivores from plot 5. Colors represent the competition strength  $\alpha_{ij}$ , which is proportional to the number of shared plant species. Diagonal elements ( $\alpha_{ii}$ ) are all set to 1 and are not colored to improve visualization. **(B)** The expected probability of species persistence of the 88 reconstructed species pools ( $\omega(\mathbf{A})$ ) separated by successional stage. Boxplots denote the median and interquartile range. **(C)** Percentile rank values ( $r$ ) of the expected probability of species persistence of observed combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and successional stage. Note that each percentile rank is computed by comparing a given observed combination within the population formed by 1,000 sampled potential combinations with the same number of species. Each point corresponds to one of 88 different empirical communities. The percentage of points above the median (dashed line) for each successional stage is: initial: 86.7%, middle: 96.7%, and late: 96.4% ( $p < 0.0001$  for all successional stages for one-sided unbiased binomial tests). In **(B)** and **(C)** colors indicate different successional stages: the lighter the later.

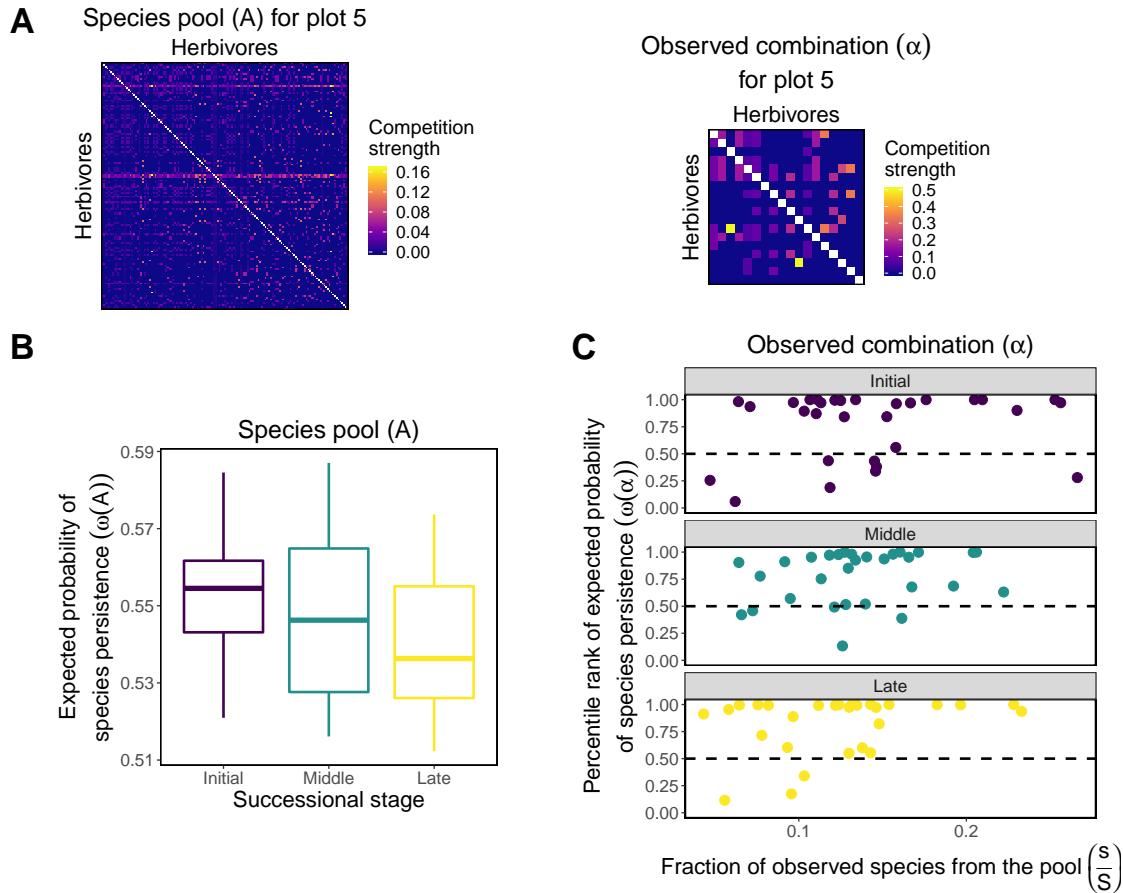


Figure A6: Empirical results building the regional pool by merging herbivores from the same stage. **(A)** Left: an illustrative example of a reconstructed regional pool (matrix  $\mathbf{A}$ ) with  $S = 139$  competing herbivore species mediated by plant resources from plot 5 (2011, middle successional stage). Right: the observed combination (matrix  $\boldsymbol{\alpha}$ ) with the  $s = 18$  competing herbivores from plot 5. Colors represent the competition strength  $\alpha_{ij}$ , which is proportional to the number of shared plant species. Diagonal elements ( $\alpha_{ii}$ ) are all set to 1 and are not colored to improve visualization. **(B)** The expected probability of species persistence of the 88 reconstructed species pools ( $\omega(\mathbf{A})$ ) separated by successional stage. Boxplots denote the median and interquartile range. **(C)** Percentile rank values ( $r$ ) of the expected probability of species persistence of observed combinations ( $\omega(\boldsymbol{\alpha})$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and successional stage. Note that each percentile rank is computed by comparing a given observed combination within the population formed by 1,000 sampled potential combinations with the same number of species. Each point corresponds to one of 88 different empirical communities. The percentage of points above the median (dashed line) for each successional stage is: initial: 73.3%, middle: 83.3%, and late: 89.3% ( $p < 0.01$  for all successional stages for one-sided unbiased binomial tests). In **(B)** and **(C)** colors indicate different successional stages: the lighter the later.

## A4 Parameterization of empirical interaction strengths

Inferring interacting coefficients is a challenging problem in ecology (Carrara et al., 2015, Harris, 2016, Morueta-Holme et al., 2016). In contrast to inferring interaction coefficients from co-occurrence data, which can yield inconsistent results depending on the method used (Barner et al., 2018), in this study we used a niche framework without tuning parameters as a simple and biologically sound estimation of the herbivore interaction coefficients. In the main text, we described a framework based on the normalized monopartite projection of the empirical bipartite matrices. Here, we explain this inference method in detail and report our results using a different method to infer interaction coefficients, which uses a tuning parameter. Furthermore, we describe our prediction analyzes intended to validate the inferred coefficients.

### A4.1 Normalizing by columns

As described in the main text, we computed the monopartite projection of the empirical bipartite matrices  $\mathbf{B}$  in order to obtain an interaction matrix  $\mathbf{A} = \mathbf{B}^\top \mathbf{B}$ . Note that the projection is done over the herbivore layer. In matrix  $\mathbf{A}$ , an element  $a_{ij}$  gives the number of shared plant species between herbivore species  $i$  and  $j$  and an element  $a_{ii}$  gives the total number of plant species consumed by herbivore  $i$ . In order to obtain a Lyapunov diagonally stable matrix  $\mathbf{A}$  to be used as a reconstructed species pool, we normalized the columns of matrix  $\mathbf{A}$  to 1 (i.e.,  $a_{ij} = \frac{a_{ij}}{\sum_{i=1}^S a_{ij}}$ ) and set its diagonal elements to 1 (i.e.,  $a_{ii} = 1$ ) afterwards. Thus,  $\mathbf{A}$  represents a set of interacting species in which each interaction coefficient  $a_{ij}$  is proportional to the number of shared plant resources between herbivores  $i$  and  $j$ .

Here, we prove that all matrices generated by the procedure above are guaranteed to be diagonally dominant, and consequently, globally stable. That is, the sum of its off-diagonal elements must be less than its diagonal element for each row (i.e.,  $\sum_{j \neq i} m_{ij} < m_{ii} \forall i$ ). In our case,  $\mathbf{A}$  is a monopartite projection and, therefore, is a symmetric matrix satisfying  $a_{ij} \leq a_{ii} \forall i$  and  $j$ . Let us consider the case for which  $a_{ij} = a_{ii} \forall i$  and  $j$ . Let us consider the matrix  $\mathbf{M}$  obtained by normalizing  $\mathbf{A}$  by diving each off-diagonal element by its column sum (i.e.,  $m_{ij} = \frac{a_{ij}}{\sum_i a_{ij}}$ ) and then setting diagonal elements to 1 (i.e.,  $m_{ii} = 1$ ). Then, we have the following expression for the sum of the off-diagonal elements of a given row  $i$ :

$$\sum_{j \neq i} \frac{a_{ij}}{\sum_i a_{ij}} = \frac{1}{S a_{ii}} \sum_{j \neq i} a_{ij} \quad (\text{A12})$$

$$= \frac{(S-1)a_{ii}}{S a_{ii}} \quad (\text{A13})$$

$$= \frac{(S-1)}{S} \quad (\text{A14})$$

Thus, when  $a_{ij} = a_{ii} \forall i$  and  $j$ , the sum of the off-diagonal elements of a given row  $i$  of  $\mathbf{M}$  is  $\frac{(S-1)}{S}$ , which is less than 1. Therefore, we have the desired condition for diagonal dominance:  $\sum_{j \neq i} m_{ij} = \frac{(S-1)}{S} < m_{ii} = 1 \forall i$ .

#### A4.2 Using a tuning parameter

We used an additional method to parameterize the empirical interaction coefficients, which uses a tuning parameter (Saavedra et al., 2014). To do so, we first computed the monopartite projection  $\mathbf{A}$  of the empirical bipartite matrices as described above. Then, we rescaled all off-diagonal elements  $a_{ij}$  by multiplying them by a tuning parameter  $\mu$  and leaving diagonal elements  $a_{ii}$  unchanged. To set the value of  $\mu$ , we first found the highest value  $\mu'$  for which  $\mathbf{A} + \mathbf{A}^\top$  is a positive definite matrix, where  $a_{ij} = \mu' a_{ij}$  for  $i \neq j$ . Recall that a positive definite matrix is Volterra-dissipative (Goh, 1977). Thus,  $\mu'$  rescales the off-diagonal elements  $a_{ij}$  in such a way that matrix  $\mathbf{A}$  has the highest possible competition strength before losing global stability in the LV model. Finally, we set  $\mu = \frac{\mu'}{2}$  to reduce interaction strength and reduce numerical instabilities when computing the expected probability of species persistence of regional pools ( $\omega(\mathbf{A})$ ) and subsets of regional pools ( $\omega(\boldsymbol{\alpha})$ ). Note that dividing  $\mu'$  by a different number would not change our results since regional pools, observed combinations, and potential combinations are all rescaled in the same way. It is also important to note that both methods described here to parameterize interaction coefficients are phenomenological and parsimonious approximations of the plant-mediated negative effects between herbivore species. The empirical results for the expected probability of species persistence of the regional pool ( $\omega(\mathbf{A})$ ) and the percentile rank of the expected probability of species persistence of observed combinations ( $\omega(\boldsymbol{\alpha})$ ) using this different parameterization of interaction coefficients are qualitatively the same as the result in Figure 3 in the main text (see Figure A7 below).

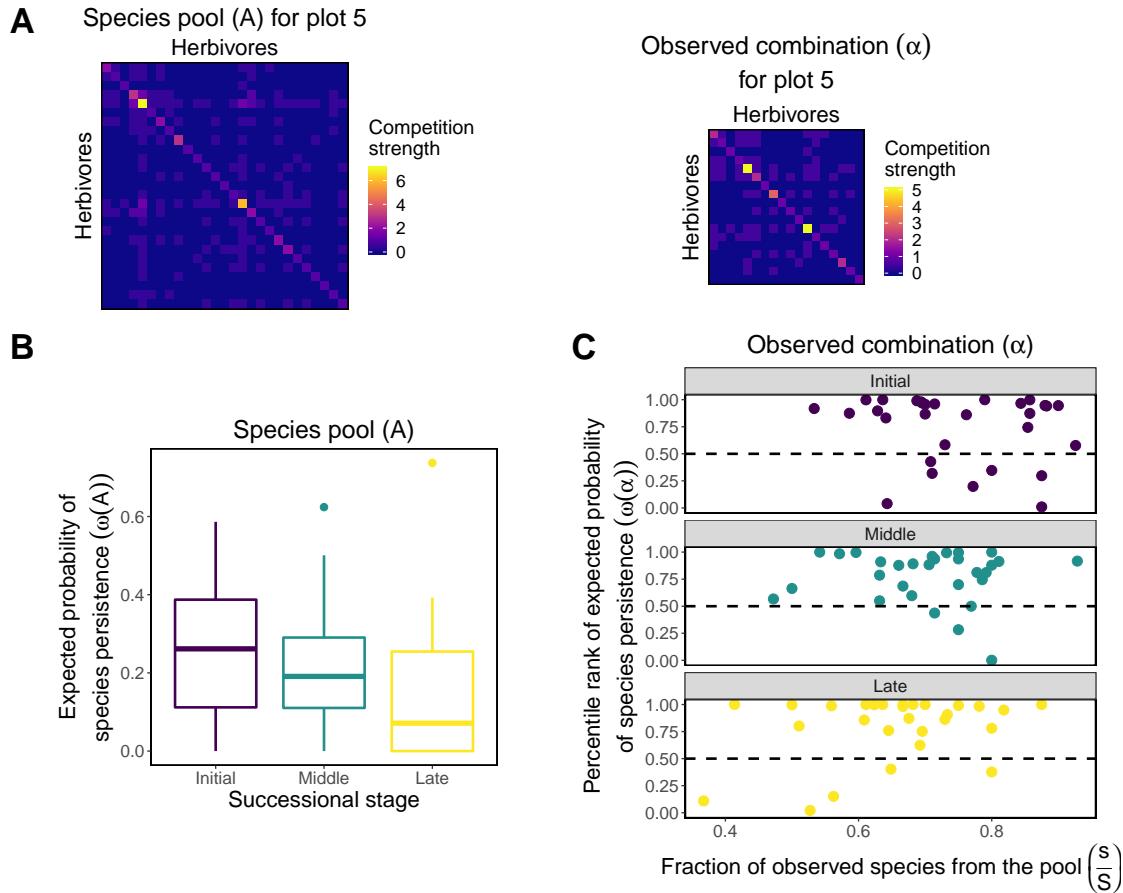


Figure A7: Empirical results using a tuning parameter instead of normalizing matrices by column sum. **(A)** Left: an illustrative example of a reconstructed regional pool (matrix  $\mathbf{A}$ ) with  $S = 27$  competing herbivore species mediated by plant resources from plot 5 (2011, middle successional stage). Right: the observed combination (matrix  $\boldsymbol{\alpha}$ ) with the  $s = 18$  competing herbivores from plot 5. Colors represent the competition strength  $\alpha_{ij}$ , which is proportional to the number of shared plant species. Diagonal elements ( $\alpha_{ii}$ ) are the number of plant species consumed by each herbivore. **(B)** The expected probability of species persistence of the 88 reconstructed regional pools ( $\omega(\mathbf{A})$ ) separated by successional stage. Boxplots denote the median and interquartile range. **(C)** Percentile rank values ( $r$ ) of the expected probability of species persistence of observed combinations ( $\omega(\boldsymbol{\alpha})$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and successional stage. Note that each percentile rank is computed by comparing a given observed combination within the population formed by 1,000 sampled potential combinations with the same number of species. Each point corresponds to one of 88 different empirical communities. The percentage of points above the median (dashed line) for each successional stage is: initial: 76.7%, middle: 86.7%, and late: 82.1% ( $p < 0.01$  for all successional stages for one-sided unbiased binomial tests). In **(B)** and **(C)** colors indicate different successional stages: the lighter the later.

#### A4.3 Prediction analyzes

To validate our inferred interaction coefficients, we performed simulations parameterized with the inferred coefficients and tested the accuracy of out-of-sample predictions of temporal changes in herbivore species composition. To do so, we first inferred the herbivore competition coefficients using the resource-partitioning method described in Section A4.1 for a given observed combination  $\alpha$  of interacting species from a given year and successional stage (e.g., plot 5, 2011, middle successional stage). Following the resource-partitioning method, we simulated LV dynamics (Equation A1) by assuming that each carrying capacity  $K_i$  is proportional to the number of plant species consumed by herbivore species  $i$ . We set the intrinsic growth rates to  $r_i = K_i$  and initial conditions to  $N_i = 1$  for every species  $i$  and verified that these conditions did not affect the simulation outcomes because of the global stability condition imposed on the matrix  $\alpha$ . After the dynamics reached stationarity, we classified a herbivore species as extinct if its abundance was below a fixed threshold of  $\epsilon = 0.02$ . Then, we compared our classification of species (i.e., survived or extinct) with the observed herbivore species composition for the next year. For example, we performed the simulation using plot 5, 2011, middle successional stage and compared our results with the true species composition in plot 5, 2012, middle successional stage (i.e., predicted out-of-sample data). In total, we performed 78 predictions of the species composition in the next year using LV simulation outcomes from the previous year.

We computed four measures of prediction accuracy: (1) true positives (TP) were computed as the number of species that we classified as "survived" and were present in the next year, (2) false positives (FP) were computed as the number of species that we classified as "survived" and were not present in the next year, (3) true negatives (TN) were computed as the number of species that we classified as "extinct" and were not present in the next year, and (4) false negatives (FN) were computed as the number of species that we classified as "extinct" and were present in the next year. Then, we measured the true positive rate as  $\frac{TP}{TP+FN}$  and the false positive rate as  $\frac{FP}{FP+TN}$ . Importantly, a given prediction is better than a random guess if the true positive rate is higher than the false positive rate (Aggarwal, 2015). It is also important to note that we tuned the threshold  $\epsilon$  in order to maximize our prediction accuracy as is typically done in classification algorithms (Aggarwal, 2015).

Figure A8 shows the true positive rate as a function of the false positive rate for all our 78 predictions of temporal changes in species composition. We found that 87.3% of our predictions

have a higher true positive rate than a false positive rate (blue points in Figure A8). This result indicates that our predictions of species survival and extinction were better than a random guess for 87.3% of the empirical combinations of interacting herbivores. Moreover, this indicates that our inferred interaction coefficients are a valid phenomenological representation of the dynamics between herbivore species in this empirical system.

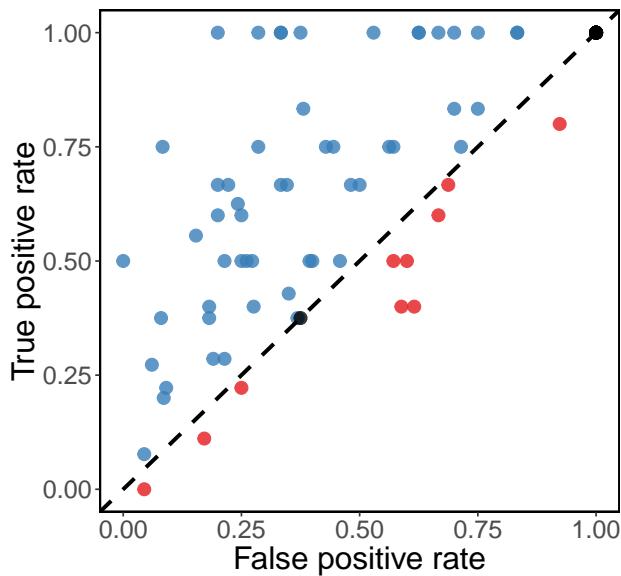


Figure A8: True positive rate and false positive rate for each out-of-sample prediction of the temporal change in species composition in our empirical data set. Each point corresponds to a prediction of the species composition of a given year (e.g., plot 5, 2012, middle successional stage) by performing a simulation of the Lotka-Volterra dynamics parameterized by the plant-herbivore interactions from the previous year (e.g., plot 5, 2011, middle successional stage). A given prediction is better than a random guess if the true positive rate is higher than the false positive rate, which is indicated as blue points above the dashed line. Overall, 87.3% of all predictions have higher true positive rate than false positive rate (blue points).

## A5 Modifications of the classic Lotka-Volterra model

As mentioned in the main text, our structural approach can be applied to any model topologically equivalent to the classic LV model (Equation (A1)). Topologically equivalent means that the unstable and stable fixed points in the classic LV model must be mapped into a pair of unstable and stable fixed points in the modified model (Cenci and Saavedra, 2018). Importantly, previous work has shown that this approach can also be extended to some modifications of the LV model with Type II functional response and to stochastic dynamics (Cenci and Saavedra, 2018). To confirm this, we repeated our analysis, but instead of using the classic LV model (Equation A1), we used a Type II LV model (Fig. A9) and a stochastic LV model (Fig. A10). The Type II LV model is given by (Case, 2000):

$$\frac{dN_i}{dt} = N_i \frac{r_i}{K_i} \left( K_i - \sum_{j=1}^S a_{ij} \frac{aN_j}{1 + ahN_i} \right), \quad (\text{A15})$$

where, following the consumer-resource terminology,  $a$  is the attack rate and  $h$  is the handling time. Here, we assume for simplicity that these two parameters are equal to 1 (Hastings and Powell, 1991). The stochastic LV model is given by adding a demographic stochasticity term to Equation (A1):

$$\frac{dN_i}{dt} = N_i \frac{r_i}{K_i} \left( K_i - \sum_{j=1}^S a_{ij} N_j \right) + \frac{v(t)}{\sqrt{S}}, \quad (\text{A16})$$

where  $v(t)$  is a Gaussian white noise with zero mean and correlations given by  $\langle v_i(t)v_j(t') \rangle = \beta_{ij}\delta(t-t')$ , where  $\beta_{ij} = N_i \frac{r_i}{K_i} \left( K_i - \sum_{j=1}^S a_{ij} N_j \right)$   $\forall i = j$  and zero otherwise, while  $\delta$  characterizes the white noise (McKane et al., 2014). Importantly, although LV dynamics was simulated using the models above, the feasibility domain and the expected probability of species persistence was computed using only matrix **A** as defined in Section A1. Overall, Figures A9 and A10 below show that our results for the modified LV models are qualitatively the same as the results in Figure 2 in the main text.

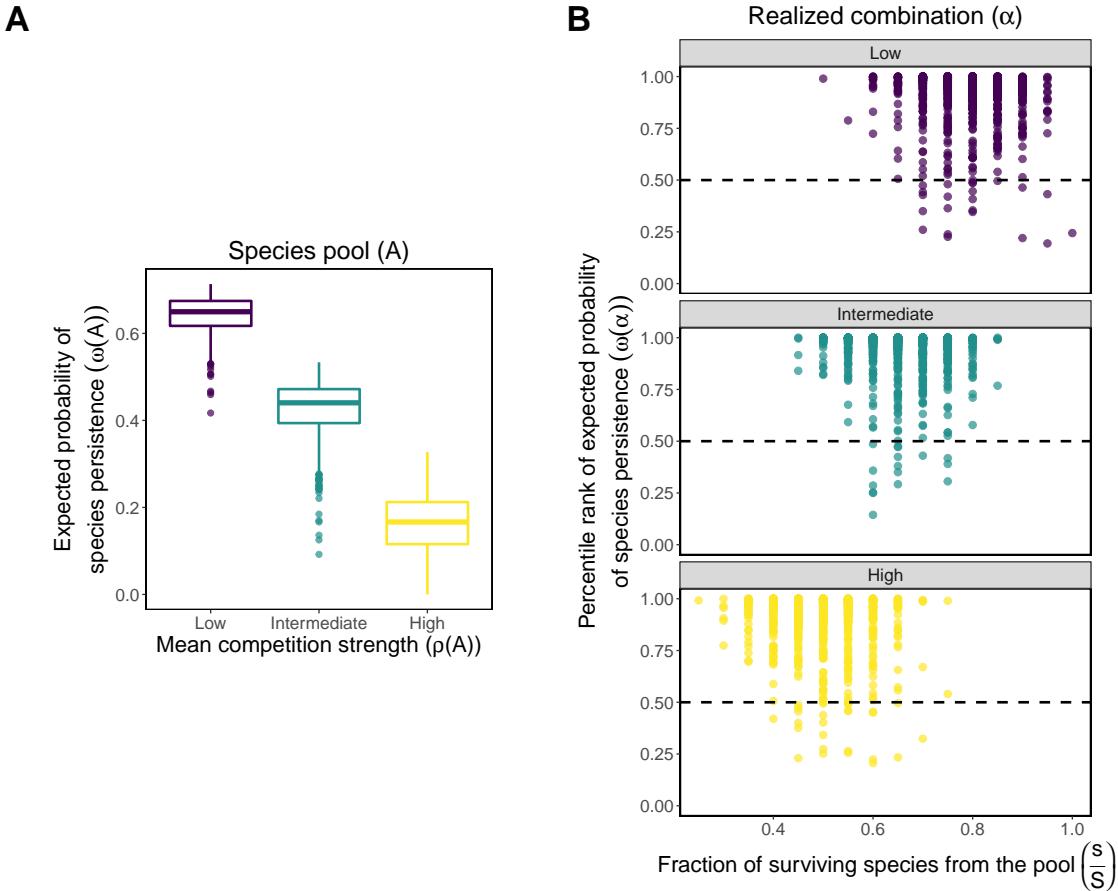


Figure A9: Theoretical results obtained by performing simulations using a LV model with Type II functional response. **(A)** The expected probability of species persistence under changing environments of 500 model-generated regional pools ( $\omega(\mathbf{A})$ ) with  $S = 20$  for each value of mean competition strength (low:  $\rho(\mathbf{A}) = 0.025$ , intermediate:  $\rho(\mathbf{A}) = 0.05$ , and high:  $\rho(\mathbf{A}) = 0.1$ ). Boxplots denote the median and interquartile range. **(B)** Percentile rank values ( $r$ ) of the expected probability of species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and level of mean competition strength. Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of  $\mathbf{K}$ -vectors) per level of mean competition strength. The percentage of points above the median (dashed line) for each value of mean competition strength is: low: 97.7%, intermediate: 98.4%, and high: 97.3% ( $p < 0.0001$  for all levels of mean competition strength for one-sided unbiased binomial tests). In **(A)** and **(B)** colors indicate different levels of mean competition strength: the lighter the higher.

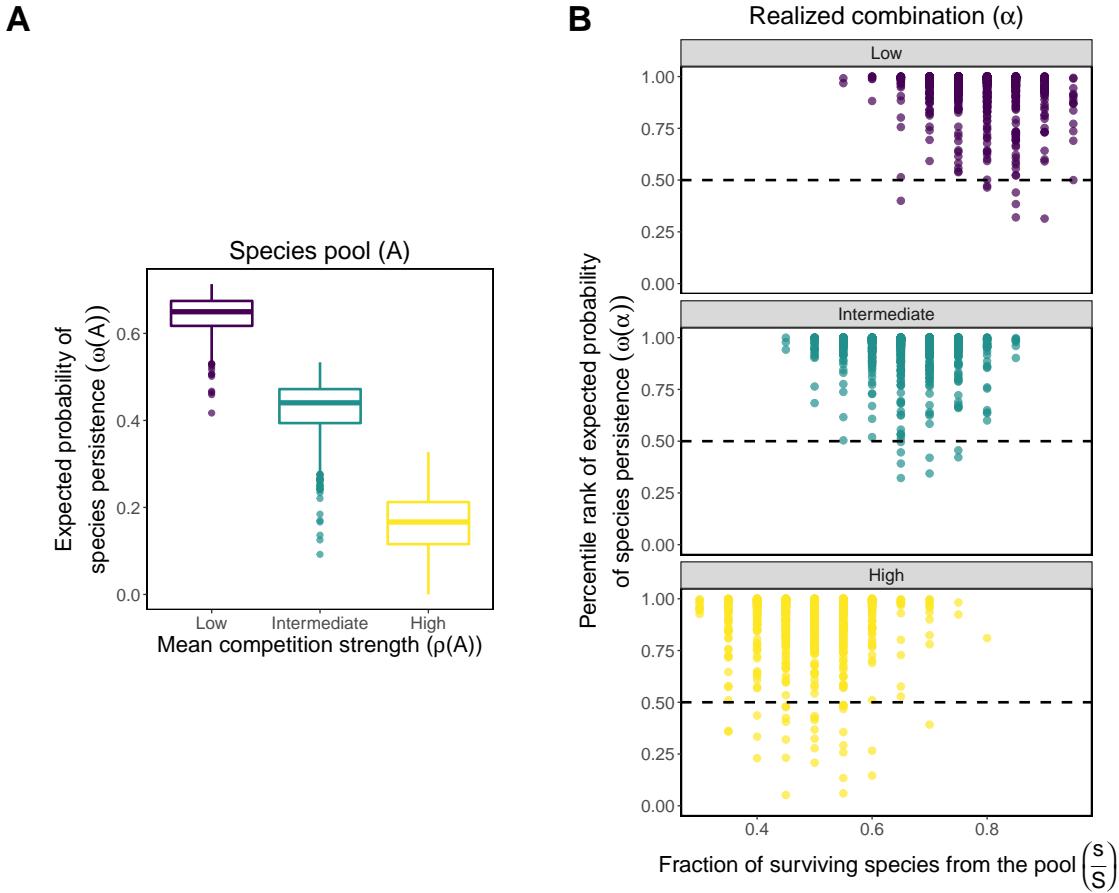


Figure A10: Theoretical results obtained by performing simulations using a stochastic LV model. **(A)** The expected probability of species persistence under changing environments of 500 model-generated regional pools ( $\omega(\mathbf{A})$ ) with  $S = 20$  for each value of mean competition strength (low:  $\rho(\mathbf{A}) = 0.025$ , intermediate:  $\rho(\mathbf{A}) = 0.05$ , and high:  $\rho(\mathbf{A}) = 0.1$ ). Boxplots denote the median and interquartile range. **(B)** Percentile rank values ( $r$ ) of the expected probability of species persistence of realized combinations ( $\omega(\alpha)$ ) as a function of the fraction of persistent species ( $\frac{s}{S}$ ) and level of mean competition strength. Note that each percentile rank is computed by comparing a given realized combination within the population formed by 500 sampled potential combinations with the same number of species. Each point corresponds to one of 1,000 different simulations (i.e., different directions of  $\mathbf{K}$ -vectors) per level of mean competition strength. The percentage of points above the median (dashed line) for each value of mean competition strength is: low: 99.2%, intermediate: 99.2%, and high: 96.8% ( $p < 0.0001$  for all levels of mean competition strength for one-sided unbiased binomial tests). In **(A)** and **(B)** colors indicate different levels of mean competition strength: the lighter the higher.

## References

- Aggarwal, C. C., 2015. Data mining: the textbook. Springer.
- Allesina, S. and Tang, S. 2015. The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology* 57:63–75.
- Barner, A. K., Coblenz, K., Hacker, S., and Menge, B. 2018. Fundamental contradictions among common estimates of non-trophic species interactions. *Journal of The Royal Society Interface* 99:557–566.
- Boege, K., Villa-Galaviz, E., López-Carretero, A., Pérez-Ishiwara, R., Zaldivar-Riverón, A., Ibarra, A., and del Val, E. 2019. Temporal variation in the influence of forest succession on caterpillar communities: A long-term study in a tropical dry forest. *Biotropica* 51:529–537.
- Carrara, F., Giometto, A., Seymour, M., Rinaldo, A., and Altermatt, F. 2015. Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity. *Methods in Ecology and Evolution* 6:895–906.
- Case, T. J., 2000. An Illustrated Guide to Theoretical Ecology. Oxford Univ. Press, Oxford.
- Cenci, S. and Saavedra, S. 2018. Structural stability of nonlinear population dynamics. *Phys. Rev. E* 97:012401.
- Cross, G. 1978. Three types of matrix stability. *Linear algebra and its applications* 20:253–263.
- Denno, R., Peterson, M., Gratton, C., Cheng, J., Langellotto, G., Huberty, A., and Finke, D. 2000. Feeding-induced changes in plant quality mediate interspecific competition between sap-feeding herbivores. *Ecology* 81:1814–1827.
- Genz, A. and Bretz, F., 2009. Computation of multivariate normal and t probabilities, volume 195. Springer Science & Business Media.
- Goh, B. S. 1977. Global stability in many-species systems. *The American Naturalist* 111:135–143.
- Harris, D. J. 2016. Inferring species interactions from co-occurrence data with markov networks. *Ecology* 97:3308–3314.
- Hastings, A. and Powell, T. 1991. Chaos in a three-species food chain. *Ecology* 72:896–903.

- Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., and Francis, C. M. 2004. Identification of birds through dna barcodes. PLoS Biology 2:e312.
- Jordano, P. 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. The American Naturalist 129:657–677.
- Logofet, D. O., 1993. Matrices and Graphs: Stability Problems in Mathematical Ecology. CRC Press.
- MacArthur, R. and Levins, R. 1967. The limiting similarity, convergence, and divergence of coexisting species. Am. Nat. 101:377–385.
- McKane, A. J., Biancalani, T., and Rogers, T. 2014. Stochastic pattern formation and spontaneous polarisation: The linear noise approximation and beyond. Bulletin of Mathematical Biology 76:895–921.
- Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B. J., Peet, R. K., Ott, J. E., Violle, C., Enquist, B. J., Jørgensen, P. M., and Svenning, J.-C. 2016. A network approach for inferring species associations from co-occurrence data. Ecography 39:1139–1150.
- Ohgushi, T. 2005. Indirect interaction webs: herbivore-induced effects through trait change in plants. Annu. Rev. Ecol. Evol. Syst. 36:81–105.
- Pallini, A., Janssen, A., Sabelis, M. W., et al. 1998. Predators induce interspecific herbivore competition for food in refuge space. Ecology Letters 1:171–177.
- Redman, A. M. and Scriber, J. M. 2000. Competition between the gypsy moth, lymantria dispar, and the northern tiger swallowtail, papilio canadensis: interactions mediated by host plant chemistry, pathogens, and parasitoids. Oecologia 125:218–228.
- Ribando, J. M. 2006. Measuring solid angles beyond dimension three. Discrete & Computational Geometry 36:479–487.
- Rohr, R. P., Saavedra, S., Peralta, G., Frost, C. M., Bersier, L.-F., Bascompte, J., and Tylianakis, J. M. 2016. Persist or produce: a community trade-off tuned by species evenness. Am. Nat. 188:411–422.

- Saavedra, S., Cenci, S., del Val, E., Boege, K., and Rohr, R. P. 2017a. Reorganization of interaction networks modulates the persistence of species in late successional stages. *J. of Animal Ecology* 86:1136–1146.
- Saavedra, S., Rohr, R. P., Bascompte, J., Godoy, O., Kraft, N. J., and Levine, J. M. 2017b. A structural approach for understanding multispecies coexistence. *Ecological Monographs* 87:470–486.
- Saavedra, S., Rohr, R. P., Gilarranz, L. J., and Bascompte, J. 2014. How structurally stable are global socioeconomic systems? *J. R. Soc. Interface* 11:20140693.
- Serván, C. A., Capitán, J. A., Grilli, J., Morrison, K. E., and Allesina, S. 2018. Coexistence of many species in random ecosystems. *Nature Ecology & Evolution* 2:1237.
- Song, C., Ahn, S. V., Rohr, R. P., and Saavedra, S. 2020. Towards a probabilistic understanding about the context-dependency of species interactions. *Trends in Ecology & Evolution* 35:384–396.
- Song, C., Rohr, R. P., and Saavedra, S. 2018. A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *J. of Theoretical Biology* 450:30–36.
- Takeuchi, Y., 1996. Global dynamical properties of Lotka-Volterra systems. World Scientific.
- Vandermeer, J. H. 1975. Interspecific competition: a new approach to the classical theory. *Science* 188:253–255.