Understanding the impact of third-party species on pairwise coexistence

Jie Deng[1], Washington Taylor[2], Serguei Saavedra[1*]

[1]Department of Civil and Environmental Engineering, MIT,
77 Massachusetts Av., 02139 Cambridge, MA, USA

[2]Center for Theoretical Physics, MIT,
77 Massachusetts Av., 02139 Cambridge, MA, USA

[*] To whom correspondence should be addressed. E-mail: sersaa@mit.edu

Keywords: species interactions | feasibility | probability | geometry | gut microbiota | dimensionality

Author contributions: J.D. and S.S designed research; J.D was responsible for investigation; all authors were responsible for conceptualization, formal analysis, methodology, software, validation, writing, reviewing, and editing; S.S was responsible for funding acquisition and supervision.

Competing financial interests The authors declare no competing financial interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data accessibility The code supporting the results is novel and will be archived on Zenodo upon publication. This code is available to reviewers on Github: `https://github.com/MITEcology/Deng_2022_1`

Preprint number: MIT-CTP/5439 (internal use only)

Acknowledgments: We thank the Editor Sara Mitri, Aming Li, and two anonymous reviewers for their highly constructive comments that led to the improvement of this work. S.S. was supported by NSF DEB-2024349. W.T. was supported in part by a grant from the Schmidt Futures Foundation.

1

Abstract

The persistence of virtually every single species depends on both the presence of other species and the specific environmental conditions in a given location. Because in natural settings many of these conditions are unknown, research has been centered on finding the fraction of possible conditions (probability) leading to species coexistence. The focus has been on the persistence probability of an entire multispecies community (formed of either two or more species). However, the methodological and philosophical question has always been whether we can observe the entire community and, if not, what the conditions are under which an observed subset of the community can persist as part of a larger multispecies system. Here, we derive long-term (using analytical calculations) and short-term (using simulations and experimental data) system-level indicators of the effect of third-party species on the coexistence probability of a pair (or subset) of species under unknown environmental conditions. We demonstrate that the fraction of conditions incompatible with the possible coexistence of a pair of species tends to become vanishingly small within systems of increasing numbers of species. Yet, the probability of pairwise coexistence in isolation remains approximately the expected probability of pairwise coexistence in more diverse assemblages. In addition, we found that when third-party species tend to reduce (resp. increase) the coexistence probability of a pair, they tend to exhibit slower (resp. faster) rates of competitive exclusion. Long-term and short-term effects of the remaining third-party species on all possible specific pairs in a system are not equally distributed, but these differences can be mapped and anticipated under environmental uncertainty.

Author Summary

It is debated whether the frequency with which two species coexist in isolation or within a single environmental context is representative of their coexistence expectation within larger multispecies systems and across different environmental conditions. Here, using analytical calculations, simulations, and experimental data, we show why and how third-party species can provide the opportunity for pairwise coexistence regardless of whether a pair of species can coexist in isolation across different environmental conditions. However, we show that this opportunity is not homogeneously granted across all pairs within the same system. We provide a framework to understand and map the long-term and short-term effects that third-party species have on the coexistence of each possible subset in a multispecies system.

Introduction

The persistence of virtually all living organisms on Earth depends to a greater or lesser extent on the presence of other living organisms and on the environmental (abiotic and biotic) conditions present in a given place and time (1, 2). This observation has established a rich research program in quantifying pairwise interactions and their impact on pairwise coexistence across environmental gradients (3–8). In this line, observational work has been focused on finding the frequency of occurrence of multiple pairs of species across different environments either in isolation or within different systems of multiple species (9–16). For example, studies have long debated whether there exists in nature pairs of species whose niches forbid their coexistence regardless of the environment—known as the checkerboard hypothesis (11, 17). Instead, experimental studies have shown that pairwise coexistence strongly depends on the details of both the system and the environment (10, 12, 18, 19). In fact, it has been shown that the coexistence expectations that may operate under controlled conditions (or unique environments) (12, 14) do not necessarily operate under uncontrolled conditions (or diverse environments) (5, 20). These observational results have shown that in order to implement successful interventions in ecological communities, it is essential to develop a testable theory to be able to explain why and how emergent processes at the system level can affect the possibility that a given pair (or community) of species will coexist under different abiotic and biotic environments.

Because the environmental conditions acting on species are typically unknown and diverse in natural settings, the majority of theoretical work has been centered on deriving the fraction of conditions (set of parameter values) leading to the persistence of a community (formed by either two or more species) (8, 21–23). Yet, species are seldom in isolation and the coexistence of a specific combination of multiple species is expected to be rare in a random environment (5, 24, 25). Moreover, the methodological and philosophical question has always been whether we can observe the entire community and, if not, what the conditions are under which an observed subset of the community can persist as part of a larger multispecies system. The answer to this question, however, may also be context-dependent due to the presence and absence of third-party species, which can be acting as ecosystem engineers (26, 27). Thus, it is unclear how far knowledge of pairwise coexistence in isolation (or full community persistence) can take us while investigating ecological systems, and how we should compare the effects of third-party species on the coexistence probability of a pair (or community) of species.

In this line, ecological research has suggested that species embedded into larger multispecies systems (more than two species) may experience higher-order effects, i.e., the effect of species i on the per capita growth rate of species j might itself depend on the abundance of a third species k due to either compensatory effects, supra-additivity, trait-mediated effects, functional effects, meta-community effects, or indirect effects (8, 28–31). Unfortunately, additional research has shown that it is virtually impossible to derive those potential effects since many parameter combinations can equally explain the observed ecological dynamics—a problem known as structural identifiability (32, 33). In fact, many of the interactions measured under multispecies systems are not direct effects, as studies often

believe (15, 16), but chains of direct effects (34, 35). Moreover, measurements of species interactions are expected to change as a function of the type of data (i.e., snapshot or time), experiment (e.g., long term or short term), perturbations (e.g., pulse or press), and dimensionality, leading to inconsistencies and inference issues (32, 36, 37). Thus, it is unclear whether the frequency with which two or more species coexist in isolation is representative of their coexistence expectation within larger multispecies systems and across different environments.

Here, we follow a geometric and probabilistic analysis based on nonlinear dynamics at equilibrium to estimate the effect of third-party species (the larger multispecies system) on the coexistence of a pair (community) of species under unknown environmental conditions using only information from pairwise species interactions. We estimate these system-level effects over the long term (using analytical calculations) and over the short term (using simulations and available experimental data on gut microbiota). Building on generalized Lotka-Volterra (gLV) dynamics (8), we provide a geometric understanding regarding whether the conditions limiting the coexistence of a pair of species in isolation (the species pool consists only of this pair) remain with third-party species (when the species pool increases). Next, we derive a system-level indicator of the extent to which third-party species can affect the probability of coexistence of a pair relative to its probability in isolation. Then, we compare the numerical and experimental effects to the analytical expectation in order to distinguish the role that third-party species play in shaping pairwise coexistence under short-term and long-term behavior, respectively. Finally, we show how our theory can be used to provide a cartographic representation (38) of the long-term and short-term effects acting on each pair within a multispecies system under environmental uncertainty. While our focus is on pairwise coexistence (following traditional work in ecology), our methodology and results are scalable to any subset (community) dimension.

Methods

Traditional indicators of pairwise coexistence

Our theoretical framework is based on the tractability of the gLV dynamics (33), where the per-capita growth rate of species $i \in \{1, \cdots, |\mathcal{S}|\} \equiv \mathcal{S}$ is written as $f_i(\mathbf{N}; \boldsymbol{\theta}) = \theta_i - \sum_{j \in \mathcal{S}} a_{ij} N_j$, where the vector $\mathbf{N} = (N_1, \ldots, N_{|\mathcal{S}|})^\intercal$ represents the density of all species in the system $\mathcal{S}$. The matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}|}$ corresponds to the interaction matrix dictating the internal structure of the system, i.e., the per-capita effect of species j on an individual of species i. The effective parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{|\mathcal{S}|})^\intercal \in \mathbb{R}^{|\mathcal{S}|}$ consist of the phenomenological effect of the internal (e.g., intrinsic growth rate), abiotic (e.g., temperature, pH, nutrients), and biotic factors (e.g., unknown species) acting on the per-capita growth rate of each particular species. That is, we consider that the *effective growth rate* $\boldsymbol{\theta}$ represents the total additional effect of all the unknown potential factors as a function of the environmental conditions acting on each species independently. Note that pairwise effects a_{ij} from a given pool of species are assumed to be known. Our essential assumptions are thus that the effective growth rates $\boldsymbol{\theta}$ are environmentally determined (i.e., change with the environment), but a pairwise

interaction a_{ij} is a property only of the species pair (i.e., invariant under change of environment or addition of further species).

Under steady-state dynamics (i.e., $\boldsymbol{f} = 0$), the necessary (but not sufficient) condition for species coexistence is the existence of a feasible equilibrium (i.e., $\boldsymbol{N}^* = \mathbf{A}^{-1}\boldsymbol{\theta} > 0$) (22). We are interested here in globally stable systems; feasible and globally stable systems fulfill the necessary and sufficient conditions for coexistence (22). The combinations of the effective growth rates $\boldsymbol{\theta}$ compatible with the feasibility of a multispecies system $\mathcal{S}$ (known as the feasibility region D_F) can be described as:

$$D_F(\mathcal{S}) = \left\{ \sum_{j \in \mathcal{S}} N_j^* \mathbf{a}_j : \mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_{|\mathcal{S}|}], N_j^* \in \mathbb{R}_{>0} \right\} \in \mathbb{R}^{|\mathcal{S}|},$$

where $\mathbf{a}_j$ is the jth column vector of the interaction matrix $\mathbf{A}$. Additionally, assuming no a priori information about how the environmental conditions affect the effective growth rate (39), the feasibility region can be normalized by the volume of the entire parameter space (25) as

$$F(\mathcal{S}) = \frac{\text{vol}(D_F(\mathcal{S}) \cap \Theta^{|\mathcal{S}|})}{\text{vol}(\Theta^{|\mathcal{S}|})} \in (0, 0.5),$$

where $\Theta^{|\mathcal{S}|}$ is the $(|\mathcal{S}| - 1)$-dimensional closed unit sphere in dimension $|\mathcal{S}|$. Geometrically, the feasibility region is the convex hull of the $|\mathcal{S}|$ spanning vectors in $|\mathcal{S}|$-dimensional space, and thus always corresponds to an $|\mathcal{S}|$-dimensional solid angle that is contained within a single hemisphere (e.g., an acute angle in the case $|\mathcal{S}| = 2$). Thus, the feasibility that is normalized by the entire $|\mathcal{S}|$-dimensional space cannot be higher than 0.5. Ecologically, this normalized feasibility region can be interpreted as the probability of feasibility of a system $\mathcal{S}$ characterized by known pairwise interactions $\mathbf{A}$ under environmental uncertainty (all possible effective growth rates are equally likely to happen). The measure $F(\mathcal{S})$ can be efficiently estimated analytically (25, 40) or using Monte Carlo methods (Sec. S4). Moreover, $F(\mathcal{S})$ is robust to gLV dynamics with linear functional responses (25), a large family of nonlinear functional responses (39, 41), gLV stochastic dynamics (39, 41), and as a lower bound for complex polynomial models in species abundances (33).

$F(\mathcal{S})$ represents the feasibility of all $|\mathcal{S}|$ species together. That is, if a system (i.e., a given pool) is composed of 2 species or of 10 species, $F(\mathcal{S})$ represents the probability of feasibility of the full set of 2 or 10 species, respectively (25). In the case where $|\mathcal{S}| = 2$, $F(\mathcal{S})$ corresponds to the traditional indicator of pairwise coexistence in isolation, which is proportional to what is typically known as niche overlap in gLV competition systems (42, 43). It is known that $F(\mathcal{S})$ will tend to decrease as a function of the dimension of the system (23, 33): $F(\mathcal{S}_1) > F(\mathcal{S}_2)$ if $|\mathcal{S}_1| < |\mathcal{S}_2|$. Thus, focusing on a given pair of species $\mathcal{Z} = \{i, j\}$ within a system $\mathcal{Z} \subset \mathcal{S}$, $F(\mathcal{S})$ will likely underestimate the probability of feasibility for the pair, whereas using only the feasibility of the pair in isolation $F(\mathcal{Z})$ can be potentially misleading. For example, considering a pair that has a rather small feasibility region in isolation ($F(\mathcal{Z}) \sim 0$, e.g., almost identical niches) will prompt us to conclude that the coexistence

of this pair should be forbidden across all possible environmental conditions—as suggested by the checkerboard hypothesis (11, 14, 17). But, can this pair of species coexist more easily with additional third-party species (44, 45)? And will different pairs within the same system of multiple species be equally benefited (or affected)?

Looking at traditional indicators from a systems perspective

To understand the extent to which $F(\mathcal{Z})$ provides a reliable indicator of the possibility of coexistence of a pair (or community) $\mathcal{Z} = \{i, j\}$ within a larger multispecies system $\mathcal{S}$ characterized by a pairwise interaction matrix $\mathbf{A}$, we study geometrically how the feasibility conditions $\boldsymbol{\theta}$ change when a pair of species is in isolation and with third-party species. We approach this question from several different perspectives. First, we start by asking to what extent can third-party species act as ecosystem engineers and modify the environment into a more suitable habitat (26, 27)? In particular, for a given pair i, j within a fixed community containing third-party species, we consider the range of environmental conditions associated with the two parameters θ_i, θ_j such that it is *possible* for the pair i, j to exist for *some* set of values for the remaining θ_k's. This gives a sense of the potential environmental range of this pair in the context of a fixed community. In the following subsection we explore the more precise question of what the *probability* of coexistence is when we average over all possible values of all the θ_k's.

Mathematically, the projection region D_{Proj} of the feasibility region D_F of a multispecies system $\mathcal{S}$ onto the 2-dimensional parameter space of a pair of species $\mathcal{Z}$ can be represented by the conical hull, i.e., the set of all conical combinations of the projection of spanning rays:

$$D_{\text{Proj}}(\mathcal{Z}, \mathcal{S}) = \left\{ \sum_{j \in \mathcal{S}} N_j^* \mathbf{b}_j : \mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_{|\mathcal{S}|}], N_j^* \in \mathbb{R}_{>0} \right\} \in \mathbb{R}^2,$$

where $\mathbf{b}_j$ is the jth column vector of $\mathbf{B}$, and it contains a subset of elements corresponding to the pair $\mathcal{Z}$ in the column vector $\mathbf{a}_j$ of the interaction matrix $\mathbf{A}$. In other words, $\mathbf{B}$ consists of the row vectors $\mathbf{a}_i$ of $\mathbf{A}$ that correspond to the pair $\mathcal{Z}$ (i.e., $i \in \mathcal{Z}$). For example, for the pair of species 1 and 3 (i.e., $\mathcal{Z} = \{1, 3\}$), for each column vector $\mathbf{a}_j = (a_{1j}, \cdots, a_{|\mathcal{S}|j})^\intercal$, $j \in \{1, \cdots, |\mathcal{S}|\}$ in $\mathbf{A}$, we have $\mathbf{b}_j = (a_{1j}, a_{3j})^\intercal$ in $\mathbf{B}$. Then, we can obtain the normalized projection by the following ratio of volumes:

$$\text{Proj}(\mathcal{Z}, \mathcal{S}) = \frac{\text{vol}(D_{\text{Proj}}(\mathcal{Z}, \mathcal{S}) \cap \Theta^2)}{\text{vol}(\Theta^2)} \in [F(\mathcal{Z}), 0.5] \cup \{1\}.$$

Recall that $F(\mathcal{Z})$ is the feasibility of pair $\mathcal{Z}$ in isolation (using the corresponding 2-dimensional submatrix $\mathbf{A}[i, j; i, j]$). This projection region captures the first ecological question raised above: what is the range of θ_i, θ_j under which it is *possible* for species i, j to coexist for a given matrix $\mathbf{A}$ and some values of the remaining parameters $\theta_k, k \neq i, j$.

If the projection region is not the entire 2-dimensional parameter space $\boldsymbol{\theta}$ of the pair, the projection

ranges within Proj $\in (0, 0.5]$ (Fig. 1A provides an illustration). Otherwise, Proj $= 1$, indicating that the region (θ_1, θ_2) incompatible with the possible coexistence of the pair disappears. In other words, in these cases pairwise coexistence can be possible under any combination of (θ_i, θ_j). The realization of such coexistence, however, still depends on the other $\boldsymbol{\theta}$ values associated with the third-party species in the system. In fact, per definition $D_{\text{Proj}}(\mathcal{Z}, \mathcal{S}) \geq D_F(\mathcal{Z})$, the entire projection region $D_{\text{Proj}}(\mathcal{Z}, \mathcal{S})$ defines equal or weaker constraints on the plane $(\theta_i,\ \theta_j)$ for the coexistence of the pair. Thus, the amount of *projection contribution* increasing the feasibility region on the plane (θ_i, θ_j) can be defined as

$$PC(\mathcal{Z}, \mathcal{S}) = \frac{\text{Proj}(\mathcal{Z}, \mathcal{S})}{F(\mathcal{Z})}. \tag{1}$$

Because the combination of spanning vectors increases as the dimension of the system $|\mathcal{S}|$ grows, the projection $\text{Proj}(\mathcal{Z}, \mathcal{S})$ is expected to approach 1 in large multispecies systems (Figs. S3-S4).

Analytical system-level effects

While the projection contribution, which measures the *possibility* of pairwise coexistence, is always beneficial to the likelihood of pairwise coexistence, it is measured in the two-dimensional (θ_i, θ_j) plane of the pair $\{i, j\}$ in question. The actual probability of pairwise coexistence, however, depends on the $\boldsymbol{\theta}$ values of the rest of the species in the community. To find long-term analytical indicators of the effect of third-party species on pairwise coexistence, first we propose to do a geometric partition of the parameter space of $\boldsymbol{\theta}$ (a closed unit sphere) into non-invadable feasibility regions of all possible species compositions within a multispecies system. These regions are disjoint when there is a global stable equilibrium. Formally, in a multispecies system $\mathcal{S}$ with an interaction matrix $\mathbf{A}$ and a stable global equilibrium, the feasibility region of a composition $\mathcal{C}$ (a subset of species found in the system, i.e., $\mathcal{C} \subseteq \mathcal{S},\ \mathcal{C} \neq \emptyset$) can be defined as:

$$D_F(\mathcal{C}, \mathcal{S}) = \left\{ \sum_{j \in \mathcal{C}} N_j^* \mathbf{a}_j - \sum_{k \in \mathcal{S} \setminus \mathcal{C}} N_k^* \mathbf{e}_k : \mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_{|\mathcal{S}|}], N_{j,k}^* \in \mathbb{R}_{>0} \right\} \in \mathbb{R}^{|\mathcal{S}|},$$

where $\mathbf{e}_k$ is a unit vector whose k^{th} entry is 1 and 0 elsewhere. In fact, for a composition $\mathcal{C}$, we only need the subset of column vectors $\mathbf{a}_j$ of $\mathbf{A}$ corresponding to the composition $\mathcal{C}$ (i.e., $j \in \mathcal{C}$) instead of the whole matrix. For example, for the species composition $\mathcal{C} = \{1, 2, 3\}$, only the 1st, 2nd and 3rd column vectors of $\mathbf{A}$ (i.e., the sub-matrix $[\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3]$) are needed to define its feasibility region. Furthermore, we can define the normalized feasibility of a composition $\mathcal{C}$ by

$$F(\mathcal{C}, \mathcal{S}) = \frac{\text{vol}(D_F(\mathcal{C}, \mathcal{S}) \cap \Theta^{|\mathcal{S}|})}{\text{vol}(\Theta^{|\mathcal{S}|})} \in (0, F(\mathcal{C})].$$

$F(\mathcal{C})$ is a function of the sub-matrix formed only by the interactions between the species in $\mathcal{C}$ (i.e., $\mathbf{A}[\mathcal{C};\mathcal{C}]$); while $F(\mathcal{C},\mathcal{S})$ is a function of this sub-matrix augmented by the system-level effects of the composition on the third-party species in the system. In general, $F(\mathcal{C},\mathcal{S}) = F(\mathcal{C})$ only when $\mathcal{C} = \mathcal{S}$. Note that when the full community $\mathcal{S}$ admits a globally stable equilibrium, non-invadable equilibria of any subsystems $\mathcal{C}$ are also stable (46).

Following our geometric partition, we define the probability of feasibility of a pair of species $\mathcal{Z}$ in a larger multispecies system $\mathcal{S}$ as the sum of all the feasibility regions where the pair forms part of the species composition ($\mathcal{Z} \subseteq \mathcal{C} \subseteq \mathcal{S}$):

$$P(\mathcal{Z},\mathcal{S}) = \sum_{\mathcal{Z}\subseteq\mathcal{C}\subseteq\mathcal{S}} F(\mathcal{C},\mathcal{S}).$$

Because the proposed probability of feasibility is related to the solution of the system at equilibrium, we estimate analytically the *long-term effect* of a multispecies system $\mathcal{S}$ on a pair $\mathcal{Z}$ by

$$LE(\mathcal{Z},\mathcal{S}) = \frac{P(\mathcal{Z},\mathcal{S})}{F(\mathcal{Z})}, \tag{2}$$

which represents the extent to which third-party species can modify the probability of pairwise coexistence relative to the probability of the pair in isolation allowing infinite time (Fig. 1B). Note that the long-term effect given by Eq. (2) describes the effect of the larger multispecies system on the *probability* that the composition $\mathcal{C}$ arises as a community of the multispecies system, while the projection contribution given by Eq. (1) denotes the effect of the larger multispecies system in weakening constraints on the range of parameters allowing *possible* coexistence of composition $\mathcal{C}$. The actual probability of pairwise coexistence depends on averaging over the effective growth rates $\boldsymbol{\theta}$ of the rest of the species in the community. The long-term effect of Eq. (2) takes that into account and therefore, measures the relative probability of a pairwise equilibrium being feasible within a specific multispecies community. Note also that while the projection contribution always expands the possible range of coexistence for a given pair, the long-term effect that measures the average coexistence range is often (roughly half of all cases) decreased; these statements are completely consistent since the probability of the additional θ_k's taking values that significantly expand the pairwise coexistence range can be nonzero but quite small compared to the probability of contracting the coexistence range, and the projection contribution only incorporates the former.

Numerical system-level effects

While the long-term behavior of a multispecies system provides an expectation of the effect of third-party species on a pair of species under an ideal setting; in practice, the study of pairwise coexistence in natural or experimental settings depends on the time scale of investigation, extinction thresholds, and the finite number of replicates. This implies that transient dynamics may play an additional role

and observed coexistence rates may be different from the long-term expectations (12, 47). Thus, to provide a practical short-term perspective regarding the effect of third-party species on the probability of pairwise coexistence, we complement the analytical calculations of feasibility with numerical calculations. We simulate gLV dynamics over a given time interval across an arbitrary number of repetitions with arbitrary initial conditions (picking from a uniform distribution between 0 and 1 for each species, but starting with the same abundance for all species yields similar results) and classify species with a final abundance $N < 10^{-6}$ (results are robust to different thresholds) as statistically extinct (Fig. 1C). The time interval 200 is chosen for consistency with the experimental data in Sec. S1. That is, while correlations between short-term and experimental effects were in general good, correlations using a total time of 200 were the strongest (Figs. S10-S11), given the specific time parameters of the experiments. Hence, for consistency, we simply set all time intervals of simulations to 200. For multispecies systems, the vector of effective growth rates $\boldsymbol{\theta}$ is sampled randomly from the closed unit sphere. Then, to control for confounding factors, we take the two values in the sampled vector $\boldsymbol{\theta}$ that correspond to the pair as the effective growth rates for the pair in isolation. All simulations are conducted by the Runge-Kutta method.

We quantify the *short-term effect* of a multispecies system $\mathcal{S}$ on a pair of species $\mathcal{Z}$ by the ratio of the simulated frequencies of pairwise coexistence with third-party species and the simulated frequencies in isolation:

$$SE(\mathcal{Z},\mathcal{S}) = \frac{\mathrm{Sim}(\mathcal{Z},\mathcal{S})}{\mathrm{Sim}(\mathcal{Z})} = \frac{\#\ \mathcal{Z} \text{ in } \mathcal{S}}{\#\text{ simulations of } \mathcal{S}} \Big/ \frac{\#\ \mathcal{Z} \text{ in } \mathcal{Z}}{\#\text{ simulations of } \mathcal{Z}}. \tag{3}$$

The simulated coexistence frequencies are always greater than or equal to the analytical, i.e., there is a short-term buffering effect $\mathrm{Sim}(\mathcal{Z},\mathcal{S}) \geq P(\mathcal{Z},\mathcal{S})$ and at $\mathrm{Sim}(\mathcal{Z}) \geq F(\mathcal{Z})$, because we always start with all species present in the system and the global equilibrium is assumed to be stable. Only after a sufficiently long time do the simulated and analytical frequencies become the same (Fig. S7).

To study whether the equilibration between short-term and long-term effects happens more quickly with third-party species or in isolation, we quantify the *buffering effect* of a simulated system by the ratio between the short-term and long-term effects

$$BE(\mathcal{Z},\mathcal{S}) = \frac{SE(\mathcal{Z},\mathcal{S})}{LE(\mathcal{Z},\mathcal{S})} = \frac{\mathrm{Sim}(\mathcal{Z},\mathcal{S})}{P(\mathcal{Z},\mathcal{S})} \cdot \left(\frac{\mathrm{Sim}(\mathcal{Z})}{F(\mathcal{Z})}\right)^{-1}, \tag{4}$$

which represents the transient benefit (if > 1) or disadvantage (if < 1) to a pair of species after a finite time simulation of the system dynamics relative to the expected long-term behavior in the analytical calculations. For a given pair, the value of the buffering effect can change according to the chosen criteria for the simulations (Fig. S10), particularly the time span of the simulation. Yet, $BE(\mathcal{Z},\mathcal{S}) = 1$ if the two simulated frequencies (within systems $\mathrm{Sim}(\mathcal{Z},\mathcal{S})$ and in isolation $\mathrm{Sim}(\mathcal{Z})$) are greater than

their analytical counterparts (within systems $P(\mathcal{Z}, \mathcal{S})$ and in isolation $F(\mathcal{Z})$) in equal proportion. On the other hand, if $BE(\mathcal{Z}, \mathcal{S})$ is greater (resp. less) than 1, the buffering effect is stronger with third-party species (resp. in isolation).

Comparing analytical and numerical effects for a given pair

To illustrate how the long-term (analytical) and short-term (numerical) effects act on a given pair with different third-party species, Figure 1D shows the distribution of analytical and numerical system-level effects calculated on a hypothetical pair of symmetrically competing species $\mathcal{Z} = \{1, 2\}$ ($a_{12} = a_{21} = 0.22$) when assembled together with different third species. Here, the interactions associated with the third species are chosen randomly from a normal distribution with $\mu = 0$ and $\sigma = 1$ (all diagonal values are set to $a_{ii} = 1$). Note that for such multispecies systems there is generally a stable global equilibrium, and $\mathcal{F}(\mathcal{S}) > 0$. First, in this example, the projection contribution (first pink distribution) displays a large range of values including its maximum 1, confirming that the necessary conditions in (θ_1, θ_2) for pairwise coexistence in isolation are substantially expanded with the third-party species. Second, the long-term effects (second blue distribution) as well as the short-term effects (third orange distribution) show that the probability of pairwise coexistence with third-party species can either decrease or increase compared to the probability in isolation. Yet, both distributions are approximately centered at 1, especially for the analytical case (95% confidence intervals for the mean of analytical effects [0.956, 1.027]). In fact, Figure 1E shows that both effects are positively associated, but not perfectly correlated, as expected due to transient behavior. We have considered a wide variety of such systems and find that this property is similar for any pair of species under various numbers of randomly-assembled third-party species (Figs. S3-S4). This reveals that the probability of pairwise coexistence in isolation is very close to the expected probability across all possible (abiotic and biotic) conditions under both short-term and long-term behaviors. This result is confirmed and further illustrated by a perturbative analysis for weakly interacting systems in Sec. S5, where the center of the probability distribution of the analytical long-term effect varies from 1 only at fourth order in the interactions.

Lastly, for this particular pair of competing species, the buffering effects (fourth green distribution in Fig. 1D) are predominantly greater than 1, showing that the short-term effects tend to be stronger (relative to the long-term effects) with third-party species than in isolation. More generally, Figure 1F shows that negative (resp. positive) long-term effects of third-party species on pairwise coexistence are associated with a stronger (resp. weaker) buffering effect with third-party species than in isolation. In other words, short-term effects tend to be higher (resp. lower) than long-term effects if these long-term effects are less (resp. greater) than one. These results also hold for cases where the environmental effects are constrained to specific regions of the parameter space (Sec. S3, Fig. S6). All these results are also confirmed by the perturbative analysis for weakly interacting systems in Sec. S5.

Comparing analytical and numerical effects across pairs

Next, we study how the long-term (analytical) and short-term (numerical) effects act across pairs with a given set of third-party species. Specifically, we randomly generate a five-dimensional interaction matrix $\mathbf{A}$, where the interspecific effects (off-diagonal values) are sampled using a normal distribution with $\mu = 0$ and $\sigma = 0.25$ (all diagonal values are set to $a_{ii} = 1$), making sure that the interaction matrix is diagonally dominant (and therefore globally stable). These values and conditions are chosen for illustrative purposes in line with experimental data in Sec. S1, but results are robust to this choice (Figs. 2 and S9). Additionally, to study how patterns can change as a function of the level of competition with third-party species, we generate a different five-dimensional interaction matrix, where the elements are the same as the first matrix except that all the off-diagonal signs are inverted. We measure the level of competition by the mean interaction strength. (We have also analyzed the association between the fraction of negative interactions and coexistence probability, but we did not find any systematic relationship; the reason is that even few strong negative interactions can generate relative negative impacts on pairwise coexistence.) Then, for each pair within the randomly-generated matrices $\mathbf{A}$, we calculate the projection contribution $PC(\mathcal{Z}, \mathcal{S})$, long-term effects $LE(\mathcal{Z}, \mathcal{S})$, short-term effects $SE(\mathcal{Z}, \mathcal{S})$, and the buffering effects $BE(\mathcal{Z}, \mathcal{S})$. In order to calculate the simulated frequency of coexistence of each pair in isolation $\mathrm{Sim}(\mathcal{Z})$, we run simulations at the pair level using the corresponding two-dimensional sub-matrices $\mathbf{A}[i, j; i, j]$ and sub-vectors $\boldsymbol{\theta}[i, j]$.

Results

Figures 2A-2B and 2C-2D summarize our results for a pair of negative-dominated ($-\langle a_{ij} \rangle = -0.051$) and positive-dominated ($-\langle a_{ij} \rangle = 0.051$) interaction matrices, respectively. Ecologically, negative-dominated interaction matrices correspond to primarily competitive systems. Instead, positive-dominated interaction matrices correspond to primarily cooperative systems. First, we found that the projection contribution (first pink distribution) is highly heterogeneous across pairs. Recall that this is a function of the combination between the pairwise interactions in $\mathcal{Z}$ and the remaining interactions in the system $\mathcal{S}$, where this combination changes from pair to pair. Second, we found that the long-term effects (second blue distribution) and short-term effects (third orange distribution) of third-party species can be disadvantageous (< 1) or beneficial (> 1) for pairs. Importantly, these effects are not equal across all pairs within the systems. Note that these distributions are only expected to be centered close to 1 for large enough systems given that each point in these distributions corresponds to a different pair with a fixed set of third-party species (not a fixed pair with different third-party species as in Fig. 1D). Third, the buffering effects (fourth green distribution) show values predominantly greater than (resp. less than) one in Figure 2A (resp. Figure 2C), revealing that the short-term effects tend to be stronger (resp. weaker) with third-party species than in isolation under negative-dominated (resp. positive-dominated) interactions. These strong (resp. weak) buffering effects are generated by the third-party species that also cause negative (resp. positive) effects over the long term.

To differentiate how long-term and short-term behaviors operate on each pair within a given set of third-party species, we construct a cartographic representation (38) of the relationship between long-term and buffering effects. Figures 2B and 2D show that the majority of pairs tend to be negatively ($LE < 1$) and positively ($LE > 1$) affected over the long-term within negative-dominated and positive-dominated random matrices, respectively. As a mirror image, the majority of pairs tend to be positively ($BE > 1$) and negatively ($BE < 1$) affected over the short-term within negative-dominated and positive-dominated random matrices, respectively. That is, on average, negative (resp. positive) long-term effects of third-party species on pairwise coexistence are typically linked with a stronger (weaker) short-term buffering effects with third-party species than in isolation, in agreement with the general results found in the previous section for a fixed pair in a variety of environments. Note that this cartographic representation can change as a function of the time span of the simulation (see Fig. S10). These effects were not correlated with pairwise coexistence in isolation (the size of points is not associated with the value of either axis).

Experimental data

To test our theory in a more realistic setting, we use a publicly available data set of an *in vivo* gut microbial system. The data set (hereafter the fruit fly data set) was generated in Ref. (15) and it comprises a set of ten-day experimental trials of five interacting microbes commonly found in the fruit fly *Drosophila melanogaster* gut microbiota: *Lactobacillus plantarum* (Lp), *Lactobacillus brevis* (Lb), *Acetobacter pasteurianus* (Ap), *Acetobacter tropicalis* (At), and *Acetobacter orientalis* (Ao). This in vivo experimental study performed monoculture experiments for each species, co-cultures experiments for each pair, and poly-culture experiments for the quintet (Fig. 3 provides an illustration), as well as triplet and quartet experiments not relevant to our analysis here. Each experiment was replicated 48 times across different hosts and each species' density was measured at the end of the observation period, providing one single record of persistence per replication for each species in monocultures, co-cultures, and poly-cultures (see Sec. S1 for details).

Experimental system-level effects

To investigate the effects of third-party species (system-level effects) on pairwise coexistence using the *in vivo* data, for each pair $\mathcal{Z}$, we compute the observational frequency of coexistence with third-party species $\mathcal{S}$ (i.e., within quintets) and in isolation $\mathcal{Z}$ (i.e., within pairs) separately. To determine species coexistence, we classify a species statistically extinct in a trial if its relative abundance was less than 1% at the end of the observation period. We also test different extinction thresholds and obtain qualitatively the same results (Fig. S12). We define the *experimental effect* of third-party species on pairwise coexistence as the ratio between the frequency of coexistence with third-party species and the frequency in isolation (Fig. 3 provides an illustration):

$$EE(\mathcal{Z},\mathcal{S}) = \frac{\#\ \mathcal{Z} \text{ in } \mathcal{S}}{\#\text{ replicates of } \mathcal{S}} \Big/ \frac{\#\ \mathcal{Z} \text{ in } \mathcal{Z}}{\#\text{ replicates of } \mathcal{Z}}. \tag{5}$$

To compare the experimental and long-term (analytical) effects, we use the experimentally-inferred five-dimensional pairwise matrix **A** (see Sec. S2 for more details). This matrix was characterized by negative-dominant interactions ($-\langle a_{ij}\rangle = -0.078$). Thus this experimental system is primarily competitive. Using this experimentally-inferred matrix, we analytically calculate the projected contributions $PC(\mathcal{Z},\mathcal{S})$ and long-term effects $LE(\mathcal{Z},\mathcal{S})$ of third-party species with $|\mathcal{S}| = 5$. We chose the quintet experiment (instead of the triplets or quartets) in order to have more pairs (10 with fives species) to perform statistics. In line with our numerical analysis, the *experimental buffering effect* $EBE(\mathcal{Z},\mathcal{S})$ is calculated by the ratio between the experimental effect $EE(\mathcal{Z},\mathcal{S})$ and long-term effect $LE(\mathcal{Z},\mathcal{S})$. We also construct a cartographic representation of the relationship between long-term effects and the experimental buffering effect in order to differentiate how long-term (analytical) and short-term (experimental) behaviors operate on each pair within the experimental system.

Comparing analytical and experimental effects across pairs

We found that the experimental effects of third-party species on pairwise coexistence were very similar to those generated by simulations (Figs. 2, 4 and S11). Specifically, Figure 4A shows that the experimental multispecies system generated heterogeneous effects across pairs as illustrated by a wide distribution of projection contributions (first pink distributions), long-term effects (second blue distributions), and experimental effects (third orange distributions). Similar to simulations for systems with negative-dominant interactions, the *in vivo* experimental system shows predominately positive values of buffering effects (fourth green distributions). This shows that despite the negative long-term effects of the third-party species on pairwise coexistence, there is a stronger buffering effect with third-party species than in isolation. Focusing on the cartographic representation of differences among pairs, Figure 4B shows that the majority of pairs tend to be negatively affected over the long-term ($LE < 1$) and positively affected over the short-term ($EBE > 1$)—consistent with the results obtained for negative-dominant matrices (Fig. 2B). This cartographic representation can also help to understand specific pairwise cases. For example, the pair (Ap, At) was observed with negative long-term effects but with a strongly positive buffering effect over the short-term (top left region). In contrast, the pair (Lp, Ap) exhibited the complete opposite effects. In turn, the pair (Ap, Ao) exhibited positive effects over both the long-term and short-term, suggesting that this pair is substantially benefited by the third-party species. This classification can change as a function of the experimental details. Importantly, these effects were again uncorrelated with pairwise coexistence in isolation (the size of points in Fig. 4B is not associated with the value of either axis), confirming the importance of third-party-species dynamics on pairwise coexistence.

Discussion

A major challenge in ecological research has been understanding how species interactions and diverse environments affect the opportunities for species coexistence (2, 4). Many research efforts have been devoted to studying the coexistence of either two species in isolation or that of a particular set of multiple species (8). Yet, species are seldom in isolation and the coexistence of a specific combination of multiple species is expected to be rare in a random environment (5, 24, 25). This has prompted us to better understand the extent to which the coexistence of a particular pair or presumed community of species can change between being in isolation to being part of a larger multispecies system. The main problem resides in the fact that in natural settings, most of the abiotic and biotic effects acting on species are unknown. As a response to this problem, we have introduced an analytical formula to estimate the extent to which third-party species can affect the probability that a particular pair of species coexist under environmental uncertainty (all possible conditions are equally likely to happen). This formula is based on information about pairwise interactions from a given pool of species and assumes that the total effect of all additional biotic and abiotic factors can be captured by a phenomenological random parameter for each species (i.e., the effective growth rate parameter θ). This kind of question is also a critical focus and challenge in bio-restoration and bio-medicine, where it is unclear whether a large system can have different effects on its different pairs of species (6, 7, 48–50). In restoration, for example, it may be that in a given ecological context where the species composition has been drastically modified or reduced, a given pair of species that could coexist in the original full ecological community may no longer have a feasible equilibrium in the modified species composition. In such a situation, better understanding the kind of analytic effects described here could help inform restoration ecologists regarding which other species must simultaneously be reincorporated to enable a specific pair to flourish.

Following our analytical framework, we have shown that the diversity of multispecies systems opens the opportunity for pairwise coexistence regardless of whether a given pair can coexist or not in isolation. This phenomenon occurs since the environmental conditions limiting the coexistence of a pair in isolation (parameter values acting on a plane) are replaced by a new set of conditions acting on all species, whose projection onto the two-dimensional space tends to cover the pair's entire plane. Thus, the possibility of pairwise coexistence with third-party species only depends on the new set of conditions in the multidimensional space of the entire system. The magnitude of projection reflects the range of compatible environments for the pair to coexist. This can be interpreted as the extent to which third-party species can act as ecosystem engineers (26, 27) and modify the environment into more suitable habitats. Yet, the probability of pairwise coexistence within larger systems still depends on the effective growth rates associated with third-party species. The long-term effects measure the difference in the probability of pairwise coexistence with and without third-party species. This can be interpreted as the relative benefit (or harm) of third-party species on pairwise coexistence. Importantly, under environmental uncertainty, the expected value of the distribution of probabilities

generated by all possible multispecies systems is approximately the same as the probability in isolation. That is, we can usefully approximate the likelihood of two species coexisting in diverse assemblages (or diverse environmental conditions) based on the understanding of their pairwise interactions. We have shown analytically that the deviation of the mean of the distribution from the isolated probability scales as a^4 for simple model systems with interactions of order a, and confirmed numerically that this effect is still very small for larger systems outside the analytic framework. This reasoning can be applied to any subset of species embedded in large multispecies systems.

While analytical measures are useful for increasing our general understanding of pairwise coexistence, the assumptions behind these measures may never be met in real-world systems. For example, our analytical measures are based on the long-term asymptotic equilibria achieved by systems in arbitrary environmental conditions. Yet, both simulations and experiments are always restricted to or biased by short-term (finite) effects. Hence, to provide an additional understanding of such short-term effects, we have compared our analytical expectations against numerical and experimental outcomes. As expected, we confirmed that the probability of pairwise coexistence over the short term (numerically and experimentally) is greater than pairwise coexistence over the long term (12). Yet, it has been unclear whether this buffering effect should be equal with third-party species and in isolation. We have found that this increase is typically greater (resp. smaller) for pairs with third-party species than in isolation if the third-party species affect the pairs negatively (resp. positively) in the long term. In other words, we have found that when third-party species tend to reduce (resp. increase) the coexistence probability of a pair, they tend to exhibit slower (resp. faster) rates of competitive exclusion. This illustrates potential differences between what we may see in experiments and the expected long-term dynamics of ecological systems. Moreover, these effects are not equal across all pairs within a system. Importantly, we have shown that we can study these differences using a cartographic representation of how third-party species are expected to affect the long-term and short-term behavior of each pair.

As a final note, it is worth mentioning that our theory assumes that environmental conditions are uniformly distributed across the entire parameter space. Instead, under experimental settings, much stronger constraints on environmental conditions can be expected. This implies that the probability of pairwise coexistence should be different between theoretical and experimental analyses. Yet, these experimental constraints should also be expected to operate on pairs with third-party species and in isolation. Therefore, system-level effects (ratio between the probability in multispecies systems and in isolation) should be comparable between theoretical and experimental analyses. Indeed, our theoretical analysis (Fig. 2) showed similar patterns to the experimental analysis (Fig. 4), suggesting that our phenomenological framework is applicable to real-world settings and can be used to increase our understanding of long-term and short-term ecological dynamics.

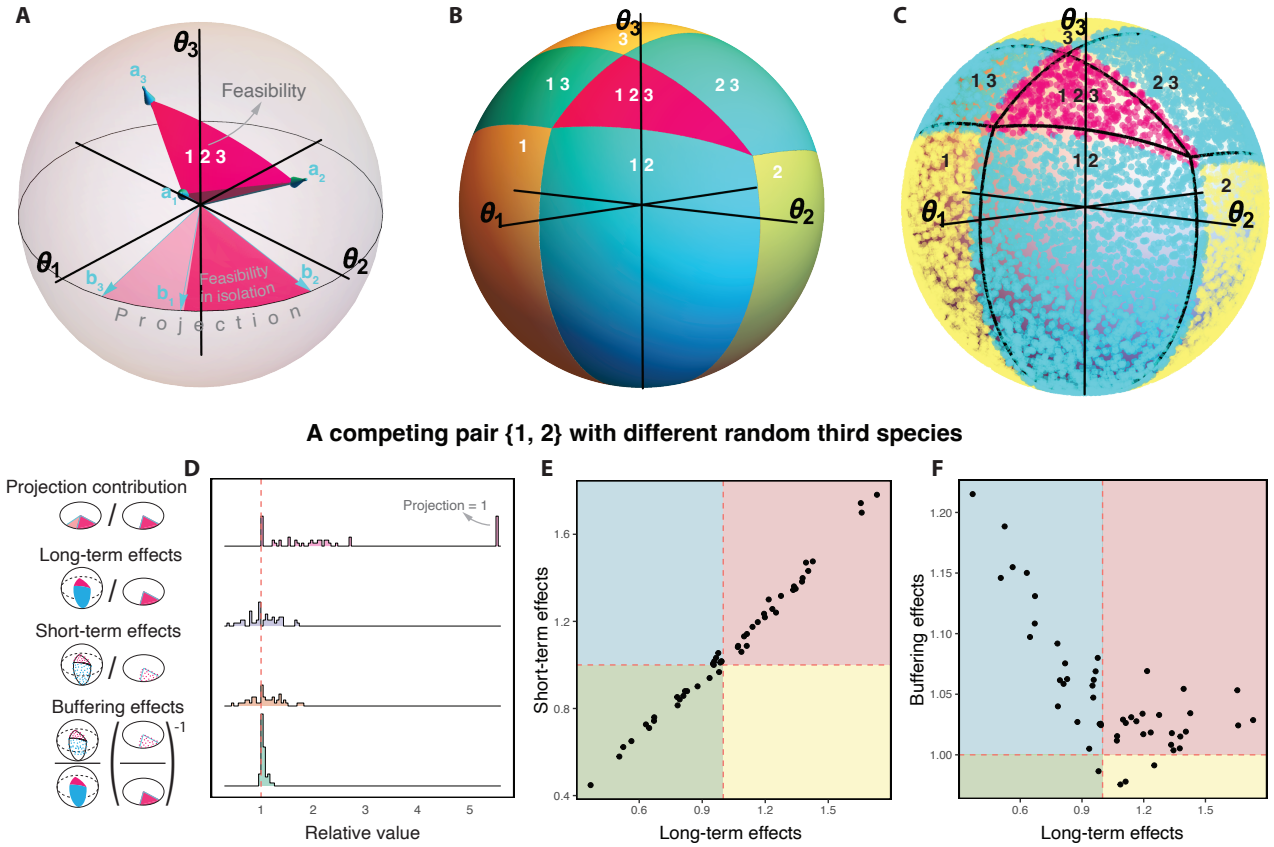

Figure 1: **Geometric perspective of the effect of third-party species on pairwise coexistence.** For a fictitious system of three species, Panel **A** shows the geometrical representation of the feasibility region $D_F(\mathcal{S})$ (i.e., pink cone spanned by $\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3$) of an illustrative three-species system $\mathcal{S}$ characterized by the interaction matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$ in the three-dimensional space of effective growth rates $(\theta_1,\ \theta_2,\ \theta_3)$. The panel also shows the projection $D_{\text{Proj}}(\mathcal{Z}, \mathcal{S})$ of the spanning rays onto the two-dimensional plane $(\theta_1,\ \theta_2)$ of the pair $\mathcal{Z} = \{1, 2\}$ (i.e., light pink and dark pink disk spanned by $\mathbf{b}_2,\ \mathbf{b}_3$). The feasibility of pair $\{1, 2\}$ in isolation (i.e., a 2-species pool) $D_F(\mathcal{Z})$ is given by $\mathbf{b}_1,\ \mathbf{b}_2$ (i.e., dark pink disk), and becomes a fraction of the projection corresponding to the possible coexistence of species 1 and 2 only. Panel **B** illustrates the geometric partition of the analytical feasibility regions of all possible compositions in the previous three-species system. For the same system, the points in Panel **C** show the resulting species composition (species with final abundances $N < 10^{-6}$ are considered statistically extinct) from 10,000 simulations over a short finite time interval (total time = 200, step size = 0.01) conducted by the Runge-Kutta method. Note that these results are a function of the time span of simulations (see Fig. S7), although the qualitative effects are similar for other choices of time span. The black lines correspond to the borders of the analytical feasibility regions in Panel **B**. Panel **D** shows the distribution of system-level effects calculated on a fixed hypothetical pair of symmetrically competing species $\mathcal{Z} = \{1, 2\}$ when assembled with different random third species. The distributions are generated using 50 random third species (see text for details). Each point in the distributions corresponds to the same pair with a different set of third-party species. For reference, the dashed line shows the value of one: the relative feasibility in isolation. Panels **E-F** show the positive and negative association of the long-term effect with the short-term and buffering effects, respectively (Panel **E**: Pearson's product-moment correlation = 0.995, p-value < 0.001; Panel **F**: Pearson's product-moment correlation = -0.707, p-value < 0.001). These results are complemented by a perturbative analysis for weakly interacting systems in Sec. S5.

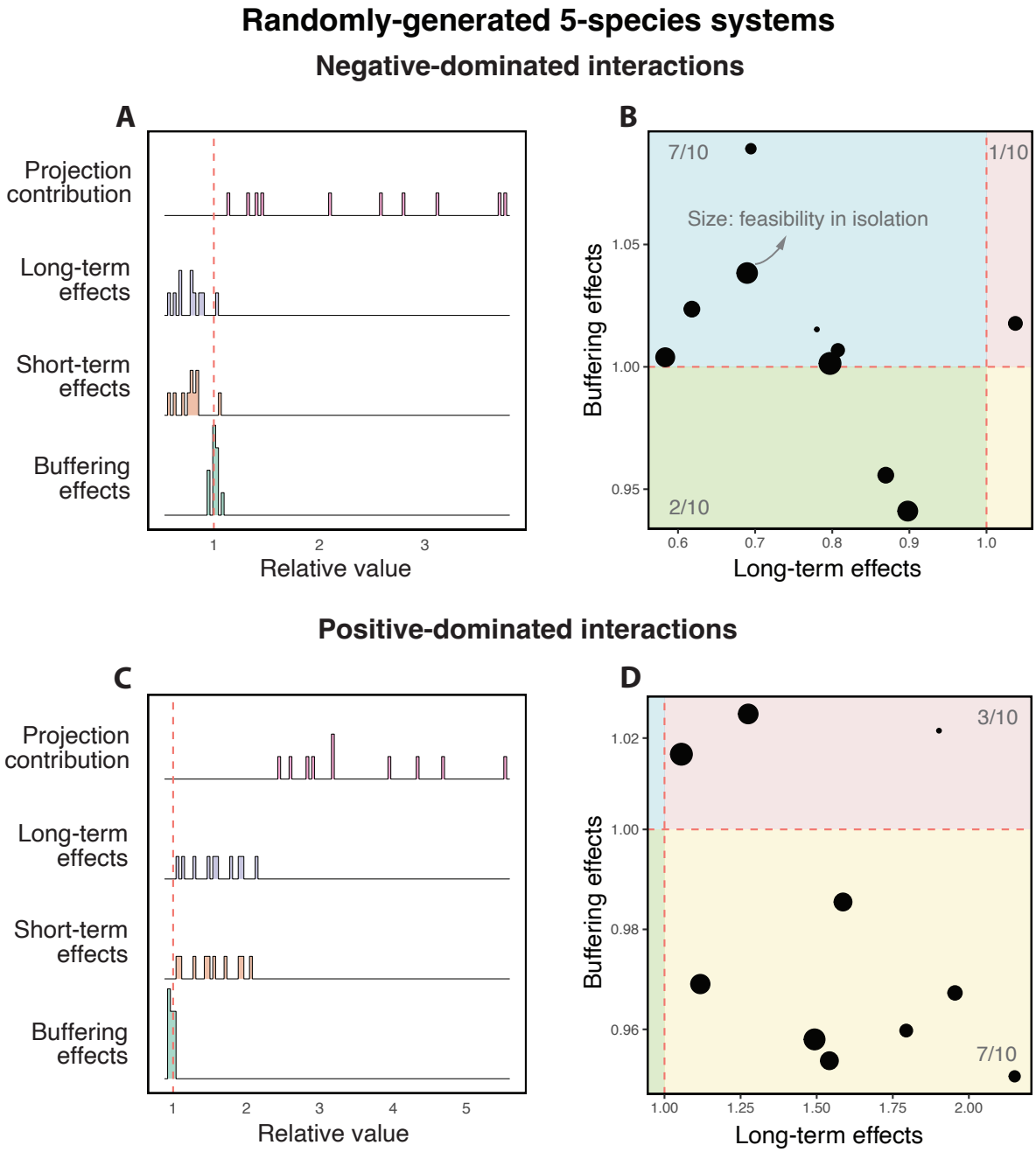

Figure 2: **Cartographic representation of long-term and short-term differences among pairs with the same set of third-party species.** Panels **A** and **C** show the system-level effects on 10 pairs within a randomly-generated five-species system characterized by negative-dominated and positive-dominated interactions, respectively (see text). Rows correspond to the projection distribution ($PC(\mathcal{Z},\mathcal{S})$), long-term effects ($LE(\mathcal{Z},\mathcal{S})$), short-term effects ($SE(\mathcal{Z},\mathcal{S})$), and buffering effects ($BE(\mathcal{Z},\mathcal{S})$), respectively. Note that each point in the distributions is a different pair within a set of third-party species. For reference, the dashed line shows the value of 1. Recall that the x-axis corresponds to the change in probability of pairwise coexistence with third-party species. Panels **B** and **D** illustrate the cartographic representation for all corresponding pairs based on beneficial ($LE(\mathcal{Z},\mathcal{S}) > 1$) or detrimental ($LE(\mathcal{Z},\mathcal{S}) < 1$) long-term effects and beneficial ($BE(\mathcal{Z},\mathcal{S}) > 1$) or detrimental ($BE(\mathcal{Z},\mathcal{S}) < 1$) short-term buffering effects. The size of points corresponds to the feasibility of pairs in isolation. The number of points in each region is annotated in gray. See Fig. S9 for a ten-species system.

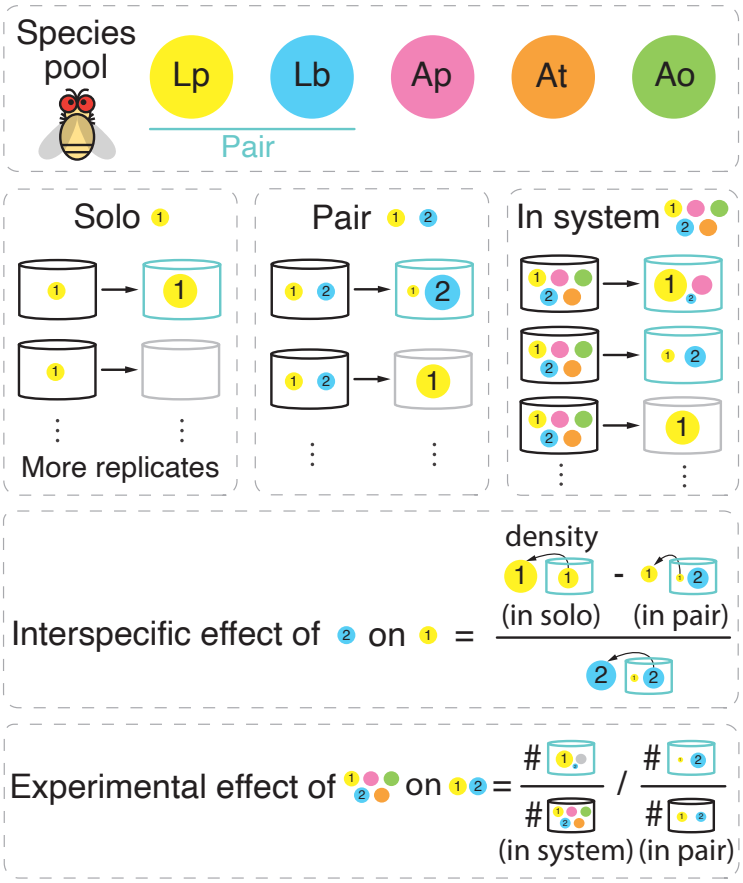

Figure 3: **Illustration of experimental data, inference of pairwise interactions, and calculation of experimental effects.** We use publicly available data of an *in vivo* gut microbial system. The data set was generated in Ref. (15) and it comprises ten-day experimental trials of five interacting microbes commonly found in the fruit fly *Drosophila melanogaster* gut microbiota: *Lactobacillus plantarum* (Lp), *Lactobacillus brevis* (Lb), *Acetobacter pasteurianus* (Ap), *Acetobacter tropicalis* (At), and *Acetobacter orientalis* (Ao). From the available experimental data, we used the results from monoculture experiments for each species, co-cultures experiments for each pair, and poly-culture experiments for the quintet. Each experiment was replicated across different hosts and each species' density was measured at the end of the observation period, providing a single record of persistence for each species in monocultures, co-cultures, and poly-cultures per replication. The experimentally-inferred pairwise matrix $\mathbf{A}$ was inferred using the data of monoculture experiments for each species (e.g., species 1) and co-cultures experiments for pairs (e.g., pair $\{1, 2\}$) by evaluating the modified abundance of species 1 with and without species 2 at the end of the observation periods (see Sec. S3). We define the experimental effects of the third-party species on pair $\{1, 2\}$ as the difference in the experimental frequencies of occurrence of the pair across all replicates with third-party species and the experimental frequencies of occurrence of the pair in isolation.

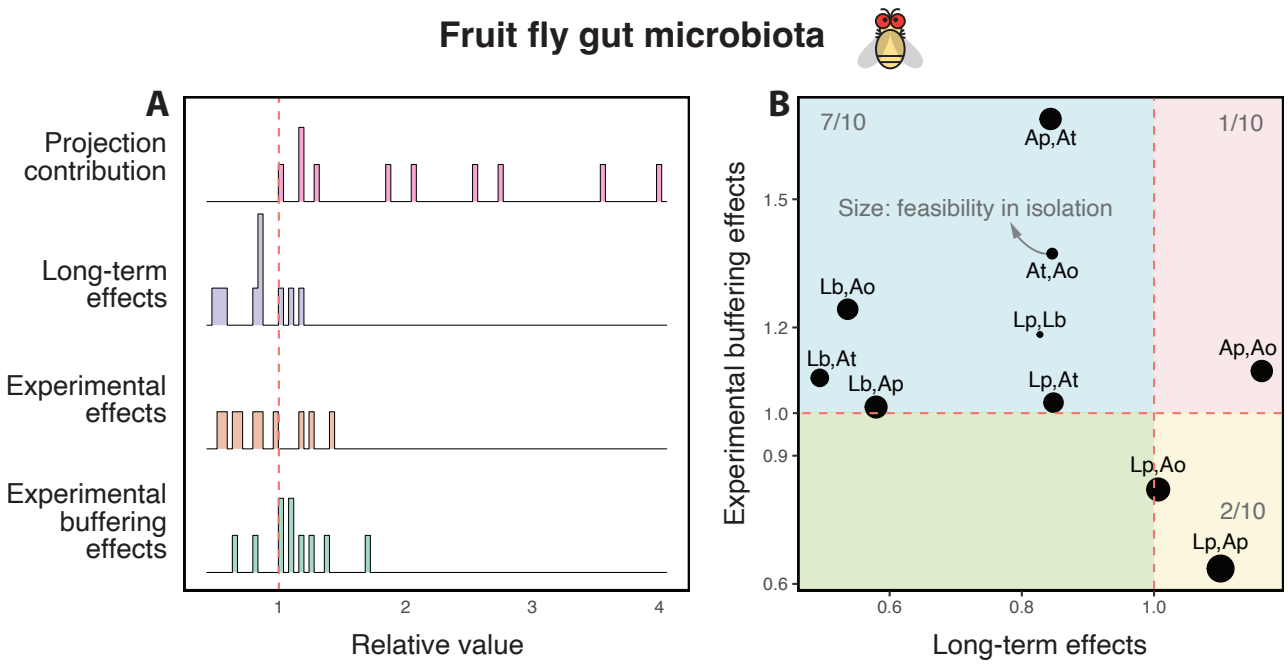

Figure 4: **Comparing system-level effects among pairs within a five-species *in vivo* experimental system.** This figure shows similar information to Fig. 2, but using the experimental data and the experimentally-inferred interaction matrix. For each of the 10 pairs in the experimental fruit fly system, Panel **A** shows the projection distribution $PC(\mathcal{Z}, \mathcal{S})$ (first row pink distribution), long-term effects $LE(\mathcal{Z}, \mathcal{S})$ (second row blue distribution), experimental effects $EE(\mathcal{Z}, \mathcal{S})$ (third row orange distribution), and experimental buffering effects $EBE(\mathcal{Z}, \mathcal{S})$ (fourth row green distribution). Panel **B** shows the cartographic representation of how third-party species affect the long-term and short-term behavior of each of its constituent pairs. Long-term effects can be beneficial ($LE(\mathcal{Z}, \mathcal{S}) > 1$) or detrimental ($LE(\mathcal{Z}, \mathcal{S}) < 1$). In this experimental system, the majority of short-term effects tend to be beneficial ($EBE(\mathcal{Z}, \mathcal{S}) < 1$) in line with negative-dominated interaction matrices (Fig. 2B). See text or Fig. 3 for all species names.

References

[1] Gould SJ (1989) *Wonderful Life* (W. W. Norton & Co, New York).

[2] Odum EP, Barrett GW (2005) *Fundamentals of Ecology* (Thomson, Brooks-Cole, Fifth Edition).

[3] Abrams PA (1987) On classifying interactions between populations. *Oecologia* 73:272–281.

[4] Vellend M (2016) *The theory of ecological communities* (Princeton University Press, NJ).

[5] Weiher E, Keddy P (2001) *Ecological assembly rules: perspectives, advances, retreats* (Cambridge University Press).

[6] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* 489:220–230.

[7] Faust K, Raes J (2012) Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10:538–550.

[8] Case TJ (2000) *An Illustrated Guide to Theoretical Ecology* (Oxford Univ. Press, Oxford).

[9] Gause GF (1932) Experimental studies on the struggle for existence. *Journal of Experimental Biology* 9:389–402.

[10] Park T (1954) Experimental studies of interspecies competition ii. temperature, humidity, and competition in two species of tribolium. *Physiological Zoology* 27:177–238.

[11] Diamond JM (1975) Assembly of species communities. *Ecology and evolution of communities* 342:444.

[12] Gilpin ME, Carpenter MP, Pomerantz MJ (1986) *The assembly of a laboratory community: multispecies competition in Drosophila*, eds Diamond JM, Case TJ (Harper and Row, New York), pp 23–40.

[13] Vandermeer JH (1969) The competitive structure of communities: An experimental approach with protozoa. *Ecology* 50:362–371.

[14] Friedman J, Higgins LM, Gore J (2017) Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* 1:0109.

[15] Gould AL, et al. (2018) Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences* 115:E11951–E11960.

[16] Sundarraman D, et al. (2020) Higher-order interactions dampen pairwise competition in the zebrafish gut microbiome. *mBio* 11:e01667–20.

[17] Gotelli NJ, McCabe DJ (2002) Species co-ocurrence: a meta-analysis of j. m. diamond's assembly rules. *Ecology* 83:2091–2096.

[18] Jones EW, Carlson JM, Sivak DA, Ludington WB (2022) Stochastic microbiome assembly depends on context. *Proc. nat. Acad. Sci. USA* 119:e2115877119.

[19] Tabi A, et al. (2020) Species multidimensional effects explain idiosyncratic responses of communities to environmental change. *Nature Ecology & Evolution* 4:1036–1043.

[20] Deng J, Angulo MT, Saavedra S (2021) Generalizing game-changing species across microbial communities. *ISME Communications* 1:1–8.

[21] Schreiber SJ, Yamamichi M, Strauss SY (2019) When rarity has costs: coexistence under positive frequency-dependence and environmental stochasticity. *Ecology* p e02664.

[22] Hofbauer J, Sigmund K (1998) *Evolutionary games and population dynamics* (Cambridge University Press).

[23] AlAdwani M, Saavedra S (2022) Feasibility conditions of ecological models: Unfolding links between model parameters. *Ecological Modelling* 466:109900.

[24] May RM (1972) Will a large complex system be stable? *Nature* 238:413–414.

[25] Song C, Rohr RP, Saavedra S (2018) A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *J. of Theoretical Biology* 450:30–36.

[26] Jones CG, Lawton JH, Shachak M (1994) Organisms as ecosystem engineers. *Oikos* 69:373–386.

[27] Fukami T (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics* 46.

[28] Schmitz OJ, Buchkowski RW, Burghardt KT, Donihue CM (2015) Functional traits and trait-mediated interactions: connecting community-level interactions with ecosystem functioning. *Advances in Ecological Research* 52:319–344.

[29] Peacor SD, Werner EE (2001) The contribution of trait-mediated indirect effects to the net effects of a predator. *Proc. Nat. Acad. Sci.* 98:3904–3908.

[30] Wooton JT (1994) Predicting direct and indirect effects: an integrated approach using experiments and path analysis. *Ecology* 75:151–165.

[31] Wilson DS (1992) Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. *Ecology* 73:1984–2000.

[32] Angulo MT, Moreno JA, Lippner G, Barabási AL, Liu YY (2017) Fundamental limitations of network reconstruction from temporal data. *Journal of the Royal Society Interface* 14:20160966.

[33] AlAdwani M, Saavedra S (2020) Ecological models: higher complexity in, higher feasibility out. *J. of the Roy. Soc. Interface* 17:20200607.

[34] Bender EA, Case TJ, Gilpin ME (1984) Perturbation experiments in community ecology: Theory and practice. *Ecology* 65:1–13.

[35] Pearl J (2009) *Causality* (Cambridge Univ. Press, Cambridge).

[36] AlAdwani M, Saavedra S (2019) Is the addition of higher-order interactions in ecological models increasing the understanding of ecological dynamics? *Mathematical Biosciences* 315:108222.

[37] Cao HT, Gibson TE, Bashan A, Liu YY (2016) Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. *Bioessays* 39:1600188.

[38] Guimerà R, Amaral LAN (2005) Functional cartography of complex metabolic networks. *Nature* 433:895–900.

[39] Saavedra S, Medeiros LP, AlAdwani M (2020) Structural forecasting of species persistence under changing environments. *Ecology Letters* 23:1511–1521.

[40] Saavedra S, Rohr RP, Olesen JM, Bascompte J (2016) Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and Evolution* 6:997–1007.

[41] Cenci S, Saavedra S (2018) Structural stability of nonlinear population dynamics. *Phys. Rev. E* 97:012401.

[42] Godoy O, Kraft NJ, Levine JM (2014) Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.* 17:836–844.

[43] Barabás G, D'Andrea R, Stump SM (2018) Chesson's coexistence theory. *Ecological Monographs* 88:277–303.

[44] Case TJ (1990) Invasion resistance arises in strongly interacting species-rich model competition communities. *Proc. Natl. Acad. Sci. U.S.A.* 87:9610–9614.

[45] Saavedra S, et al. (2017) A structural approach for understanding multispecies coexistence. *Ecological Monographs* 87:470–486.

[46] Takeuchi Y (1996) *Global dynamical properties of Lotka-Volterra systems* (World Scientific).

[47] Hastings A (2004) Transients: the key to long-term ecological understanding? *TREE* 19:39–45.

[48] Smith MB, Kelly C, Alm EJ (2014) How to regulate faecal transplants. *Nature* 506:290–291.

[49] Moreno-Mateos D, et al. (2017) Anthropogenic ecosystem disturbance and the recovery debt. *Nature Comm.* 8:14163.

[50] Rohr JR, et al. (2020) Towards common ground in the biodiversity – disease debate. *Nature Ecology & Evolution* 4:24–33.

[51] Xiao Y, et al. (2017) Mapping the ecological networks of microbial communities. *Nature communications* 8:1–12.