PNAS

Supporting Information for

A theory of ecological invasions and its implications for eco-evolutionary dynamics

Zhijie Feng, Emmy Blumenthal, Pankaj Mehta, Akshit Goyal

Akshit Goyal
E-mail: akshitg@icts.res.in

This PDF file includes:

Supporting Information Text

1. Derivation of linear response theory for the Generalized Lotka-Volterra (GLV) models

In this section, we provide a detailed derivation of Eq. (6) in main text which relates community shift to invader abundance and the pre-invasion abundance of extinct species. Then we derive Eq. (5) in main text, which predicts invader abundance given the knowledge of pre-invasion community, invader-community interaction, and the knowledge of what survives after invasion. Finally, in the last subsection we go beyond the scope of the main text and extend the framework to allow multiple invaders and environmental perturbations.

A. Primer: Environmental perturbations without extinctions. A simple primer to consider before considering invasion and extinction is when environmental perturbations alter species abundance without causing extinctions. The steady state before invasion X_j^{old} is given by linear equality $r_i - \sum_j A_{ij} X_j^{\text{old}} = 0$, solving the linear equation leads to

$$X_i^{\text{old}} = A_{ij}^{-1} r_j. \tag{1}$$

Now suppose there is an environmental perturbation $r_i \to r_i + \delta r_i$, the new steady state condition is simply $r_i + \delta r_i - \sum_j A_{ij} X_j^{\text{old}} = 0$, leading to new abundance

$$X_i^{\text{new}} = A_{ij}^{-1}(r_j + \delta r_j). \tag{2}$$

Therefore, the change in steady state abundance is simply

$$\delta X_i^{\text{no extinction}} = A_{ij}^{-1} \delta r_j. \tag{3}$$

Here this relationship is exact instead of perturbative; the steady-state condition for surviving species in the Lotka-Volterra model is exactly linear.

B. Derivation of Eq. (6) in main text: Invasions and extinctions as effective perturbations. In this work, instead of actual environmental perturbation, we make use of the insight that in a background of large community, a species feels the addition of a new interaction or a deletion of an old interaction no differently than an environmental perturbation. More explicitly, solving Eq. (3) in main text leads to

$$r_{\mathcal{S}} - A_{\mathcal{S}\mathcal{S}} \vec{X}_{\mathcal{S}}^{\text{new}} - A_{\mathcal{S}0} X_0 = 0, \tag{4}$$

$$\vec{X}_{\mathcal{S}}^{\text{new}} = A_{\mathcal{S}\mathcal{S}}^{-1}(r_{\mathcal{S}} - A_{\mathcal{S}0} X_0). \tag{5}$$

At the same time, the old steady state condition for these surviving species is

$$r_{\mathcal{S}} - A_{\mathcal{S}\mathcal{S}} \vec{X}_{\mathcal{S}}^{\text{old}} - A_{\mathcal{S}\mathcal{E}} \vec{X}_{\mathcal{E}} = 0, \tag{6}$$

$$\vec{X}_{\mathcal{S}}^{\text{old}} = A_{\mathcal{S}\mathcal{S}}^{-1}(r_{\mathcal{S}} - A_{\mathcal{S}\mathcal{E}} \vec{X}_{\mathcal{E}}). \tag{7}$$

Therefore, their change in abundance is

$$\delta \vec{X}_{\mathcal{S}} = \vec{X}_{\mathcal{S}}^{\text{new}} - \vec{X}_{\mathcal{S}}^{\text{old}}, \tag{8}$$

$$= A_{\mathcal{S}\mathcal{S}}^{-1}(-A_{\mathcal{S}0} X_0 + A_{\mathcal{S}\mathcal{E}} \vec{X}_{\mathcal{E}}). \tag{9}$$

This recovers Eq. (6) in main text. Moreover, comparison with Eq. (3) shows the role of effective environmental perturbation played by invader and extinct species influence, $\vec{r}^{\text{eff}} = -A_{\mathcal{S}0} X_0 + A_{\mathcal{S}\mathcal{E}} \vec{X}_{\mathcal{E}}$.

C. Derivation of Eq. (5) in main text: Predicting invader abundance. The invader introduces an extra constraint on the steady state condition, namely

$$r_0 - A_{0\mathcal{S}} \vec{X}_{\mathcal{S}}^{\text{new}} - A_{00} X_0 = 0 \tag{10}$$

$$r_0 - A_{0\mathcal{S}} \vec{X}_{\mathcal{S}}^{\text{old}} - A_{0\mathcal{S}} \delta \vec{X}_{\mathcal{S}} - A_{00} X_0 = 0 \tag{11}$$

We can then substitute Eq. (6) in main text to obtain

$$r_0 - A_{0\mathcal{S}} \vec{X}_{\mathcal{S}}^{\text{old}} - A_{0\mathcal{S}} A_{\mathcal{S}\mathcal{S}}^{-1}(-A_{\mathcal{S}0} X_0 + A_{\mathcal{S}\mathcal{E}} \vec{X}_{\mathcal{E}}) - A_{00} X_0 = 0 \tag{12}$$

Solving for X_0 arrives at

$$X_0 = \frac{r_0 - A_{0\mathcal{S}} \vec{X}_{\mathcal{S}}^{\text{old}} - A_{0\mathcal{S}} A_{\mathcal{S}\mathcal{S}}^{-1} A_{\mathcal{S}\mathcal{E}} \vec{X}_{\mathcal{E}}}{A_{00} - A_{0\mathcal{S}} A_{\mathcal{S}\mathcal{S}}^{-1} A_{\mathcal{S}0}}, \tag{13}$$

which recovers Eq. (5) in main text.

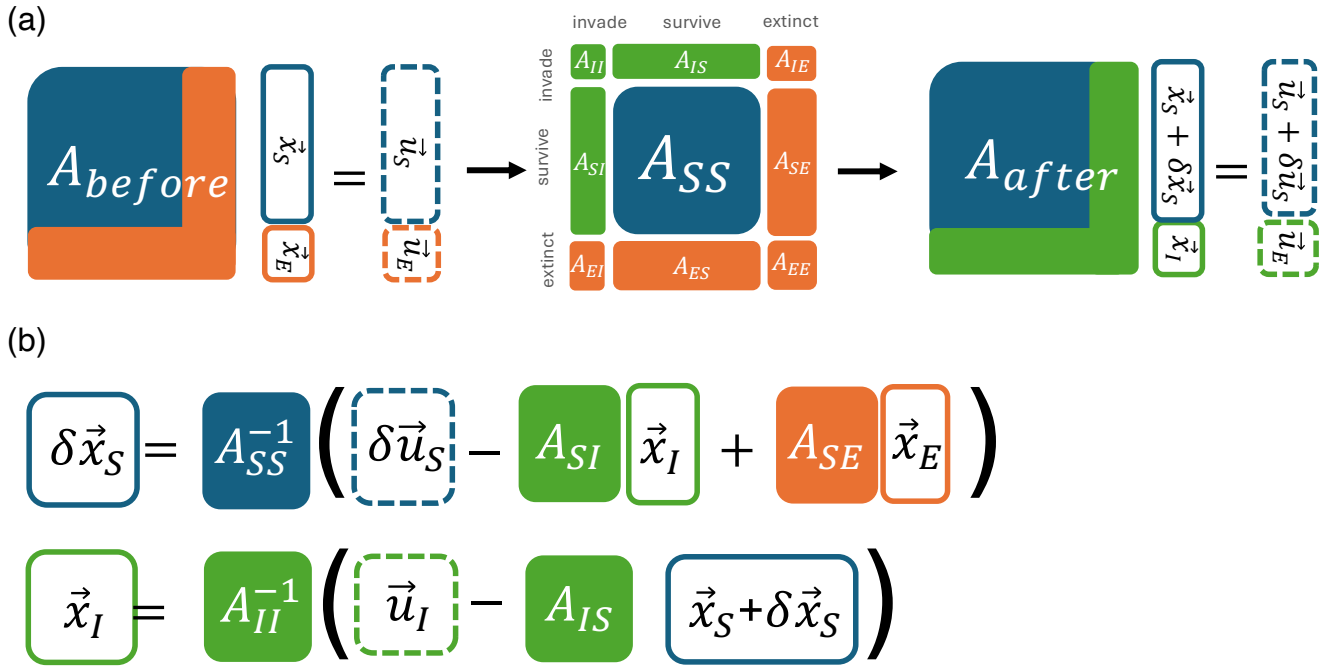

Fig. S1. Equations showing how our theory partitions an ecosystem into invaders, survivors and extinct species. (a) shows how the structure of interactions changes in the Lotka-Volterra model before and after an invasion, and (b) shows how the community shifts and invader abundance is related to each part of the system shown in Eq. (14)

D. Extended framework for multiple invaders and environmental perturbations. To extend our theory to accommodate multiple invaders, we need to separate indices of all species into those invading ($\mathcal{I}$), extinct ($\mathcal{E}$) and surviving ($\mathcal{S}$), illustrated in Fig. S1. In explicit block matrix notation, the interaction matrix of the new system consists of 9 block matrices,

$$A = \begin{pmatrix} A_{\mathcal{S}\mathcal{S}} & A_{\mathcal{S}\mathcal{I}} & A_{\mathcal{S}\mathcal{E}} \\ A_{\mathcal{I}\mathcal{S}} & A_{\mathcal{I}\mathcal{I}} & A_{\mathcal{I}\mathcal{E}} \\ A_{\mathcal{E}\mathcal{S}} & A_{\mathcal{E}\mathcal{I}} & A_{\mathcal{E}\mathcal{E}} \end{pmatrix}. \tag{14}$$

This block structure and relationship with the self-consistency equations is illustrated in the diagram in Fig. S1 (b).

In general, any perturbation to the system is composed of 3 parts, including direct environmental perturbation, invasions, and extinctions respectively. These effects can all be absorbed into an effective environmental perturbation $\vec{\delta r}^{(eff)}$,

$$\vec{\delta r}^{\text{eff}} = \vec{\delta r} + \vec{\delta r}^{\text{inv}} + \vec{\delta r}^{\text{ext}}, \tag{15}$$

where

$$\vec{r}^{\text{inv}} = -A_{\mathcal{S}\mathcal{I}}\vec{X}_{\mathcal{I}}, \quad \vec{r}^{\text{ext}} = A_{\mathcal{S}\mathcal{E}}\vec{X}_{\mathcal{E}} \tag{16}$$

The corresponding linear perturbations in abundance, the generalized version of Eq. (6) in main text, is

$$\delta\vec{X}_{\mathcal{S}} = A^{-1}_{\mathcal{S}\mathcal{S}}\delta\vec{r}^{(eff)}_{\mathcal{S}} = \delta\vec{X}_{\mathcal{S}} = A^{-1}_{\mathcal{S}\mathcal{S}}(\delta\vec{r}_{\mathcal{S}} - A_{\mathcal{S}\mathcal{I}}\vec{X}_{\mathcal{I}} + A_{\mathcal{S}\mathcal{E}}\vec{X}_{\mathcal{E}}). \tag{17}$$

At the same time, abundance of surviving invaders needs to follow the set of new steady state conditions

$$r_0 - A_{0\mathcal{S}}\vec{X}^{\text{new}}_{\mathcal{S}} - A_{\mathcal{I}\mathcal{I}}\vec{X}_{\mathcal{I}} = 0 \tag{18}$$

Substituting Eq. (17) into Eq. (18) like before, we obtain the generalized version of Eq. (5) in main text

$$\vec{X}_{\mathcal{I}} = \mathbf{M}^{-1}_{\mathcal{I}\mathcal{I}}\left(\vec{f}_I + A_{\mathcal{I}\mathcal{E}}\vec{X}_{\mathcal{E}} - A_{\mathcal{I}\mathcal{S}}A^{-1}_{\mathcal{S}\mathcal{S}}\delta\vec{r}_{\mathcal{S}} - A_{\mathcal{I}\mathcal{S}}A^{-1}_{\mathcal{S}\mathcal{S}}A_{\mathcal{S}\mathcal{E}}\vec{X}_{\mathcal{E}}\right), \tag{19}$$

where

$$\mathbf{M}_{\mathcal{I}\mathcal{I}} = A_{\mathcal{I}\mathcal{I}} - A_{\mathcal{I}\mathcal{S}}A^{-1}_{\mathcal{S}\mathcal{S}}A_{\mathcal{S}\mathcal{I}}, \tag{20}$$

$$\vec{f}_I = \vec{r}_{\mathcal{I}} - A_{\mathcal{I}\mathcal{S}}\vec{X}_{\mathcal{S}} - A_{\mathcal{I}\mathcal{E}}\vec{X}_{\mathcal{E}}. \tag{21}$$

2. Predicting invasion in the presence of multi-stability

Our framework can predict the correct invasion outcome even if the dynamical system has multiple stable fixed points, given that identity of surviving species are not significantly changed. We demonstrate this using Lotka-Volterra model with high enough variance in interaction strength. It has been shown that a phase transition change the system from unique stable steady states to multiple stable steady states (1) in random Lotka-Volterra models when the variance of insteraction increases. We demonstrate this by explicitly sampling many initial state to check if the system exhibit multiple fixed points, then estimates the portion of steady-states exhibiting multistability given the level of variance. Finally we choose a high enough variance that guarantee multi-stability happens more than 80 percent of the time.

In the main text, we sample the invader using the same distribution as that of the regional species pool, which forms the resident community before invasion. This choice models the natural occurrence of invaders originating from a statistically similar pool of species. However, when we increase the variance of interactions in the regional species pool, some interactions become negative, representing "cooperative" rather than competitive interactions. This cooperation allows species from the pool to become large-impact invaders that can potentially drive most resident species to extinction. While the self-consistency equation remains exact in such situations, the iterative scheme assumes that a considerable fraction of species survive after invasion. Therefore, in the demonstration in Fig. S2, we reduced the variance of the interaction coefficients for the pool from which we sample invaders. In this simulation, we set the mean interaction to be $\mu_A = 0.1$, the variance of the regional species pool for the resident community as $\sigma_A = 3.16$, and the variance for sampling invaders as $\sigma_A^{\text{invader}} = 0.316$. The size of the regional species pool is 10, with approximately 3 species surviving in the resident community. This demonstrates that the self-consistency equations and the iterative approach remain applicable in the presence of multiple steady states.

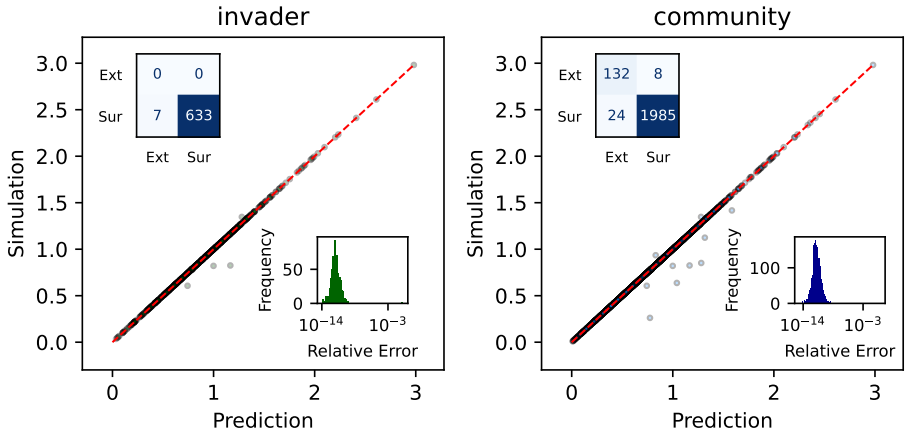

Fig. S2. Comparison of simulation outcomes and theoretical predictions for Lotka-Volterra models exhibiting multiple stable equilibria prior to invasion. The resident community's interaction matrix is sampled with high variance ($\sigma_c = 3.16$), generating systems with multiple locally stable fixed points identified by simulating dynamics from diverse initial conditions. Invader–community interaction strengths are drawn from a distribution with lower variance ($\sigma_{\text{inv}} = 0.316$), ensuring that invaders exert a limited perturbative effect on the resident community.

3. General linear response theory

We propose the following generalized form of ecological models:

$$\frac{dX_i}{dt} = f_i(X_i) g_i(\vec{X}, \vec{r}), \tag{22}$$

where X_i represents the abundance of each species, and $\vec{r}$ parameterizes the environmental influences on growth. The function $f_i(X_i)$ governs species extinction dynamics. It is positive for surviving species, and it captures the concept of extinction for biotic species as $f_i(X_i)$ must satisfy

$$f_i(0) = 0. \tag{23}$$

It simply means that biotic species cannot grow again once their abundance reaches zero, as biotic species grow by reproduction, an organic self-replication. In most models, $f_i(X_i)$ is simply X_i. Instead, for abiotic species like nutrients or metabolites that grow through supply, extinction should not apply, and we can simply set $f_i(x) = 1$.

In contrast, the function $g_i(\vec{X}, \vec{r})$ governs species growth and interactions. The steady-state condition for species abundance $\vec{X}^*$ is determined by:

$$g_i(\vec{X}^*, \vec{r}) \begin{cases} = 0 & \text{if species } i \text{ survives} \\ < 0 & \text{if species } i \text{ becomes extinct} \end{cases}, \tag{24}$$

where the second condition corresponds to the stability to re-introduction by extinct species. This tells us that instead of linearizing the complete dynamics in Eq. (22), it is more natural to linearize the steady-state condition if we consider perturbations to the ecosystem. Note that the full dynamics are *always nonlinear* because of the presence of $f_i(X_i)$. When we refer to linearization in this appendix, we refer to approximating g_i as a linear function by expanding it around a given steady-state set of abundances $\vec{X}$ and growth rates $\vec{r}$. As we show below, this approximation is mathematically equivalent to assuming that the effective interaction matrix strengths do not change too much between the pre- and post-invasion communities. We might refer to this general class of models with nonlinear g_i as "nonlinear ecological models" for brevity; to clarify all models considered in this manuscript are nonlinear in their full dynamics. We merely mean that the per-capita growth rates are nonlinear when we refer to this shorthand.

We first study the scenario where all species survive, following the first condition in Eq. (24). We want to study the abundance change to each species, $\vec{X} \to \vec{X} + \delta\vec{X}$, in response to perturbation of environmental parameter $\vec{r} \to \vec{r} + \delta\vec{r}$. The new steady state needs to satisfy

$$g_i(\vec{X} + \delta\vec{X}, r_i + \delta r_i) = 0. \tag{25}$$

Define the effective interaction matrices as the derivatives of the non-zero steady state conditions,

$$\mathcal{A}_{ij} = \left.\frac{\partial g_i(\vec{X}, \vec{r})}{\partial X_j}\right|_{\vec{X}=\vec{X}^{\rm old}}, \quad \mathcal{B}_{ij} = \left.\frac{\partial g_i(\vec{X}, \vec{r})}{\partial r_j}\right|_{\vec{X}=\vec{X}^{\rm old}} \tag{26}$$

Note that in general effective interactions $\mathcal{A}$ and $\mathcal{B}$ depend on $\vec{X}$, but assuming they remain constant at the new steady state $\vec{X}^*$ is precisely the same as performing a first-order Taylor expansion of the new steady state condition in Eq. (25). This linear approximation of the change in the steady state function around $\vec{X}^*$ yields

$$\mathcal{A}\delta\vec{X} + \mathcal{B}\delta\vec{r} = 0. \tag{27}$$

By simplifying and labeling

$$A^{-1} = -\mathcal{A}^{-1}\mathcal{B},$$

the abundance shift satisfies the linear approximation

$$\delta\vec{X} = A^{-1}\delta\vec{r}. \tag{28}$$

Here, the linearity expressed through the derivatives $\mathcal{A}$ and $\mathcal{B}$ corresponds exactly to the statement that changes in $\vec{X}$ can be well approximated by applying these effective interaction matrices to changes in $\vec{r}$, illustrating the equivalence between locally linear functions and their differential approximations.

In the context of invasion theory, we have effective environmental perturbation $\delta\vec{r}^{\rm eff}$ consists of invasion, extinction and actual environmental perturbation, the same way as linear model in Eq. (14) and Fig. S1:

$$\delta\vec{r}^{\rm eff} = \delta\vec{r} + \delta\vec{r}^{\rm inv} + \delta\vec{r}^{\rm ext}, \tag{29}$$

where

$$\delta\vec{r}^{\rm inv} = -A_{\mathcal{SI}}\vec{X}_{\mathcal{I}}, \quad \delta\vec{r}^{\rm ext} = A_{\mathcal{SE}}\vec{X}_{\mathcal{E}}. \tag{30}$$

Therefore the abundance shift to surviving species is

$$\delta X = A_{\mathcal{SS}}^{-1}(-A_{\mathcal{SI}}\vec{X}_{\mathcal{I}} + A_{\mathcal{SE}}\vec{X}_{\mathcal{E}} + \delta\vec{r}) \tag{31}$$

To summarize, to predict the linear response of the steady state in any nonlinear model, the first step is to compute the effective interaction matrix, and compute the matrix inverse given by A^{-1}. Once this matrix is determined, the problem simplifies and becomes equivalent to the generalized Lotka-Volterra model. This simple exercise of Taylor expansion justifies the use of the Lotka-Volterra model despite its simplicity - it captures the linear response of ecosystems near their steady states. Finally, the result maps to the one derived for the generalized Lotka-Volterra model in Eq. (6) in main text.

A similar derivation for Eq. (5) in main text will require solving only the equation for invader steady state abundance, i.e., $g_{i\in\mathcal{I}}(\vec{X}^{\rm new}, r_i) = 0$.

$$g_{i\in\mathcal{I}}(\vec{X}_{\mathcal{S}}^{\rm old} + \delta\vec{X}, \vec{X}_{\mathcal{E}}, X_i, r_i) = 0, \tag{32}$$

$$\sum_{j\in\mathcal{S}} \frac{\partial g_{i\in\mathcal{I},j}}{\partial X_j}\delta x_j + \sum_{k\in\mathcal{I}} \frac{\partial g_{i\in\mathcal{I},k}}{\partial X_k} x_k + g_{i\in\mathcal{I}}(\vec{X}_{\mathcal{S}}^{\rm old}, \vec{X}_{\mathcal{E}}^{\rm old} = 0, X_i = 0, r_i) = 0. \tag{33}$$

The last part of the expression corresponds almost exactly to the naive invasion fitness, except that contributions from extinct species are set to zero. This portion uniquely captures the nonlinear effects present in the growth rate, whereas the first two terms arise purely from the linear response of the system.

In matrix notation, it becomes simply

$$A_{IS}\delta\vec{X}_S + A_{\mathcal{II}}\vec{X}_{\mathcal{I}} + \vec{g}_{\mathcal{I}}(\vec{X}^{\rm old}, X_i = 0, r_0) = 0. \tag{34}$$

Therefore, the invader abundance formula in terms of dressed invasion fitness (Eq. (10) in main text) applies even for nonlinear consumer growth rate.

4. Extinct species have negative invasion fitness

If we do a perturbation around a stable steady state by adding a little bit of the extinct species i, the new growth rate follows

$$\frac{d}{dt}(x_i + \delta x_i) = (x_i + \delta x_i)(k_i - \sum_{j\neq i} A_{ij}x_j - A_{ii}\delta x_i) \tag{35}$$

for extinct species, $x_i = 0$, thus

$$\frac{d\delta x_i}{dt} = \delta x_i(k_i - \sum_{j\neq i} A_{ij}x_j - A_{ii}\delta x_i) \tag{36}$$

The terms in the bracket are almost the invasion fitness. It needs to be negative for a stable fixed point, i.e.,

$$(k_i - \sum_{j\neq i} A_{ij}x_j) < A_{ii}\delta x_i. \tag{37}$$

For a general model, the perturbation to extinct species is

$$\frac{d\delta X_i}{dt} = g_i(\vec{X},\vec{r})\frac{\partial f_i}{\partial X_i}\delta X_i + f_i(X_i)\frac{\partial g_i}{\partial X_i}\delta X_i \tag{38}$$

The analogous inequality is

$$g_i(\vec{X},\vec{r})\frac{\partial f_i}{\partial X_i}\Big|_{X_i=0} < 0 + f_i(X_i)\frac{\partial g_i}{\partial X_i}\Big|_{X_i=0} < 0 \tag{39}$$

But $f_i(0) = 0$ by definition, and we should expect $\frac{\partial f_i}{\partial X_i}\big|_{X_i=0} > 0$ because it needs to allow growth for positive abundance. Therefore, for the general model, an analogous quantity to invasion fitness is simply $g_i(\vec{X},\vec{r})\big|_{X_i=\delta X_i}$

5. Derivation of linear response in CRMs

Notice that the consumer resource model can be reformulated in a matrix form

$$\begin{pmatrix}\frac{d\vec{N}}{dt}\\ \frac{d\vec{R}}{dt}\end{pmatrix} = \begin{pmatrix}\vec{N}\\ \vec{R}\end{pmatrix} \odot \left[\begin{pmatrix}-\vec{m}\\ \vec{K}\end{pmatrix} - \begin{pmatrix}0 & -C\\ E^{\mathrm{T}} & Q\end{pmatrix}\begin{pmatrix}\vec{N}\\ \vec{R}\end{pmatrix}\right] \tag{40}$$

Comparing with the Lotka-Volterra model in matrix form,

$$\frac{d\vec{X}}{dt} = \vec{X} \odot (\vec{r} - A\vec{X}) \tag{41}$$

We have the mapping

$$\vec{X} = \begin{pmatrix}\vec{N}\\ \vec{R}\end{pmatrix}, \quad \vec{r} = \begin{pmatrix}-\vec{m}\\ \vec{K}\end{pmatrix}, \quad A = \begin{pmatrix}0 & -C\\ E^{\mathrm{T}} & Q\end{pmatrix} \tag{42}$$

For the surviving subset of species, interaction matrices are invertible, and can be derived in block matrix form

$$A^{-1}_{\mathcal{SS}} = -\begin{pmatrix}-I & B^{-1}C\\ Q^{-1}E^{\mathrm{T}} & I-P\end{pmatrix}_{\mathcal{SS}}\begin{pmatrix}B^{-1} & 0\\ 0 & Q^{-1}\end{pmatrix}_{\mathcal{SS}} \tag{43}$$

$$= \begin{pmatrix}B^{-1} & B^{-1}CQ^{-1}\\ Q^{-1}E^{\mathrm{T}}B^{-1} & (I-P)Q^{-1}\end{pmatrix}_{\mathcal{SS}} \tag{44}$$

where B and P are defined to simplify the expressions

$$B = CQ^{-1}E^{T}, \quad P = Q^{-1}E^{T}B^{-1}C. \tag{45}$$

For consumer invader, we have the invader

$$X_0 = \begin{pmatrix}\vec{N_0}\\ 0\end{pmatrix}, \quad \vec{r_0} = \begin{pmatrix}-m_0\\ 0\end{pmatrix}, \quad A_{0S/E} = \begin{pmatrix}0 & -C_{0S/E}\end{pmatrix}, \quad A_{S/E0} = \begin{pmatrix}0\\ E^{\mathrm{T}}_{S/E0}\end{pmatrix} \tag{46}$$

We can substitute these block-form Lotka-Volterra quantities to Eq. (5) and Eq. (6) in main text to obtain

$$A_{0\mathcal{S}}A^{-1}_{\mathcal{SS}}A_{\mathcal{S}0} = C_{0\mathcal{S}}(1-P)Q_{\mathcal{SS}}E^{\mathrm{T}}_{\mathcal{S}0}, \tag{47}$$

$$A_{0\mathcal{S}}A^{-1}_{\mathcal{SS}}A_{\mathcal{S}\mathcal{E}} = C_0E^{T}B^{-1}C_{\mathcal{S}\mathcal{E}}\vec{R}_{\mathcal{E}} - C_{0S}(1-P)Q_{\mathcal{S}\mathcal{E}}\vec{N}_{\mathcal{E}}, \tag{48}$$

$$A_{\mathcal{I}\mathcal{E}}x_{\mathcal{E}} = -C_{0\mathcal{E}}R_{\mathcal{E}}, \tag{49}$$

which pieces together the invader abundance

$$N_0 = \frac{f_0 - C_{0\mathcal{E}} R_{\mathcal{E}} - C_{0\mathcal{S}} E_{\mathcal{S}\mathcal{S}}^T B^{-1} C_{\mathcal{S}\mathcal{E}} \vec{R}_{\mathcal{E}} + C_{0\mathcal{S}} P Q_{\mathcal{S}\mathcal{E}} \vec{N}_{\mathcal{E}}}{C_{0\mathcal{S}}(1-P) Q_{\mathcal{S}\mathcal{S}}^{-1} E_{\mathcal{S}0}^{\mathrm{T}}} \tag{50}$$

where invader fitness is $f_0 = C_{0\mathcal{E}} R_{\mathcal{E}} + C_{0\mathcal{S}} R_{\mathcal{S}} - m_0$ And the abundance shifts to consumers and resources are

$$\begin{pmatrix} \delta N_{\mathcal{S}} \\ \delta R_{\mathcal{S}} \end{pmatrix} = \begin{pmatrix} B^{-1} & B^{-1} C Q^{-1} \\ Q^{-1} E^{\mathrm{T}} B^{-1} & (I-P) Q^{-1} \end{pmatrix}_{\mathcal{S}\mathcal{S}} \left[\begin{pmatrix} 0 & -C \\ E^{\mathrm{T}} & Q \end{pmatrix}_{\mathcal{S}\mathcal{E}} \begin{pmatrix} N_{\mathcal{E}} \\ R_{\mathcal{E}} \end{pmatrix} - \begin{pmatrix} 0 & -C \\ E^{\mathrm{T}} & Q \end{pmatrix}_{\mathcal{S}0} \begin{pmatrix} N_0 \\ 0 \end{pmatrix} \right] \tag{51}$$

$$= \begin{pmatrix} B^{-1} & B^{-1} C_{\mathcal{S}\mathcal{S}} Q_{\mathcal{S}\mathcal{S}}^{-1} \\ Q_{\mathcal{S}\mathcal{S}}^{-1} E_{\mathcal{S}\mathcal{S}}^{\mathrm{T}} B^{-1} & (I-P) Q_{\mathcal{S}\mathcal{S}}^{-1} \end{pmatrix} \begin{pmatrix} -C_{\mathcal{S}\mathcal{E}} R_{\mathcal{E}} \\ E_{\mathcal{S}\mathcal{E}}^T N_{\mathcal{E}} + Q_{\mathcal{S}\mathcal{E}} R_{\mathcal{E}} - E_{\mathcal{S}0}^T N_0 \end{pmatrix} \tag{52}$$

Expanding, we have

$$\delta N_{\mathcal{S}} = -B^{-1} C_{\mathcal{S}\mathcal{S}} Q_{\mathcal{S}\mathcal{S}}^{-1} E_{\mathcal{S}0}^{\mathrm{T}} N_0 - B^{-1} C_{\mathcal{S}\mathcal{E}} R_{\mathcal{E}} + B^{-1} C_{\mathcal{S}\mathcal{S}} Q_{\mathcal{S}\mathcal{S}}^{-1} (E_{\mathcal{S}\mathcal{E}}^{\mathrm{T}} N_{\mathcal{E}} + Q_{\mathcal{S}\mathcal{E}} R_{\mathcal{E}}) \tag{53}$$

$$\delta R_{\mathcal{S}} = -(I-P) Q_{\mathcal{S}\mathcal{S}}^{-1} E_{\mathcal{S}0}^{\mathrm{T}} N_0 - Q_{\mathcal{S}\mathcal{S}}^{-1} E_{\mathcal{S}\mathcal{S}}^{\mathrm{T}} B^{-1} C_{\mathcal{S}\mathcal{E}} R_{\mathcal{E}} + (I-P) Q_{\mathcal{S}\mathcal{S}}^{-1} (E_{\mathcal{S}\mathcal{E}}^{\mathrm{T}} N_{\mathcal{E}} + Q_{\mathcal{S}\mathcal{E}} R_{\mathcal{E}}) \tag{54}$$

6. Linear response of nonlinear consumer resource model

We can apply the linearization framework to specially non-linear consumer resource models, similarly to how we map general nonlinear ecological models to generalized Lotka-Volterra models.

Consider general consumer resource models of the form

$$\frac{dN_i}{dt} = f_i^N(N_i) g_i^N(\vec{N}, \vec{R}, \vec{m}) \tag{55}$$

$$\frac{dR_\alpha}{dt} = f_\alpha^R(R_\alpha) g_\alpha^R(\vec{N}, \vec{R}, \vec{k}) \tag{56}$$

In such general models, the per capita growth rate functions $g_i^N(\vec{N}, \vec{R}, \vec{m})$ and $g_\alpha^R(\vec{N}, \vec{R}, \vec{k})$ of consumer species i and resources α respectively are nonlinear functions of consumer abundances N_i and/or resource abundances R_α. Note that even when g_i^N and/or g_α^R are linear, the full dynamics are *always nonlinear* because of the presence of the function $f_i^N(N_i)$ which is usually simply $f_i^R(N_i) = N_i$ (Eqs. (55) and (56)). Similarly, $f_\alpha^R(R_\alpha)$ is typically—though not always—R_α. In the remaining appendices, for brevity we might sometimes refer to such models as "nonlinear consumer-resource models": we mean here that the per-capita growth rate functions g_i^N and g_α^R are nonlinear in N_i and/or R_α (Eq. (55) and Eq. (56))

The steady state condition for surviving species is

$$0 = g_i^N(\vec{N}, \vec{R}, \vec{m}) \tag{57}$$

$$0 = g_\alpha^R(\vec{N}, \vec{R}, \vec{k}) \tag{58}$$

The new steady state condition is

$$0 = g_i^N(\vec{N} + \delta\vec{N}, \vec{R} + \delta\vec{R}, \vec{m} + \delta\vec{m}) \tag{59}$$

$$0 = g_\alpha^R(\vec{N} + \delta\vec{N}, \vec{R} + \delta\vec{R}, \vec{k} + \delta\vec{k}) \tag{60}$$

The linear expansion leads to

$$\sum_j \left.\frac{\partial g_i^N(\vec{N}, \vec{R}, \vec{m})}{\partial N_j}\right|_{\vec{N}=\vec{N}_{\mathcal{S}\mathcal{S}}} \delta N_j + \sum_\alpha \left.\frac{\partial g_i^N(\vec{N}, \vec{R}, \vec{m})}{\partial R_\alpha}\right|_{\vec{R}=\vec{R}_{\mathcal{S}\mathcal{S}}} \delta R_\alpha + \sum_j \frac{\partial g_i^N(\vec{X}_{\mathcal{S}\mathcal{S}}, \vec{m}_i)}{\partial m_j} \delta m_j = 0. \tag{61}$$

$$\sum_j \left.\frac{\partial g_\alpha^R(\vec{N}, \vec{R}, \vec{m})}{\partial N_j}\right|_{\vec{N}=\vec{N}_{\mathcal{S}\mathcal{S}}} \delta N_j + \sum_\alpha \left.\frac{\partial g_\alpha^R(\vec{N}, \vec{R}, \vec{m})}{\partial R_\beta}\right|_{\vec{R}=\vec{R}_{\mathcal{S}\mathcal{S}}} \delta R_\beta + \sum_j \frac{\partial g_\alpha^R(\vec{X}_{\mathcal{S}\mathcal{S}}, \vec{m}_i)}{\partial k_\beta} \delta k_\beta = 0. \tag{62}$$

Defining each effective interaction matrix,

$$\left.\frac{\partial g_i^N(\vec{N}, \vec{R}, \vec{m})}{\partial N_j}\right|_{\vec{N}=\vec{N}_{\mathcal{S}\mathcal{S}}} = \mathcal{Q}_{ij}^N, \quad \left.\frac{\partial g_i^N(\vec{N}, \vec{R}, \vec{m})}{\partial R_\alpha}\right|_{\vec{R}=\vec{R}_{\mathcal{S}\mathcal{S}}} = \mathcal{C}_{i\alpha}, \quad \frac{\partial g_i^N(\vec{X}_{\mathcal{S}\mathcal{S}}, \vec{m}_i)}{\partial m_j} = \mathcal{B}_{ij}^N. \tag{63}$$

$$\left.\frac{\partial g_\alpha^R(\vec{N}, \vec{R}, \vec{m})}{\partial N_j}\right|_{\vec{N}=\vec{N}_{\mathcal{S}\mathcal{S}}} = -\mathcal{E}_{\alpha j}, \quad \left.\frac{\partial g_\alpha^R(\vec{N}, \vec{R}, \vec{m})}{\partial R_\beta}\right|_{\vec{R}=\vec{R}_{\mathcal{S}\mathcal{S}}} = -\mathcal{Q}_{\alpha\beta}^R, \quad \frac{\partial g_\alpha^R(\vec{X}_{\mathcal{S}\mathcal{S}}, \vec{m}_i)}{\partial k_\beta} \delta k_\beta = \mathcal{B}_{\alpha\beta}^R. \tag{64}$$

In matrix notation the linear equation becomes

$$\mathcal{Q}^N \delta\vec{N} + \mathcal{C}\delta\vec{R} + \mathcal{B}^N \delta\vec{m} = 0. \tag{65}$$

$$-\mathcal{E}\delta\vec{N} - \mathcal{Q}^R\delta\vec{R} + \mathcal{B}^R\delta\vec{k} = 0. \quad [66]$$

$$\begin{pmatrix} Q^N & \mathcal{C} \\ -\mathcal{E} & Q^R \end{pmatrix}\begin{pmatrix} \delta\vec{N} \\ \delta\vec{R} \end{pmatrix} = \begin{pmatrix} \mathcal{B}^N & 0 \\ 0 & \mathcal{B}^{\mathcal{R}} \end{pmatrix}\begin{pmatrix} \delta\vec{m} \\ \delta\vec{k} \end{pmatrix} \quad [67]$$

$$\begin{pmatrix} (\mathcal{B}^N)^{-1} & 0 \\ 0 & (\mathcal{B}^{\mathcal{R}})^{-1} \end{pmatrix}\begin{pmatrix} Q^N & \mathcal{C} \\ -\mathcal{E} & Q^R \end{pmatrix}\begin{pmatrix} \delta\vec{N} \\ \delta\vec{R} \end{pmatrix} = \begin{pmatrix} \delta\vec{m} \\ \delta\vec{k} \end{pmatrix} \quad [68]$$

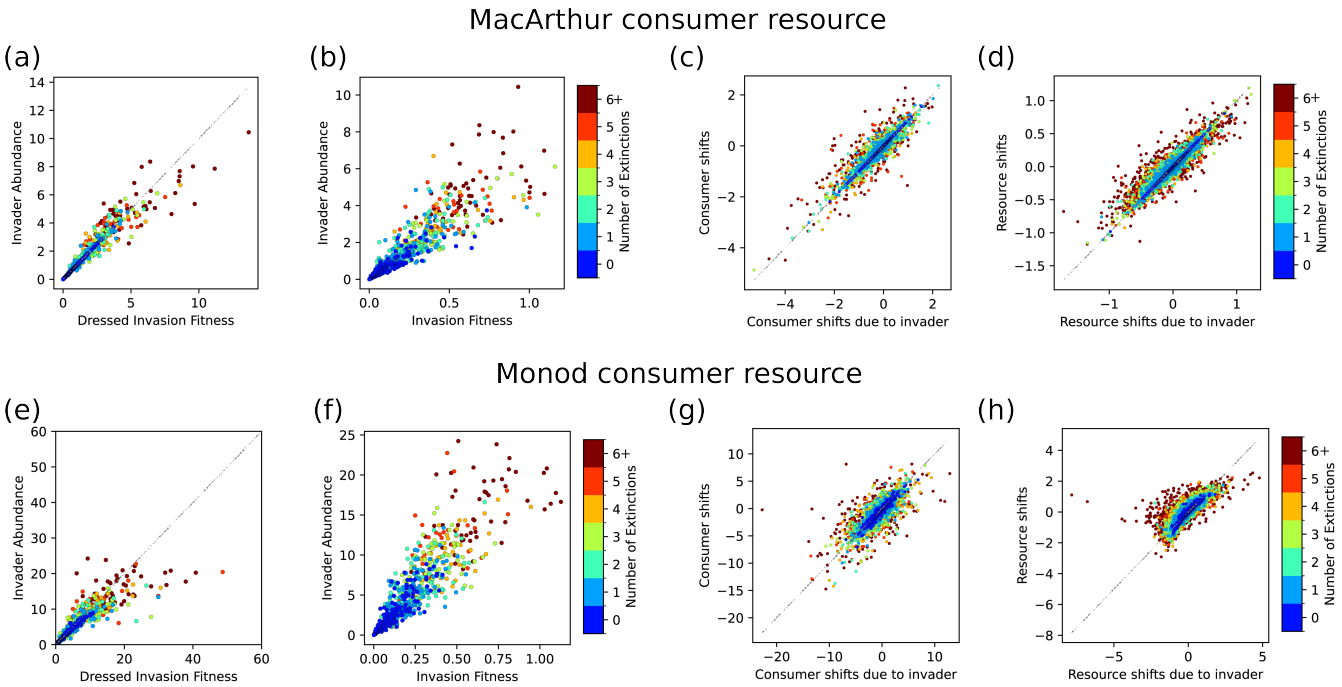

Fig. S3. Dressed invasion fitness and extinction-induced feedbacks govern invasion outcomes in additional consumer-resource models. Simulations of (a–d) the MacArthur Consumer Resource Model (MCRM) and (e–h) a Monod growth consumer-resource model show qualitatively similar results as the Microbial Consumer Resource Model (MiCRM) presented in the main text (Fig. 3).

It can be solved with block matrix inversion

$$\begin{pmatrix} \delta\vec{N} \\ \delta\vec{R} \end{pmatrix} = \begin{pmatrix} (\mathcal{B}^N)^{-1}\mathcal{Q}^N & -(\mathcal{B}^N)^{-1}\mathcal{C} \\ (\mathcal{B}^{\mathcal{R}})^{-1}\mathcal{E} & (\mathcal{B}^{\mathcal{R}})^{-1}\mathcal{Q}^R \end{pmatrix}^{-1}\begin{pmatrix} \delta\vec{m} \\ \delta\vec{k} \end{pmatrix} \quad [69]$$

For the generalized consumer resource model discussed previously, $(\mathcal{B}^N)^{-1} = -I$, $(\mathcal{B}^R)^{-1} = I$, $\mathcal{Q}^N = 0$, $\mathcal{Q}^R = Q$, $\mathcal{C} = C$, $\mathcal{E} = E$.

In general, if $\mathcal{Q}^N = 0$, we can retrieve the result for nonlinear consumer resource model by substitution

$$C \to -(\mathcal{B}^N)^{-1}\mathcal{C}, \quad [70]$$

$$E \to (\mathcal{B}^R)^{-1}\mathcal{E}, \quad [71]$$

$$Q^R \to (\mathcal{B}^{\mathcal{R}})^{-1}\mathcal{Q}^R. \quad [72]$$

7. Linear reponse in Monod CRM with type II functional response

This section explicitly demonstrates the applicability of our framework to ecological models of consumer-resource dynamics where the effective interactions depend on community state as described in last section. One generic way this can occur is when the per capita growth rate functions $g_i^N(\vec{N}, \vec{R}, \vec{m})$ of consumer species i are nonlinear functions of consumer abundances N_i and/or resource abundances R_α (see Eqs. (55) and (56) for definition).

A paradigmatic case of such consumer-resource models is an extension of the MacArthur consumer-resource model (where g_i is a linear function, even though the full dynamics are nonlinear). In this extension, consumer species dynamics $\dot{N}_i$ are governed by Monod growth rate functions (Type II functional responses) instead of linear functions of R_α, as

$$\frac{dN_i}{dt} = N_i\left(\sum_\alpha \frac{C_{i\alpha}R_\alpha}{h+R_\alpha} - m_i\right), \quad [73]$$

$$\frac{dR_\alpha}{dt} = R_\alpha\left(K_\alpha - R_\alpha - \sum_j \frac{E_{j\alpha}N_j}{h+R_\alpha}\right). \quad [74]$$

The steady state condition is

$$0 = g_i^N(\vec{N},\vec{R},\vec{m}) = \sum_\alpha \frac{C_{i\alpha}R_\alpha}{h+R_\alpha} - m_i, \quad [75]$$

$$0 = g_\alpha^R(\vec{N},\vec{R},\vec{m}) = K_\alpha - R_\alpha - \sum_j \frac{E_{j\alpha}N_j}{h+R_\alpha}. \quad [76]$$

Effective interactions. The main difference between Type I and Type II responses is that the consumer preferences $C_{i\alpha}$, impacts $E_{i\alpha}$, and $Q_{\alpha\beta}$ must now be replaced by effective quantities $\mathcal{C}_{i\alpha}$, $\mathcal{E}_{\alpha j}$, and $Q^{(R)}_{\alpha\beta}$ that have to be calculated from the full non-linear dynamics. These effective quantities are precisely what we call *effective interactions*, since they represent how consumer species interact with resources and vice versa. Note also that $Q^{(R)}_{\alpha\beta}$ captures the effective interactions between resources. More generic models might similarly have effective interactions between species, say $B^{(N)}_{ij}$, but here $B = 0$. One therefore has

$$\mathcal{C}_{i\alpha} = \left.\frac{\partial g_i^N(\vec{N},\vec{R},\vec{m})}{\partial R_\alpha}\right|_{\vec{R}=\vec{R}_{\mathcal{SS}}} = \frac{C_{i\alpha}h}{(h+R_\alpha)^2}, \quad [77]$$

$$\mathcal{Q}^R_{\alpha\beta} = -\left.\frac{\partial g_\alpha^R(\vec{N},\vec{R},\vec{m})}{\partial R_\beta}\right|_{\vec{R}=\vec{R}_{\mathcal{SS}}} = \delta_{\alpha\beta} - \delta_{\alpha\beta}\sum_j \frac{E_{j\alpha}N_j}{(h+R_\alpha)^2}, \quad [78]$$

and

$$\mathcal{E}_{\alpha j} = -\left.\frac{\partial g_\alpha^R(\vec{N},\vec{R},\vec{m})}{\partial N_j}\right|_{\vec{N}=\vec{N}_{\mathcal{SS}}} = \frac{E_{j\alpha}}{h+R_\alpha}. \quad [79]$$

Using these effective interactions, we can subsequently solve for the invader abundance in the same way as performed with the MCRM and MiCRM models. The remaining challenge lies in addressing the nonlinear invasion fitness:

$$f_0 = \sum_\alpha \frac{C_{0\alpha}R_\alpha^{\text{old}}}{h+R_\alpha} - m \quad [80]$$

For convenience, in this model, this nonlinearity can be absorbed in the previous expression of invasion fitness simply by $C_{0\alpha} \to \frac{C_{0\alpha}}{h+R_\alpha}$. In summary, there are 4 replacements to normal MCRM solutions, including

$$C_{i\alpha} \to \frac{C_{i\alpha}h}{(h+R_\alpha)^2}, \quad [81]$$

$$E_{j\alpha} \to \frac{E_{j\alpha}}{(h+R_\alpha)}, \quad [82]$$

$$Q_{\alpha\beta} \to \delta_{\alpha\beta} - \delta_{\alpha\beta}\sum_j \frac{E_{j\alpha}N_j}{(h+R_\alpha)^2}, \quad [83]$$

$$C_{0\alpha} \to \frac{C_{0\alpha}}{h+R_\alpha}, \quad \text{for invasion fitness only}, \quad [84]$$

Since all effective interaction matrices are proportional to $1/(h+R_\alpha)$, changes in this quantity after invasion serve as a good measure of the model's nonlinearity. In Fig. S5, we demonstrate that the mean relative error of the framework increases linearly with the mean relative shift of this value.

8. Linear response in the microbial consumer resource model

The Microbial consumer resource model (MiCRM) explicitly incorporates cross-feeding by modeling how metabolic byproducts from one species become resources for others. It describes the dynamics of microbial abundances N_i and resource concentrations R_α,

$$\frac{dN_i}{dt} = N_i(\sum_\alpha (1-l_\alpha) w_\alpha c_{i\alpha} R_\alpha - m_i) \quad [85]$$

$$\frac{dR_\alpha}{dt} = K_\alpha - \omega R_\alpha - \sum_j c_{j\alpha} N_j R_\alpha + \sum_{j,\beta} l_\beta \frac{w_\alpha}{w_\beta} D_{\alpha\beta} c_{j\beta} R_\beta N_j, \quad [86]$$

where l_α is the fraction of consumed resource secreted as byproducts, $D_{\alpha\beta}$ is the cross-feeding matrix describing conversion between resources, w_α are resource weights reflecting energetic values of resources, $c_{i\alpha}$ are resource preferences, and environmental factors are contained in resource supply K_α and decay ω. The steady state of this model, specifically the resource part, is not linear, thus requiring us to first linearize the model.

Defining effective depletion rate of resource α by species i:

$$\kappa_{i\alpha}^{\text{eff}} = -c_{i\alpha} R_\alpha + \sum_\beta l_\beta \frac{w_\alpha}{w_\beta} D_{\alpha\beta} c_{i\beta} R_\beta \quad [87]$$

The steady state condition can be written as

$$0 = \sum_\alpha (1-l_\alpha) w_\alpha c_{i\alpha} R_\alpha - m_i \quad [88]$$

$$0 = \frac{K_\alpha}{\omega} - R_\alpha + \sum_j \frac{\kappa_{i\alpha}^{\text{eff}}}{\omega} N_j \quad [89]$$

Applying the formulation in the last section to arrive at the effective interactions in this Microbial consumer resource model, we arrive at the following effective interactions by following the same steps as in the last two appendices:

$$C_{i\alpha} \to (1-l_\alpha) w_\alpha c_{i\alpha} \quad [90]$$

$$E_{i\alpha} \to -\kappa_{i\alpha}^{\text{eff}}/\omega \quad [91]$$

$$Q_{\alpha\beta} \to -\frac{\partial}{\partial R_\beta}(\frac{K_\alpha}{\omega} - R_\alpha + \sum_j \frac{\kappa_{i\alpha}^{\text{eff}}}{\omega} N_j) \quad [92]$$

$$= \delta_{\alpha\beta}(1 + \frac{1}{\omega}\sum_j c_{j\alpha} N_j) - l_\beta \frac{w_\alpha}{w_\beta} D_{\alpha\beta} \sum_j c_{j\beta} N_j \quad [93]$$

9. Algorithm to solve the self-consistency equations

Here we present a simple algorithm to solve the self-consistency equations iteratively. Algorithm 1 outlines the basic logic and implementation steps. We assess convergence by monitoring the extinction boolean indicators that determine the extinction and survival status of each species. The iteration terminates when these indicators remain unchanged following an update, indicating that the extinction and survival states have stabilized. If convergence is not reached, we impose a maximum of 50 iterations, after which we accept the last extinction boolean prediction as the final result. While Algorithm 1 performs reasonably well on its own, we incorporate several numerical improvements to enhance stability and better accommodate complexities arising in simulation results and experimental data:

First, when calculating the abundance shifts of resident species, we introduce a momentum hyperparameter at step 6. We chose its value to be 0.1, meaning the new iteration contributes 10% weight while the accumulated value carries 90% weight. This parameter incorporates a weighted contribution from the abundance shift of the previous iteration into the current update, smoothing fluctuations and accelerating convergence. Such momentum-based updates are a well-known technique in numerical optimization to improve stability. However, to truly reflect the self-consistency equations, we use the final extinction boolean to recompute the abundances of both invaders and surviving species, so the memory from previous iterations is only retained in this boolean state.

Second, numerical uncertainties in abundance data before invasion mean that extinct species may not reach exact zero abundances in practice. To address this, at step 7 we implement a survival threshold, a small cutoff value ranging from 10^{-3} to 10^{-7} depending on the magnitude of species abundances in the particular model. Species with predicted abundances below this threshold are treated as extinct, instead of below exactly zero as described in step 7. This step effectively handles numerical noise and improves extinction boolean predictions.

Third, for models with nonlinear invasion fitness terms, such as Monod consumer resource model, we provide an additional input argument representing the effective consumer preference matrix used to compute nonlinear invasion fitness. This matrix

is distinct from the ordinary interaction matrix A and extend to complex models (see Appendix 7 for the detail example for Monod model).

Together, these modifications improve the robustness and applicability of the iterative algorithm across diverse systems, allowing accurate prediction of species abundances and extinction patterns following invasion.

The full code implementation underlying all results in this paper is publicly available at https://github.com/Emergent-Behaviors-in-Biology/invasion-theory

Algorithm 1 Iterative solver for self-consistency equations

Input: Pre-invasion system interaction matrix A, species abundances $\vec{X}$, growth rates $\vec{r}$, invader interaction matrices A_{II}, A_{IS}, A_{SI}, and invader growth rates u_I, maximum number of iterations `max_num_iters`.
Initialize: Set extinction boolean *Ebool* to `False` for all species (i.e., assume no species are extinct)
for $i = 1$ to `max_num_iters` **do**
    According to *Ebool*, extract submatrices and vectors for surviving, extinct, and invader species as described in Eq. (14)
    Solve invader abundances using Eq. (5) in main text
    Compute abundance shifts for the surviving community using Eq. (6) in main text
    Update extinction boolean *Ebool*:
    - Set *Ebool* = `True` for species with negative abundance
    - Set *Ebool* = `False` for species with positive invasion fitness

    if convergence condition met then
        Break loop
return Final predicted invader abundances, community shift, and extinction boolean

10. Error analysis

The errors observed in the predictions compared to simulation results in Fig. 2 originate from multiple sources. Broadly, the error of the framework prediction arise from two components: (1) the self-consistency equations themselves and (2) the algorithm used to solve them. The primary source of error may vary depending on the particular model. Theoretically, the self-consistency equations are exact for models with linear steady-state conditions, including GLV (Eq. (5), Eq. (6) in main text) and MCRM Eq. (50) to Eq. (54) . To demonstrate this, Fig. S6 presents the results of solving the equations for the MCRM model using the same simulations as in Fig. 2(e)–(g), but with the extinction and survival states of each species after invasion provided explicitly rather than inferred through our iterative algorithm. This is done to disambiguate errors in our algorithm and approximations in our self-consistency equations. Notably, all visible errors in Fig. 2(e)–(g) disappear once the error introduced by the algorithm is removed (except numerical errors coming from ODE simulations themselves, between a relative 10^{-9} to 10^{-2}. Since our iterative algorithm explores the boolean space of extinction and survival states locally, beginning from an initial condition where all resident species are assumed to survive, its performance improves when the true states do not deviate substantially from this starting point. Consequently, the mean relative error of invader abundance in Fig. S4(b) declines with number of consumers after the invasion, despite the error distribution remaining bimodal, with peaks centered around 10^{-7} (likely due to numerical simulation uncertainty) and 10^{-2} (attributable to the algorithm) respectively. A similar trend is observed in the Lotka-Volterra model shown in Fig. S4(a), although only two instances of substantial error occur, both at small consumer numbers of 4 and 5.

For nonlinear models such as MiCRM and Monod, the primary source of error likely arises from the self-consistency equations using a linear response approximation. This approximation assumes that the effective interactions among surviving species do not change significantly after invasion as described from Eq. (25) to Eq. (28). However, since these effective interactions depend on species or resource abundances, the resulting solutions from the self-consistency equations deviate from the actual invasion outcomes. This effect is clearest to test in the Monod model. As demonstrated in Fig. S5, the mean of the logarithm of the relative error increases nearly linearly with the mean of the logarithm of relative change in the effective interactions, confirming that deviations in these interactions contribute directly to the error. Moreover, this linear response approximation becomes more accurate in more species-diverse ecosystems, as the relative shift in species abundance caused by a single invader diminishes with increasing species number. This consistent trend is also evident in Fig. S4 (c)–(d), where the logarithm of the relative error decreases as the species diversity (number of surviving consumer species) increases.

11. Numerical details for simulations

A. Lotka-Volterra Model. In Fig. 2(b)-(c), and Fig. 3(c)-(e), we simulated an ensemble of 640 Lotka-Volterra models. We assembled each system's pre-invasion community with $n = 10$ species, which resulted in an average of 8 surviving species. The interaction matrices followed statistics

$$\langle A_{ij} \rangle = \mu_A/n + \delta_{ij}, \langle A_{ij}A_{kl} \rangle = \delta_{ij}\delta_{kl}\frac{\sigma_A^2}{n} + \delta_{il}\delta_{jk}\frac{\rho\sigma_A^2}{n}. \quad [94]$$

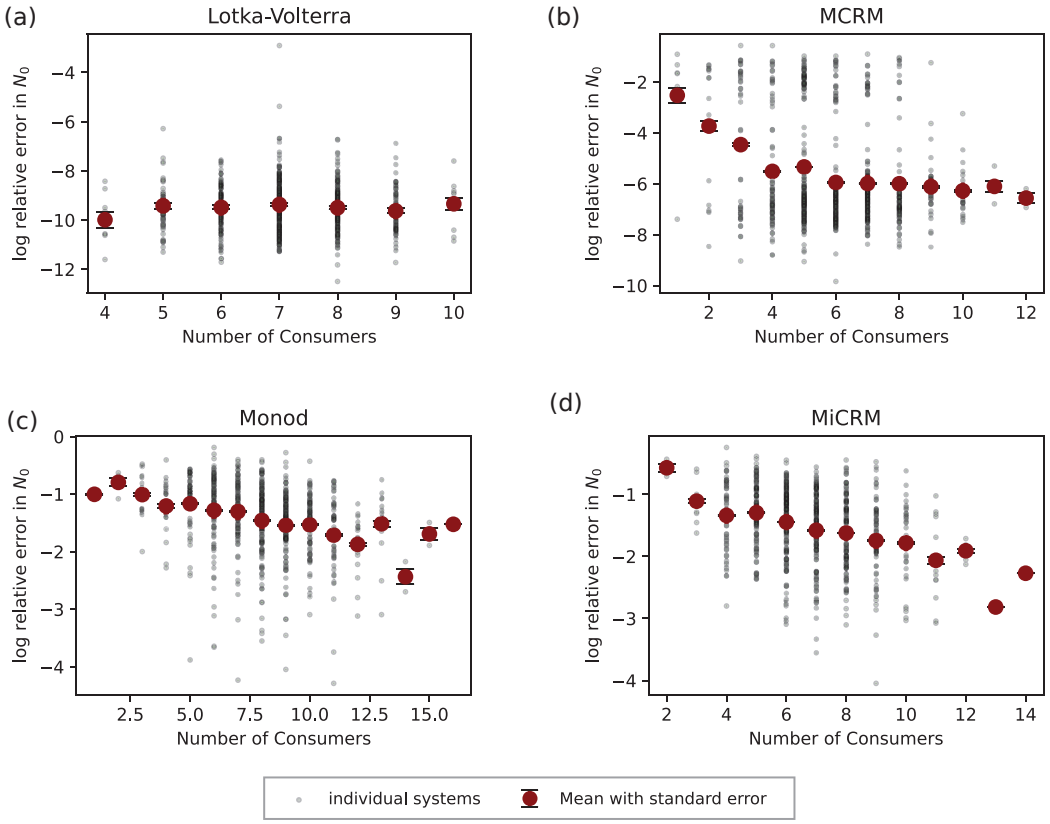

Fig. S4. Error decreases with system size With the same simulation data used in Fig. 2, we grouped predictions into different number of consumers after the invasions. For (a) Lokta-Volterra and (b) MacArthur consumer resource models, the decreasing trend saturate when the system is large enough because the self-consistency equation is exact and errors are only coming from the algorithm to deduce identity of extinct species, which also improves with system size. For (c) consumer resource models with Monod functional response and (d) Microbial consumer resource models, the error shows consistent decreasing trend because of the linear-response approximation to non-linear models.

And the environmental parameters similarly were independent and had mean μ_r/n and variance σ_r^2/n. In this figure we chose $\mu_A = 2$, $\sigma_A = 0.2$, $\rho = 1$ for a symmetric model, and $\mu_r = 2$, $\sigma_r = 0.1$. We chose the threshold for extinction as 10^{-5}. Invaders were uncorrelated to the system but shared the same statistics of interaction matrix and environmental parameters. We performed simulations with numerical ODE simulator 'LSODA', with parameters atol=1e-12, rtol=1e-12, and checked convergence by ensuring the maximum derivative was less than 1e-12.

B. MacArthur consumer resource model. In Fig. 2(e)–(g) and Fig. S3(a)–(d), we simulated an ensemble of 640 generalized MacArthur consumer resource models. We assembled each system's pre-invasion community with 60 consumers and $n = 30$ resources, which resulted in an average of 7 surviving consumers and 22 surviving resources. In this model, we set consumer extinction threshold at 10^{-3} and resource extinction threshold at 10^{-9}. The consumer preference matrix followed statistics:

$$\langle C_{i\alpha}\rangle = \frac{\mu_C}{n}, \langle C_{i\alpha}C_{j\beta}\rangle = \delta_{i\alpha}\delta_{j\beta}\frac{\sigma_C^2}{n} \quad [95]$$

In the figure we used $\mu_C = 60$, $\sigma_C = 0.5$. For demonstration purpose of the generality, the resource had a self-interaction matrix Q with statistics,

$$\langle Q_{ij}\rangle = \mu_Q/n + \delta_{ij} \quad [96]$$

$$\langle Q_{ij}Q_{kl}\rangle = \delta_{ij}\delta_{kl}\frac{\sigma_Q^2}{n} + \delta_{il}\delta_{jk}\frac{\rho_Q\sigma_Q^2}{n}, \quad [97]$$

where we used $\mu_Q = 0.1, \sigma_Q = 0.1, \rho_Q = 0.1$.

For environmental parameters, we had mean and variance of consumer mortality rate, μ_m/n, σ_m^2/n, with $\mu_m = 1$, $\sigma_m = 0.1$. And mean and variance of carrying capacity were, μ_k/n, σ_k^2/n, with $\mu_k = 2$, $\sigma_k = 0.1$. Again, the invader parameters were uncorrelated with the system and followed the same statistics. We used a constraint optimization algorithm instead of numerical ODE for simulation. The algorithm was similar to the one used in (2).

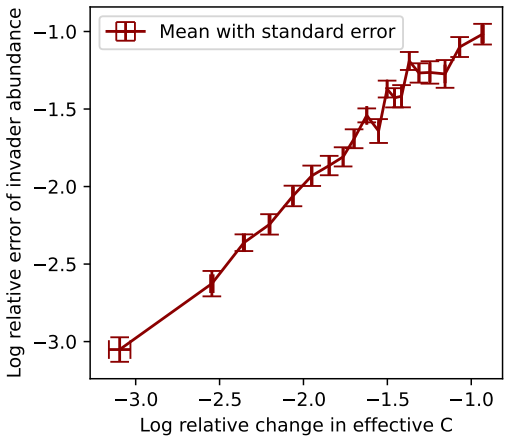

Fig. S5. Error increases with degree of nonlinearity in Monod consumer resource model The effective interaction in Monod consumer resource model are all proportional to $1/(h+R)$, The mean relative shift in this quantity serves as a direct measure of nonlinearity, quantifying the extent to which our linear-response approximation breaks down as nonlinearity increases.

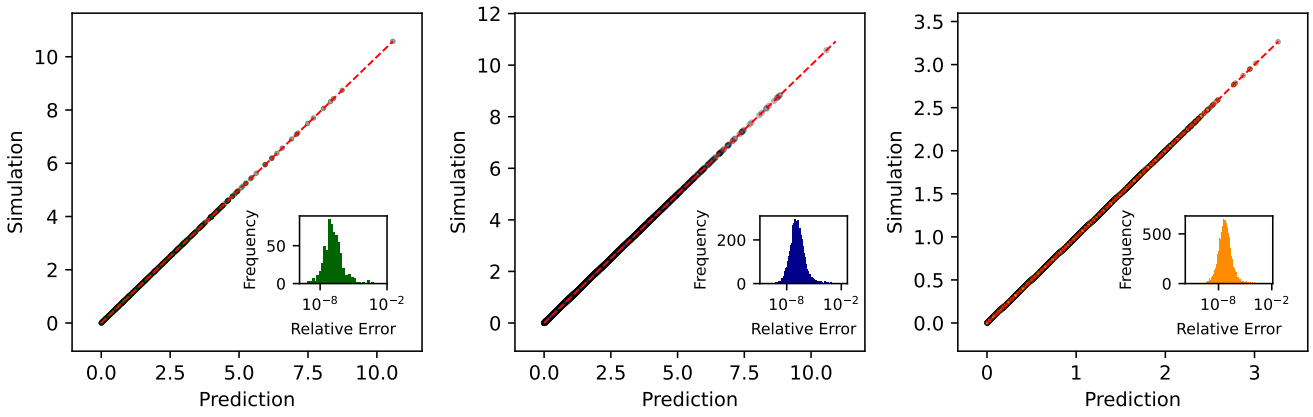

Fig. S6. Perfect prediction of MCRM when correct information of extinction is given. We demonstrate that the error distribution in Fig. 2(e)–(g) only has a larger second peak because of wrong boolean prediction of extinctions of species. When this information is given, the larger peak disappear and left with the smaller peak that only comes from numerical uncertainty of simulations.

C. Microbial consumer resource model. In Fig. 2 (m)-(o), Fig. 3 (f)-(i), and Fig. 4 (b)-(e) with legend "Assembled", we simulated an ensemble of 640 microbial consumer resource models (MiCRM). We assembled each system's pre-invasion community with 60 consumers and $n = 30$ resources, which resulted in an average of 8 surviving consumers, while resources were abiotic and did not go extinct. It had a cross-feeding structure where we supplied one of the resources at a high rate $k = 1000$, while others were by-products. Similar to before, the consumer preference matrix was

$$\langle c_{i\alpha}\rangle = \frac{\mu_c}{n}, \langle c_{i\alpha}c_{j\beta}\rangle = \delta_{i\alpha}\delta_{j\beta}\frac{\sigma_c^2}{n} \quad [98]$$

with $\mu_c = 10$, $\sigma_c = 3$, the leakage rate was a constant $l = 0.8$ across consumers. We chose the sparsity parameter for cross-feeding matrix to be 0.2. We fixed extinction threshold abundance for consumers at 0.1. We performed the simulation with community simulator, which used an Expectation-Maximization algorithm that utilized a constraint optimization algorithm to find the ecological steady states.

D. Eco-evolutionary dynamics of MiCRM. In Fig. 4 (b)-(h), we performed a series of simulations for eco-evolutionary dynamics for 40 successful invasions on an ensemble of 640 microbial consumer resource models. We first initialized the communities with parameters whose statistics were identical to communities from the previous section on the microbial consumer resource model (MiCRM); we removed all extinct species at an ecological steady state. Then, at each iteration, we randomly selected a parent from the surviving consumers, and sampled a mutant as an invader with consumer preference vector $\vec{c}_{\text{mutant}}$ correlated with a randomly selected parent $\vec{c}_{\text{parent}}$ according to the equation

$$\vec{c}_{\text{mutant}} = \mu_c + \rho(\vec{c}_{\text{parent}} - \mu_c) + \sqrt{1-\rho^2}\vec{z}_c \quad [99]$$

where ρ is the correlation between mutant and parent traits, and z_c is a random vector composed of uncorrelated Gaussian variables with zero mean and standard deviation $\sigma_c/\sqrt{n}$. For our analysis, we used $\mu_c = 10$ and $\rho = 0.8$. After we sampled the invader and added it to the community, we solved for the new ecological steady-state using the iterative optimization algorithm common to all MiCRM simulations. We removed extinct species if they were below a threshold abundance of 0.1. We repeated this iterative eco-evolution until we observed 40 successful invasions in each community.

In Fig. 4(b), we applied hierarchical clustering to the consumer preference matrix to group consumers and resources based on their pairwise correlations. We computed Pearson correlation coefficients for rows (consumers) and columns (resources) separately, and performed clustering using the average linkage method. We reordered the matrix based on the resulting dendrograms to highlight correlation of similar consumers and resources. We visualized the reordered matrix as a heatmap with interaction strengths represented by color intensity.

In Fig. 4(c), we calculated the average extinction probability versus consumer abundance by grouping the consumers into bins according to their pre-invasion abundance and computing the average extinction probability for consumers within each bin. The bins are grouped with approximately the same number of samples per bin. Number of bins in this plot is chosen to be 13, with around 350 samples per bin for the evolved systems, and around 460 for the assembled systems. The error bars represent the standard error for both axes, calculated for each bin.

E. Consumer resource model with (Monod) Type-II functional response . In Fig. 2 (i)–(k), and Fig. S3(e)–(h), we simulated an ensemble of 640 consumer-resource models employing Monod (Type-II) functional responses. The pre-invasion communities were sampled from a regional species pool consisting of 30 consumers, with the number of resources set to $n = 30$. The Monod constant h was chosen as $h = 3$, comparable in magnitude to the resource abundances, thereby introducing a moderate degree of nonlinearity. Consumer preferences were sampled similarly to other models, with parameters $\mu_c = 3$ and $\sigma_c = 1$. Environmental variables were held constant across species, with values $k = 3$ and $u = 1$.

In Fig. 5, we varied h to sample models with different non-linearity. We choose 10 values of h linearly from $h = 3$ to $h = 10$. For each value we sampled 100 systems. Other parameters are kept the same as before. All simulations for this model is performed with numerical ODE simulator 'LSODA', with parameters atol=1e-11, rtol=1e-11, and we checked convergence by ensuring the maximum derivative was less than 1e-10.

12. Methods for Data Analysis and Visualization

In Fig. 5, we used species abundance data obtained from three different experimental works from Kuebbing et al.(3), Rakowski and Cardinal (4), and Pennekamp et al. (5). In addition, we used the Lotka-Volterra interaction matrix fitted for these datasets in a previous work (6).

We repeated the same set of procedures for these 3 different datasets as follows: (1) we first took the median abundance of each species in the largest ecosystem (containing all species) as the true values corresponding to new community abundance X_S^{new} and invader abundance X_0, plotted on the y-axis; (2) then we considered the abundance of smaller ecosystems (assemblages) with exactly one fewer species than the largest system and treated the missing species as an invader to those smaller systems as X_S^{old}; (3) Using our framework described by Eq. (2)-(7) in main text, we predicted X_S^{new} and X_0. Since for each dataset, researchers repeated the same experiment with the same community composition (assemblages) 5 to 10 times, we had a prediction for each of these replicas. In Fig. 5, we plotted the median values on x-axis, and used the 25th and 75th percentiles as error bars.

13. Additional experimental data

To further assess the applicability of our framework to experimental systems, we present results for three additional sets of biodiversity-ecosystem functioning (BEF) experimental data (7–9). We used Lotka-Volterra interaction matrices fitted in prevoiusly work (10), where phylogenetic relationships are assumed to shape species competition. Among the four models fitted with similar performance, we selected Model 3, which provides an intermediate level of complexity. Each dataset comprised species assemblages from distinct ecosystems. Compared to the datasets analyzed in the main text, we found two key differences. First, most assemblages lacked replicates, preventing us from estimating variability or error bars. Second, the datasets did not include exhaustive combinations of all possible species; consequently, pairs of assemblages used for comparison may differ by multiple species, with several considered as invaders or extinctions.

To align with the assumptions of our framework, we filtered pairs of assemblages based on "similarity", defined as the fraction of shared species relative to the total in either assemblage. In Fig. S7, to enable comparable analyses to the main text, we first limited to systems with a clearly defined single invader, then applied different similarity thresholds to show a comparable number of invasion results on the plots. For (a), covering 9 years of the Biodiversity II experiment (7) and 9 years of the Wageningen experiment (8), a similarity threshold of 0.85 was applied. For (c), corresponding to the Cadotte 2013 dataset (9), we used a threshold of 0.6. Additionally, in Fig. S8, we plotted all pairs of assemblages with a similarity above 0.5, coloring points by similarity value. There is a clear trend showing that the similarity value positively correlates with the prediction performance, meaning that pairs of assemblages with higher species overlap tend to yield more accurate predictions within our framework.

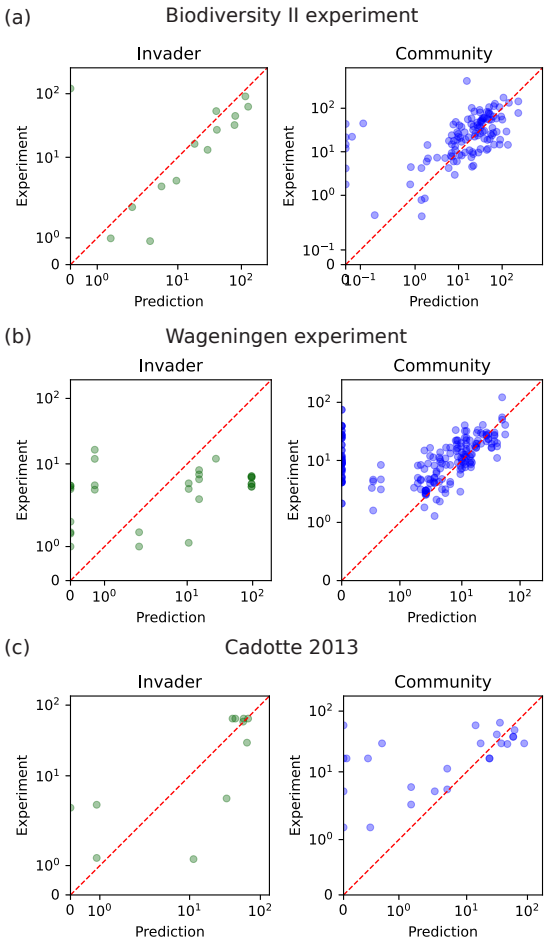

Fig. S7. Theory applicable to bio-diversity ecosystem and functioning (BEF) experiments. Scatter plots comparing experimentally measured abundances with theoretical predictions for three datasets of assemblages of plant ecosystems. We considered pairs of assemblages which differed from each other by the introduction of exactly one species which we call the invader, and variable numbers of extinctions events. Using our theory for the GLV model on these datasets, we predict both the invader abundance (left) and post-invasion community abundances (right). There is no replicates in general and therefore no error bars. Both axes use a sym-log scale, with linear scaling applied below the smallest non-zero data point. The datasets are all plant communities from different experiments, filtered to pairs of assemblages with high species similarity (defined as ratio of shared species in two assemblages) for visual clarity, (a) 9 years of Biodiversity II experiment (7) with similarity larger than 0.85, (b) 9 years of Wageningen experiment(8) with similarity larger than 0.85 and (c) one dataset Cadotte 2013 with similarity larger than 0.6. See Fig. S8 for all pairs of assemblages with similarity larger than 0.5 (9).

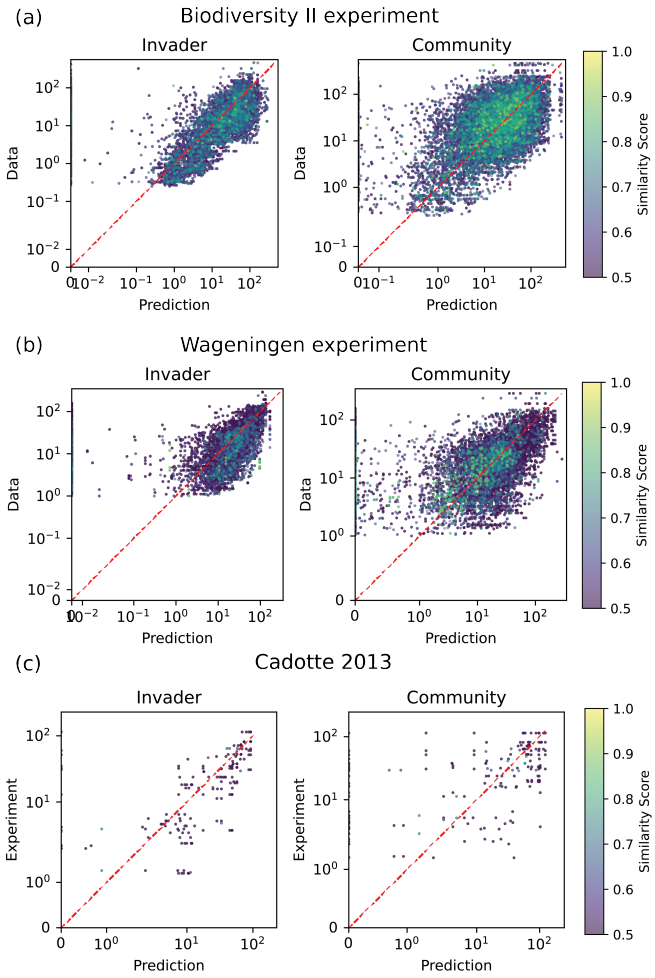

Fig. S8. Extended version of Figure S7 with variable number of invaders and number of extinction events. The marker color indicates the similarity of species between the pairs of assemblages, defined as the number of shared species divided by the total number of species in these assemblages. Only assemblages with more than half of shared (similarity> 0.5) species are included. The plots are (a) 9 years of Biodiversity II experiment (7) with (b) 9 years of Wageningen experiment (8), and (c) one dataset Cadotte 2013 (9).

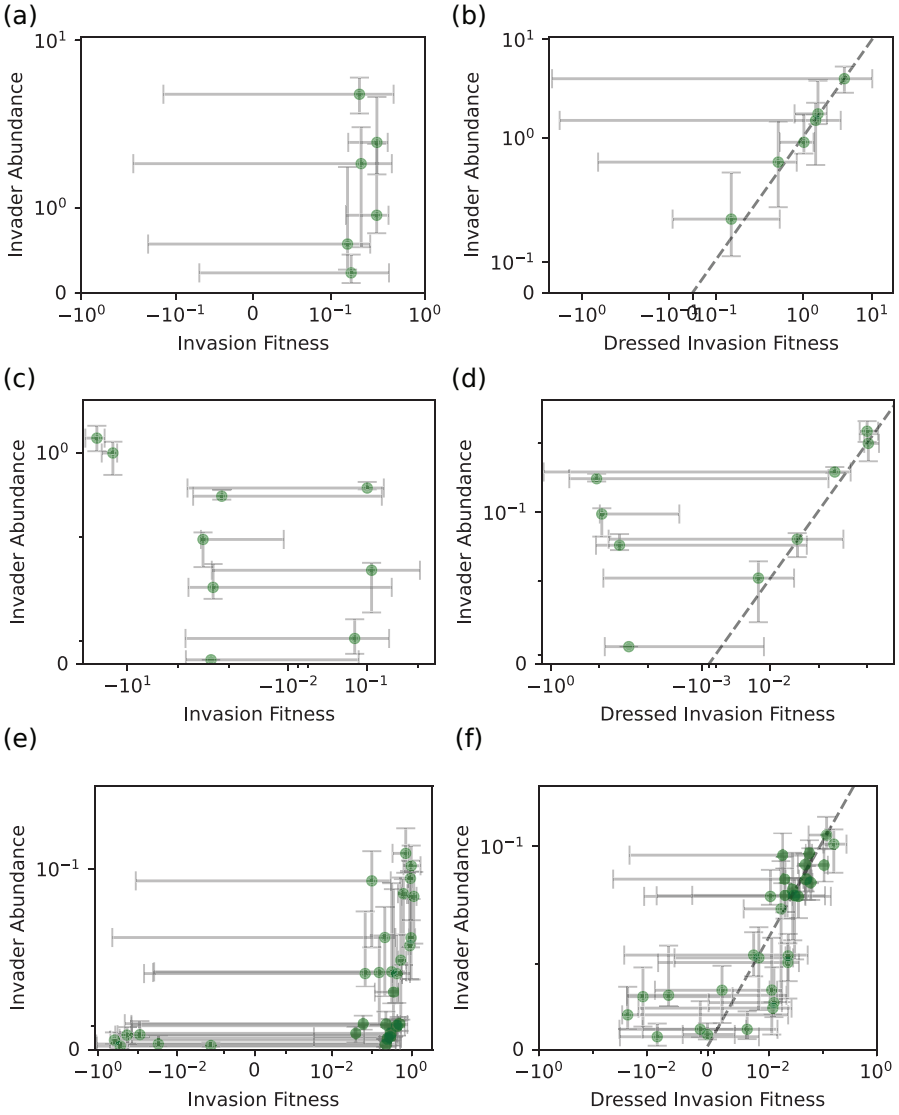

Fig. S9. Dressed invasion fitness correlates more strongly with invader abundance in experimental data. Scatter plots comparing experimentally measured invader abundances with invasion fitness and dressed invasion fitness, which are both computed from the fitted LV models and the pre-invasion abundance. Points indicate median value for measurements, with error bars indicating variability across replicates. Both axes use a sym-log scale, with linear scaling applied below the smallest non-zero data point. The datasets span (a)–(b) plant communities (3), (c)–(d) herbivore–algae communities (4), and (e)–(f) ciliated protist communities (5).

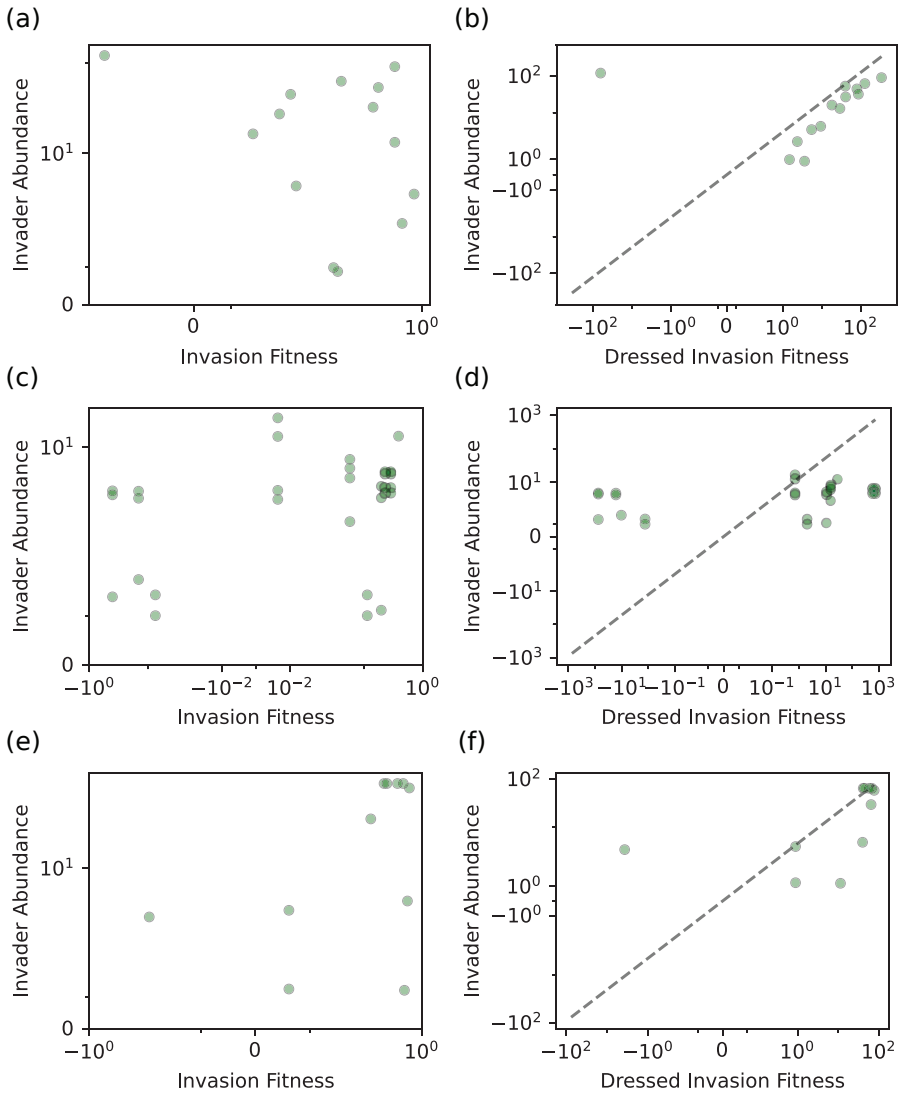

Fig. S10. Additional data of dressed invasion fitness correlates more strongly with invader abundance in experimental data. Comparison of the relationship between invader abundance and both dressed invasion fitness and invasion fitness, shown in analogous plots to Fig. S9, for additional plant experiments: (a)–(b) nine years of the Biodiversity II experiment (7) with similarity larger than 0.85, (c)–(d) nine years of the Wageningen experiment (8) with similarity larger than 0.85, and (e)–(f) one dataset from Cadotte 2013 with similarity larger than 0.6. See Fig. S8 for all pairs of assemblages with similarity larger than 0.5 (9).

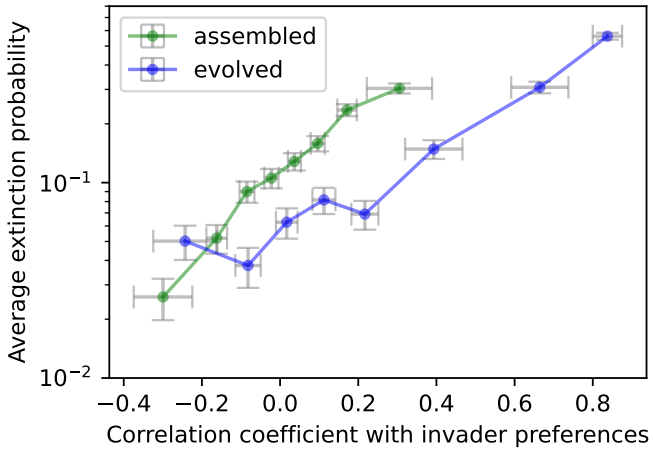

Fig. S11. Extinction probability increases with invader-resident correlation, consistent with competition exclusion. We used the same simulation results from the MiCRM model in Fig. 4 to plot how the extinction probability of a resident species depends on its phenotypic similarity with an invader. Similarity is measured using correlation between consumption preferences of residents and invaders. For evolved communities, the highest correlation would be between parents (residents) and their mutants (invaders). The plot shows that resident species are indeed more likely to become extinct when their consumer preferences are more correlated with an invader, consistent with competitive exclusion (10, 11). However, note that even at high correlation, extinction probability is still $\sim 60\%$. This indicates that in roughly 40% of cases, invaders can still coexist with highly similar species as long as they are embedded in a diverse community.

References

1. G Bunin, Ecological communities with lotka-volterra dynamics. *Phys. Rev. E* **95**, 042414 (2017).
2. R Marsland, W Cui, P Mehta, The minimum environmental perturbation principle: A new perspective on niche theory. *Am. Nat.* **196**, 291–305 (2020).
3. SE Kuebbing, AT Classen, NJ Sanders, D Simberloff, Above-and below-ground effects of plant diversity depend on species origin: an experimental test with multiple invaders. *New Phytol.* **208**, 727–735 (2015).
4. C Rakowski, BJ Cardinale, Herbivores control effects of algal species richness on community biomass and stability in a laboratory microcosm experiment. *Oikos* **125**, 1627–1635 (2016).
5. F Pennekamp, et al., Biodiversity increases and decreases ecosystem stability. *Nature* **563**, 109–112 (2018).
6. DS Maynard, ZR Miller, S Allesina, Predicting coexistence in experimental ecological communities. *Nat. ecology & evolution* **4**, 91–100 (2020).
7. D Tilman, et al., Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
8. J Van Ruijven, F Berendse, Diversity enhances community recovery, but not resistance, after drought. *J. Ecol.* **98**, 81–86 (2010).
9. MW Cadotte, Experimental evidence that evolutionarily diverse assemblages result in higher productivity. *Proc. Natl. Acad. Sci.* **110**, 8996–9000 (2013).
10. P Lemos-Costa, ZR Miller, S Allesina, Phylogeny structures species' interactions in experimental ecological communities. *Ecol. Lett.* **27**, e14490 (2024).
11. M Sireci, MA Muñoz, J Grilli, Environmental fluctuations explain the universal decay of species-abundance correlations with phylogenetic distance. *Proc. Natl. Acad. Sci.* **120**, e2217144120 (2023).