Artificial selection of communities drives the emergence of structured interactions

Jules Fraboul,[1,*] Giulio Biroli,[1] and Silvia De Monte[2,3]

[1] *Laboratoire de Physique de l'École normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France.*

[2] *Institut de Biologie de l'ENS (IBENS), Département de Biologie, Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France.*

[3] *Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany.*

(Dated: December 13, 2021)

Species-rich communities, such as the microbiota or environmental microbial assemblages, provide key functions for human health and ecological resilience. Increasing effort is being dedicated to design experimental protocols for selecting community-level functions of interest. These experiments typically involve selection acting on populations of communities, each of which is composed of multiple species. Numerical explorations allowed to link the evolutionary dynamics to the multiple parameters involved in this complex, multi-scale evolutionary process. However, a comprehensive theoretical understanding of artificial selection of communities is still lacking. Here, we propose a general model for the evolutionary dynamics of species-rich communities, each described by disordered generalized Lotka-Volterra equations, that we study analytically and by numerical simulations. Our results reveal that a generic response to selection for larger total community abundance is the emergence of an isolated eigenvalue of the interaction matrix that can be understood as an effective cross-feeding term. In this way, selection imprints a structure on the community, which results in a global increase of both the level of mutualism and the diversity of interactions. Our approach moreover allows to disentangle the role of intraspecific competition, interspecific interactions symmetry and number of selected communities in the evolutionary process, and can thus be used as a guidance in optimizing artificial selection protocols.

Several ecosystem services and functions are emergent properties of biological communities formed by many different species [1]. Particularly interesting in this respect are microbial communities – such as the microbiota, plankton and soil communities – that dispense highly relevant functions, spanning from human health [2] to the Earth biogeochemical cycles [3]. Exploitation of microbial diversity for providing services, such as detoxification, bioremediation, or efficient production of a compound of interest, is moreover an open challenge for both evolutionary biology and engineering [4, 5]. *Artificial selection* appears to be the most promising avenue to steer microbial communities by repeatedly screening for a target collective function [5]. Experimentally, this principle has been applied to evolve plant-associated communities that provide increased plant biomass [6], and rhizosphere microbiomes improving leaf greenness [7] or salt-tolerance [8]. The widespread application of such approach is nonetheless hampered by the large number of parameters that have potential bearings on the efficiency of the selection protocol, and that must be evaluated in designing these highly demanding experiments. Simplified models have been therefore used as a guide to identify general principles underpinning community evolution and to predict the effect of specific experimental choices [4, 9–13].

Both experiments and models demonstrated that artificial selection operating at the community level can give rise to evolved communities performing that function better than the ancestor. However, they also reveal that changes in collective functions are associated to a myriad of possible rearrangements of species-level features, making it extremely difficult to draw general conclusions on the effect of community selection on the ecology of the underlying community. This calls for general approaches able to tackle the challenging problem of community selection. Results have been achieved for two-species communities, indicating that selection for collective properties leads to directional changes in the strength of species interactions, and to evolved communities which have specific ecological features [14–16]. However, most real-world microbial ecosystems are species-rich, and nested models for community evolution need a large number of parameters to encompass such diversity. This feature makes the problem very complex. Even a purely numerical approach is limited by the impossibility to perform exhaustive explorations, and thus to assess the degree of generality of the conclusions.

One way to tackle this challenge is the so-called 'disorder approach', which has been successfully used in many complex systems. Pioneered by Wigner [17] in physics, it was extended by May to study the ecology of species-rich communities [18, 19]. It consists in assuming that interspecific interactions are randomly distributed. Several recent works used this framework to study the ecological dynamics of communities described by generalized Lotka-Volterra equations and by McArthur equations [9, 11, 20–34].

Here, we extend this approach to study the evolutionary dynamics of species-rich communities undergoing artificial selection for increased total abundance. This variable of ecological relevance is directly related to ecosystem biomass and export, and to primary production. Its maximization can be achieved by different routes, so that predicting the outcome of the evolutionary process, and in particular the composition of the community, is a challenge. Different from previous works [10, 35], where the source of variation among communities was mutations in interaction parameters of single species, we consider 'community-level' mutations that affect the whole interaction matrix at community reproduction. This choice allows us to derive an analytical description for how interactions evolve while community-level properties change smoothly along an evolutionary trajectory. With such equations, we address in general terms what changes in community structure are associated to the evolution of collective function, what is the pace of community evolution, and how

* jules.fraboul@phys.ens.fr

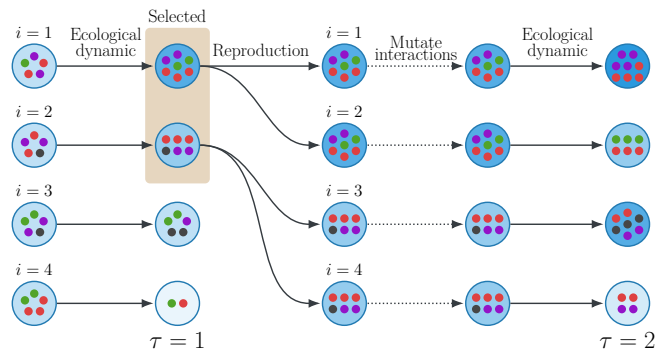

Figure 1. **Structure of the model for the evolution of species-rich communities.** Each community i in a population of n (here, $n = 4$) communities is represented by a circle and is composed of a set of individuals (represented by the dots), belonging to different species (represented by colour), initially sampled by a same metacommunity. The intensity of the circle shade indicates the value of a community-level function (for instance the total abundance). The $m = 2$ communities that maximize that function are selected for reproduction. Newborn communities are generated by copying the state (vector of species abundances), but modifying the parameters of the interactions among species, as detailed in the text. In the course of a community generation, these changes can result in ecological variations of community composition and of its selected function.

these depend on the selection protocol. Our main result is that selection for total abundance leads to the emergence of a collective term in the interaction matrix (corresponding to an effective cross-feeding effect), and that is associated to an isolated eigenvalue. This leads to a global increase of mutualism, diversity and total abundance – far beyond those achievable with fully disordered interactions.

RESULTS

Model for selection of species-rich communities

We model a population of n communities that undergo cycles of ecological growth, selection and reproduction, corresponding to the classic protocol used in experimental evolution [4, 9, 11], as explained in Fig. 1. The ecological dynamics of species is described as a function of a continuous time variable t. Reproduction occurs via monoparental seeding of the next community generation ('propagule' reproduction [9, 11]). The successive community generations (also simply referred to as 'generations') are indexed with a discrete variable τ. Selection is applied at the community level by letting the probability that a community reproduce depend upon a collective function, typically evaluated at $t = T$, with T the duration of one generation. The evolutionary dynamics that we aim to describe consists in the change of the community composition – and of the corresponding function – across multiple generations. For simplicity, we assume that mutations only occur in newborn communities, so that within one collective generation species abundances are only ruled by the ecological dynamics. In the following, we first detail the model for the ecological dynamics within one community generation, and then the rules for community reproduction and mutation.

Ecological dynamics. Each community is composed of S species whose abundances $(N_i)_{i=1,\dots,S}$ are described by continuous variables. The ecological dynamic is described by the generalized Lotka-Volterra equations, a classic model for studying species-rich communities [36]. The variation in time of the abundance of species $i \in 1, \dots, S$ in one community is:

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = \frac{N_i}{K_i}\left(K_i - N_i - \sum_{j\neq i} \alpha_{ij} N_j\right), \tag{1}$$

where the constants (K_i) are the carrying capacities (the maximum abundances the species achieve in isolation) and the interaction coefficients α_{ij} of the matrix α represent the effect of species j on the growth of species i.

Species interactions. Following May's disorder approach [18], we choose initial communities with random interactions. Specifically, the coefficients α_{ij} are drawn, as in [22], from a multi-variate normal distribution of parameters:

$$\begin{aligned} \mathbb{E}(\alpha_{ij}) &= \mu/S \\ \mathrm{Var}(\alpha_{ij}) &= \sigma^2/S \\ \mathrm{Corr}(\alpha_{ij}, \alpha_{ji}) &= \gamma. \end{aligned} \tag{2}$$

Here, μ represents the total interaction strength faced by one species from all of its partners, whereas σ measures the diversity of interactions. The scaling of the moments with $\frac{1}{S}$ is chosen so that the moments of the distribution do not diverge or vanish for communities with many species[1]. The parameter $\gamma \in [-1, 1]$ determines the symmetry of the ecological interactions. Competition and mutualism are characterized by symmetric effects on growth of the interacting species, corresponding to $\gamma = 1$. In this case, as $\alpha_{ij} < 0$ accounts for a beneficial interaction, a negative μ indicates prevalence of mutualism, whereas a positive one indicates prevalence of competition. On the opposite, exploitative interactions like predator-prey and parasitic interactions are characterized by parameters whose deviations from the mean μ are skew-symmetric, corresponding to $\gamma = -1$. As a start, we consider $\gamma = 0$, for which α_{ij} and α_{ji} are uncorrelated random variables, and discuss generalizations to arbitrary values later. In the limit of large number S of species, the asymptotic ecological dynamics of such complex communities depends only on the statistics μ, σ and γ of the interaction strength distribution [22, 24, 32].

Composition of newborn communities. We initialize communities by sampling interactions from a Gaussian distribution in the region where the system has a unique, stable coexistence equilibrium (see Section 1 of SI). Such equilibrium community is a global attractor that is independent of the initial community composition, as long as all species are present [22]. Nonetheless, the transient dynamic leading to such an attractor depends on the initial state of the community, and can have important evolutionary implications [11]. In order to avoid such long-term effects of ecological transients, generations are chosen sufficiently long for the community to approach equilibrium. In order to speed up simulations, newborn communities are initialized with the same composition

[1] The results we present below are robust with respect to other choices of the probability law of the α_{ij}'s, as long as their first and their second moments exist.

N_i of their parent community. This way, the equilibrium position is attained very quickly, and evolution changes adiabatically its position. Numerical simulations where at each generation all species were introduced at low abundance (not shown) indicate that this assumption does not qualitatively alter the results.

Community-level selection. In order to study the effect of selection for community-level properties, we rank communities according to a single scalar community-level property, the total abundance at the end of the growth phase $N_T = \sum_i N_i(T)$. The m best-ranking communities are chosen to seed the following generation, and the rest discarded (Fig.1). In our simulations, newborn communities are all copies of the single best community ($m = 1$), whose offspring substitute all other less performing communities. A non-exhaustive numerical exploration of cases when $m > 1$ suggests that the qualitative results illustrated later are robust to this parameter.

Community-level mutations. When an offspring community is born, it acquires the same composition (both in terms of types and abundance of species) of the parent community. In the absence of variation in the community parameters, this guarantees that community functions are perfectly inherited. For evolution by natural *and artificial* selection to occur at the level of communities, however, there must be variation in the collective functions [37]. In our model, variation is replenished at each community generation by changes in the interaction matrix that we call 'community-level mutations'. Such mutations affect the statistics of the interaction matrix, that are the only determinants of collective function in large communities [22, 24, 32].

One fundamental property of mutations (at any level of organization) is that they should provide an unbiased variation of the trait under selection. To this purpose, we write the interaction matrix at generation τ as:

$$\alpha_{ij}(\tau) = \text{mean}[\alpha(\tau)] + \text{std}[\alpha(\tau)]\, b_{ij}(\tau) \tag{3}$$

where:

$$\text{mean}[\alpha] = \frac{1}{S^2}\sum_{ij}\alpha_{ij}$$
$$\text{std}[\alpha] = \sqrt{\frac{1}{S^2}\sum_{ij}(\alpha_{ij} - \text{mean}[\alpha])^2}$$

are the empirical mean and standard deviation of the matrix α, and the reduced matrix b has empirical mean 0 and empirical variance 1.

We define the mutated interaction matrix as:

$$\alpha_{ij}(\tau+1) = \text{mean}[\alpha(\tau)] + \text{std}[\alpha(\tau)]\, \hat{b}_{ij}(\tau) \tag{4}$$

with:

$$\hat{b}_{ij}(\tau) = \frac{b_{ij}(\tau) + \varepsilon\eta_{ij}(\tau)}{\sqrt{1+\varepsilon^2}}, \tag{5}$$

where $\eta(\tau)$ is a Gaussian random matrix of expected value 0, variance 1 and symmetric correlation γ. The reduced matrices at two successive generations have the same expectations. Community mutations therefore statistically preserve the mean and the variance of the interaction matrix, so that total abundance is on average constant. In the absence of selection, thus, community function will not improve. Statistical fluctuations at finite S, however, make the empirical mean and variance of η different from zero and one, respectively. Hence, newborn communities have a range of differences in their interactions, providing the variation in community function selection acts upon. The variation of the selected function corresponding to such a mutation of the interaction matrix is also of order ε. Considering small mutational steps ε, we will be able to describe the evolutionary dynamics as a quasi-continuous process.

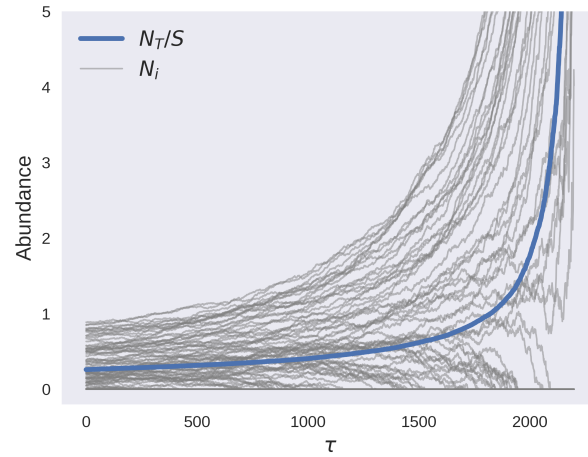

Figure 2. **Changes of species abundance along an evolutionary trajectory.** Selection for increased total abundance leads to an increase in the abundances of most species (**grey lines**), and, as a consequence, of the average abundance $\sum_i N_i/S$ (**blue line**), that is proportional to the selected function N_T. See Material and Methods for the details of the numerical simulations.

Selection for larger total abundance favors mutualism and imprints a structure on the interaction matrix

We now present and discuss the salient features of the evolutionary dynamics obtained by numerically simulating the model previously introduced. We illustrate with a specific example the general behavior later addressed analytically. The simulation algorithm and parameters are detailed in the Materials and Methods.

Community response to selection. Unsurprisingly, and as previously observed when selection was imposed on collective functions [9, 10, 38, 39], communities evolve so that their total abundance gradually increases (Fig. 2). Interestingly, although this goal could be achieved by just one species, the ecosystem systematically responds as a whole with a simultaneous raise in the abundance of a large fraction of species, while the rest goes extinct. As a consequence, the diversity of the community ϕ, measured as the number of co-existing species divided by the initial number of species, decreases, but does not collapse (Fig. 7 of SI). Such increase accelerates on longer time scales, and the ecological dynamics is pushed in a region where the system eventually diverges.

Evolution of the statistics of interactions. Additional insights on how selection affects the community can be gained by analyzing the evolution of the statistics of the interaction matrix α such as its scaled mean $\mu(\tau) = \text{mean}[\alpha(\tau)]/S$ and standard deviation $\sigma(\tau) = \text{std}[\alpha(\tau)]/\sqrt{S}$.

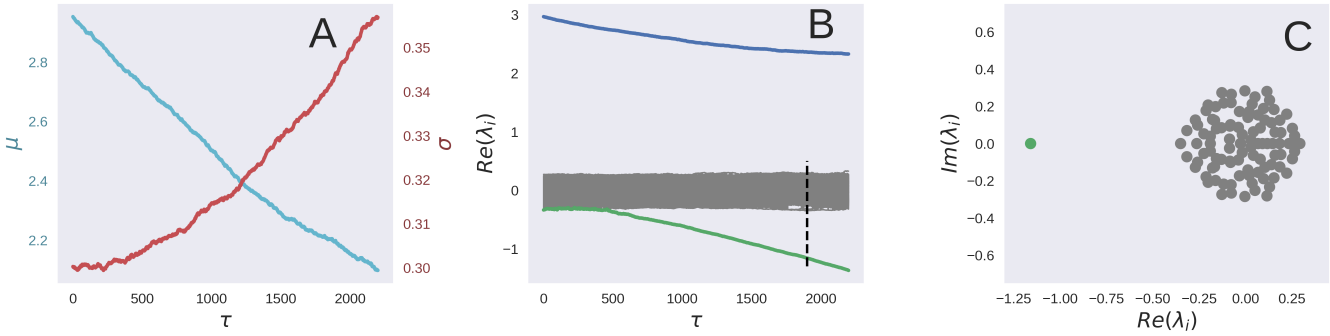

Figure 3. **Changes of the statistics of the interaction matrix along an evolutionary trajectory.** The interaction matrix α of the best community evolves so that the average interaction strength decreases linearly in time (A, cyan), while its variance increases (A, red). Such changes correspond to a modification in structure, manifest in the spectrum of its eigenvalues. The change of their real part across community generations (B) reveals the appearance of an isolated negative real eigenvalue (green). A zoom of the spectrum in the complex plane (C) at generation $\tau = 1900$ represented by the dotted line in (B) reveals that, apart from the emergence of this mutualistic collective mode, the matrix retains its initial random structure characterized by a circular law for the eigenvalues.

As shown in Fig. 3 A, the mean decreases, indicating that interactions become – on average – progressively more mutualistic. At the same time, their variance increases, so that diversity of interactions within the community is enhanced. The results obtained for disordered communities with an infinite number of species are helpful to rationalize these findings. In fact, as shown in [22], their total population size N_T is a function of μ and σ alone. Thus, one could envision selection as a process in which the empirical moments of α change across community generations so as to climb the gradient of the fitness function $N_T(\mu, \sigma)$. This explanation, however, falls short of explaining quantitatively the numerical results, as we show in Fig. 4 by plotting $N_T(\mu, \sigma)$ for purely random communities together with the evolutionary trajectory of Fig. 2. Indeed, the evolutionary dynamics deviates from the gradient-climbing process predicted for a disordered matrix with the same moments as those observed along the trajectory. This clearly shows that evolution shapes the interaction matrix α and selection makes it deviate from its initial random form (eq. 2). By imprinting a structure on α, selection produces communities characterized by higher values of the total abundance than those obtainable for purely disordered interaction matrices.

Emergent community structure. In order to understand the emergent structure of the interaction matrix, we study its eigenvalues. The spectrum of the random initial interaction matrix is, in the complex plane, a circle of radius σ centered in the origin, corresponding to the random component of the interactions [40], plus an isolated positive eigenvalue (blue in Fig. 3 B) of value μ, which reflects the average strength of competitive interactions. The initial effect of selection is to reduce this value. After some time, however, an isolated *negative* eigenvalue λ (green in Fig. 3 B and C) emerges from the circle and detaches from it linearly in time. Numerical simulations indicate that it is a general feature of the model, as it occurs for all parameter sets we explored, including different values of γ – with the notable exception of $\gamma = -1$, that we discuss later. Apart from this isolated component, the matrix remains fully random. In fact, the circle of eigenvalues doesn't change much during evolution. This phenomenon is strongly reminiscent of the so-called BBP transition [41, 42] studied in the context of signal processing and random matrix theory. In that case, by increasing the signal-to-noise ratio the signal

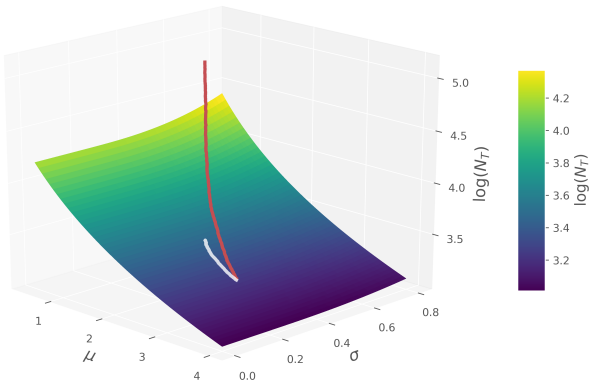

Figure 4. **Purely random interactions cannot explain the evolution of total community abundance.** Variation of the interaction moments $\mu(\tau)$, $\sigma(\tau)$, and of the total abundance $\log(N_T(\tau))$ (red line) along an evolutionary trajectory. The abundance of a random interaction matrix (2) with moments μ, σ (surface) is plotted for comparison. The white line is the predicted total abundance if the matrix of moments $\mu(\tau)$, $\sigma(\tau)$ was completely random, indicating that along the trajectory the matrix becomes progressively structured.

leads to an isolated eigenvalue in the noisy correlation matrix. In our case, selection adds to the random part a new rank-one term which can be written as $\lambda \mathbf{q}^T \mathbf{r}$, where q and r are the left and right eigenvectors relative to the isolated eigenvalue λ.

The structure that evolution imprints on the interaction matrix can be visualized by displaying the entries of α along the evolutionary trajectory presented in Figs. 2 and 3. Fig. 5 represents the interaction matrix for early and late stages of community evolution. Species are ordered in terms of their carrying capacity from larger to smaller. At the beginning of evolution, there is no visible structure in the interaction matrix, except the diagonal that has zero entries. After 2000 generations, it is clear that species who have become more mutualistic, thus have more negative interaction coefficients, are mostly those that initially had higher carrying capacity. This is a direct manifestation of the emergence of the isolated

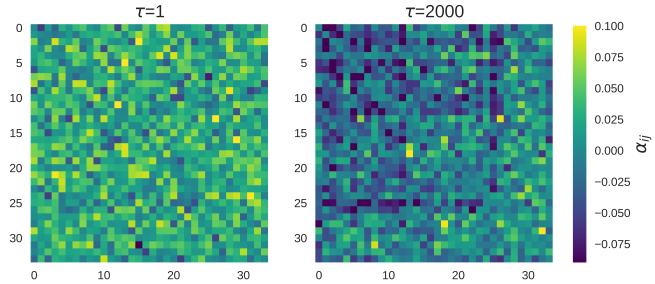

Figure 5. **Evolution of the interaction matrix.** Coefficients of the interaction matrix α with rows and columns sorted by decreasing carrying capacities at generations 1 (left) and 2000 (right) for the same simulation as Fig. 2. Only the species that have positive abundance at generation 2000 are shown.

eigenvalue, as we find that its eigenvectors are correlated to both $\mathbf{K}$ and to the equilibrium abundance. Aside from this structure, the matrix remains random. Indeed, subtracting $\lambda \mathbf{q}^T \mathbf{r}$ from α, we retrieve a purely random matrix, indicating that selection has little effect on the random component of the spectrum.

In summary, selecting ecosystems with larger total abundance leads the initial purely random interaction matrix α_{ij} to be summed up with an emergent term that makes interactions globally more mutualistic. An increase in total abundance is thus achieved by species developing collective positive interactions, rather than increasing the abundance of one single member of the community. Simulations for a number of different parameter values and for other target functions (Section 10 of SI) suggest that the mechanism illustrated above, by which the interaction matrix evolves into a random part plus a rank one perturbation is generic.

Analytical description of the evolutionary trajectory

The generality of the evolutionary dynamics observed in numerical simulations can be addressed by developing an appropriate analytical description. In this section we provide equations for the evolution of the interaction matrix. The emergence of structure can then be understood in a simple case where these equation are solvable exactly. We sketch here the derivation of difference equations for the total abundance and the interaction matrix in the case $\gamma = 0$, and point the reader to the Material and Methods and the SI for details on the general derivation.

Given a community with interaction matrix $\alpha(\tau)$ and equilibrium $\mathbf{N}(\tau)$ at a given generation τ, we aim to identify what will be the interaction matrix $\alpha(\tau+1)$ of the selected offspring community. Such matrix is the one giving the largest possible total abundance at equilibrium. The vector of abundances at equilibrium depends both on the interspecific and intraspecific interactions and is the solution of $\mathbf{N}^\star(\tau) = (\mathbb{I}^\star + \alpha^\star(\tau))^{-1}\mathbf{K}^\star$, where $\alpha^\star(\tau)$ and $\mathbf{N}^\star$ are the interaction matrix and the vector of abundances reduced to extant species (the same notation is used for other matrices and vectors below). In order to obtain the dynamical equation for the evolution of the interaction matrix we proceed as follows: mutations are, for $\epsilon \ll 1$, equivalent to small random perturbations of the carrying capacities, so we use linear response theory to obtain the change they induce in the equilibrium abundances. The total abundance of each of the n communities is therefore modified by a random contribution that we can fully characterize. The selection process, then, singles out the largest contribution, i.e. the largest among several independent random variables, a problem that we solve using extreme value statistics [43]. The outcome of the computation is the equation for the evolution of the total abundance $N_T(\tau)$ across one collective generation:

$$N_T(\tau+1) = N_T(\tau) + M_n(\tau)\frac{\varepsilon\sigma(\tau)}{\sqrt{S}}\|\mathbf{v}(\tau)\|\|\mathbf{N}(\tau)\|, \qquad (6)$$

where $\mathbf{v}(\tau)$ is a nonlinear function of the interaction matrix $\alpha(\tau)$: $\mathbf{v}^\star(\tau) = (\mathbb{I}^\star + \alpha^\star(\tau)^\top)^{-1}\mathbf{1}^\star$ and is zero for extinct species. This vector is the gradient of the total abundance at equilibrium with respect to the carrying capacities: $\mathbf{v}(\tau) = \frac{\partial N_T}{\partial \mathbf{K}}(\tau)$ and represents the response of the selected community function to a perturbation of individual species' carrying capacity K_i. The random variable $M_n(\tau)$ (drawn independently at each generation) follows the statistic of the maximum value of n Gaussian variables [43], with expected value $\overline{M_n}$ (see the distribution of M_n in Fig. 2 of SI). Eq. 6 explains why evolution tends to improve the value of the target community function. Indeed, the product of the norms is always positive, and so is M_n for sufficiently large n. However, the capacity of a community to change in response to a given selective pressure can vary, as quantified by the vector $\mathbf{v}$.

The change in the interaction matrix across one collective generation can be in turn decomposed in a directional term – contributing to the evolution of N_T – and its complement B_{ij}, that acts as a random fluctuation[2]. The interactions thus evolve according to:

$$\alpha_{ij}(\tau+1) = \alpha_{ij}(\tau) - \frac{\varepsilon\sigma(\tau)}{\sqrt{S}}\left(M_n(\tau)\frac{v_i(\tau)}{\|\mathbf{v}(\tau)\|}\frac{N_j(\tau)}{\|\mathbf{N}(\tau)\|} + B_{ij}\right). \qquad (7)$$

This expression has a simple interpretation: among the Gaussian random mutations of the interaction matrix, only those in the special direction associated to the function N_T matter. The selected community is hence the one having the largest random Gaussian contribution associated to the first term (thus explaining the statistics of $M_n(\tau)$). For $\gamma \neq 0$, Eq. 7 generalizes to Eq. 22 of SI. The two equations only differ for the matrix component $v_i N_j$ being replaced by a corresponding symmetric term, which preserves the correlation structure of the interaction matrix. Similarly, the noise term B has a symmetric correlation of value γ. From Eq. 7, we can understand that change in pair-wise interaction coefficients has no prescribed sign, and species are not all equivalent in the face of selection. Species whose potential variation contributes more to the function (those with larger v_i) and with larger equilibrium abundance will face a larger decrease in the interaction strength, as pointed out in Fig. 5.

Eq. 6 and Eq. 7 allow us to draw general conclusions on the effect of the system's parameters on the evolutionary dynamics. First, since the expected value $\overline{M_n}$ is an increasing

[2] Formally, this amounts to decomposing η_{ij} in two parts: one along the direction (in the space of the S^2 indices) $\frac{v_i(\tau)}{\|\mathbf{v}(\tau)\|}\frac{N_j(\tau)}{\|\mathbf{N}(\tau)\|}$, and a remainder $B_{ij} = \eta_{ij} - \frac{v_i(\tau)}{\|\mathbf{v}(\tau)\|}\frac{N_j(\tau)}{\|\mathbf{N}(\tau)\|}\sum_{k,l}\frac{v_k(\tau)}{\|\mathbf{v}(\tau)\|}\eta_{kl}\frac{N_l(\tau)}{\|\mathbf{N}(\tau)\|}$.

function of n, evolution occurs faster when selection screens a larger number of communities. The growth of $\overline{M_n}$ with the number of communities, however, scales as $\sqrt{\log(n)}$, which increases slowly for large n. Expanding the number of communities under selection, thus, brings diminishing returns, and for large n may be of little avail to speed up evolution. An interesting extreme case is when there is only one community and thus no selection. M_1 is then just a Gaussian random variable with zero mean. In such a 'neutral' scenario, the total abundance and the interaction matrix undergo, in the long run, unbiased changes akin to a community-level drift (see section 7 of SI). Second, the variation of the interaction matrix and the total abundance across one community generation is characterized by a time scale $dt = \varepsilon/\sqrt{S}$ that depends on the number of species S and the mutation strength ε. Evolution occurs faster in communities with a smaller number of species and for larger mutational steps.

Eq. (7), and its generalization Eq. (22) of SI, are nonlinear recursive equations that can't be exactly solved. The key dynamical features illustrated above by numerical simulation, notably the emergence of an isolated eigenvalue, are however reproduced by an analytical continuous approximation. Indeed, when ε is small and S is large, the time scale dt is small, so that the evolution of α and N_T is quasi-continuous[3]. Let us first consider that all the carrying capacities are equal ($K_i = 1$) and $\sigma(\tau)$ remains small throughout evolution (which in the large S limit occurs whenever $\sigma(0)$ is small[4]). Then, the evolution of the interaction matrix (derived in section 6 of SI) simply reads:

$$\alpha_{ij}(\tau) = \alpha_{ij}(0) + \frac{\mu(\tau) - \mu(0)}{S}, \tag{8}$$

which is exclusively determined by its mean value:

$$\mu(\tau) = \mu(0) - \tau \cdot \frac{\varepsilon\sigma}{\sqrt{S}} \overline{M_n} \sqrt{1+\gamma} \tag{9}$$

As shown in Fig. 3 A, the average interaction strength thus decreases linearly across collective generations. Interestingly, the speed of evolution has a simple dependence on the symmetry of the interaction matrix, slowing down as the matrix becomes more asymmetric. In the limit $\gamma = -1$ of perfectly asymmetric interactions, selection on total abundance is unable to modify the interaction matrix. Ecological interactions in this limit are indeed zero-sum: the increase of the abundance of one species implies the same decrease for another species, and thus has no net effect on total abundance. The variation of such abundance along the evolutionary trajectory depends, like the interaction matrix, only on the mean interaction strength:

$$\frac{N_T(\tau)}{S} = \langle N_i(\tau) \rangle = \frac{1}{1+\mu(\tau)} + \mathcal{O}(\sigma). \tag{10}$$

It increases as interactions become progressively less competitive, and diverges when $\mu \to -1$. Fig. 5 of SI shows that Eq. 8 and Eq. 9 reproduce accurately the simulated evolutionary dynamics of the interaction matrix for small values of $\sigma(0)$. In this simple case, where species are almost equivalent, the diversity ϕ remains largely unchanged (and $\alpha \sim \alpha^\star$). Concerning the eigenvalues, Eq. 8 shows that along an evolutionary trajectory the interaction matrix maintains its initial random component of order $\sigma(0)$, and is modified by the rank-one term $\mu(\tau)/S$. For large τ this contribution, associated to mutualism, leads to a left eigenvalue popping out by a BBP transition [41] as we found numerically.

When $\sigma(\tau)$ is not small and $\mathbf{K} \neq \mathbf{1}$, the effect of evolution on the structure of the interaction matrix is qualitatively similar. Eq. (7) tells us that selection for increased total abundance indeed results in the repeated addition of rank-one matrices. Now, however, the right and left eigenvectors associated to the smallest eigenvalue depend on $\mathbf{v}(\tau)$ and $\mathbf{N}(\tau)$, and thus change along the evolutionary trajectory. Therefore, the direction of evolution is only predictable across one generation. In order to have an insight into the evolutionary dynamics, we can exploit the inference method recently introduced in [44] to find the probability distribution of the interactions coefficients between extant species given that the mean μ and equilibrium abundances $\mathbf{N}$ are known. For large S, one finds (see [44] and SI):

$$\alpha^*_{ij}(\tau) = \tilde{\alpha}^*_{ij} + \frac{l_i(\tau) N_j(\tau)}{\sum_k N_k^2(\tau)} \tag{11}$$

where $l_i(\tau) = K_i - N_i(\tau) - \mu(\tau) N_T(\tau)/S$, and $\tilde{\alpha}_{ij}$ can be taken as[5] a Gaussian random matrix like Eq. (2). This synthetic matrix thus inferred from the observation of the community at a given generation τ and the true one obtained by selection turn out to have very similar spectral properties (Fig. 8 of SI). In particular, the isolated eigenvalue and the associated eigenvector are remarkably matching. This result indicates that the evolved structure is indeed the most likely one, given the observations, and that such structure is generically the superposition of a random and a rank-one component. The latter component varies along the evolutionary trajectory, underpinning the BBP-like transition and the emergence of the isolated smallest eigenvalue observed in numerical experiments.

Eq. (11), moreover, allows us to infer the structure imprinted on the interaction matrix by selection from measures which are much more accessible than individual interaction rates. Such structure is characterized by a correlation between the right and left eigenvectors associated to the isolated eigenvalue and $\mathbf{N}$, as we found in Fig. 3. The correlation of the equilibrium abundances to the carrying capacities $\mathbf{K}$ results, as evolution proceeds, in an additional irreversible effect on the community: species with smaller carrying capacity will be more likely to go extinct, thus affecting diversity to a larger extent.

DISCUSSION

This study is devoted to identifying key and general features of the evolutionary dynamics in species-rich communi-

[3] In the limit $dt \to 0$, one can write Eq. 7 as a differential equation with all random variables replaced by their mean value (because their variances go as dt and not $\sqrt{dt}$ as in stochastic differential equations).

[4] When $S \gg 1$, $\sigma(\tau)$ changes slowly with respect to $\mu(\tau)$, consistently with the scaling evidenced in the neutral regime discussed in section 7 of SI.

[5] Here we are only interested in the spectrum and the isolated eigenvalue.

ties, when selection is applied to a collective function: the total abundance. Following previous work that brought fundamental insight in community ecology [18], we considered an initial population of communities where interspecific interactions were chosen at random. These communities can evolve across successive community generations due to selection (death of a fraction of communities) and community reproduction with mutation. Mutations are here modelled by non-biased changes of the interaction matrix, whereby multiple interaction coefficients are simultaneously modified in one community generation, affecting the community composition at equilibrium. Such composition in turns determines the function upon which selection is applied. Starting from predominantly competitive communities, predicting the outcome of the selection process is challenging, as total abundance can be improved both by an increase in the abundance of a single species, or by a coordinated increase in multiple species.

We showed that the interaction matrix evolves in response to selection, and that it results generically in interspecific interactions becoming progressively less competitive. We interpret this as the evolution of facilitation, similar to what was observed in a two-species model when the sign of interactions was allowed to change [16]. If we consider the Lotka-Volterra model as a limit case of the MAcArthur equations, which describe not only species abundance, but also the resources they consume, the emergent facilitation term can be viewed as an effective, global cross-feeding. At the same time as the average strength of interspecific interactions decreases, they become more variable. Notably, the evolutionary process imprints a structure on the interaction matrix. The key to this structure is an isolated eigenvalue, which emerges as a 'collective mode' that positively impacts the abundances of all species. In the analytical description, this corresponds to an order-one perturbation of the interaction matrix, that otherwise retains its original, disordered nature. Our finding ties with similar phenomena found in other situations where interactions between degrees of freedom are adjusted dynamically to lead to a specific collective property, such as a lower ground state energy in spin-glasses and learning in neural networks [45–49].

The nonlinear Eqs. 7 for the change of the interaction matrix across one collective generation account for the interplay between the selection target (in our case, the total abundance) and the intraspecific interactions (the carrying capacities), and explain the emergence of the imprinted structure. This formulation reveals when and how selection is able to drive the ecological dynamics of the community towards achieving the target function. Moreover, it quantifies how the speed of evolution scales with the number of communities, the intensity of selection, number of species, the symmetry of interactions and the magnitude of the mutational steps. Notably, we found that the nature of ecological interactions in the community (which imposes a constrain on the random interaction matrix) affects the efficiency of selection for total abundance: evolution occurs faster when interactions are symmetric, which is the case for competitive and mutualistic interactions, whereas it is slower when interactions are more exploitative (predator-prey, host-parasite). Analyses of microbial communities in conditions where selection can be expected to occur at the community-level, such as the gut microbiota, revealed a prevalence of cross-feeding interactions [50, 51]. Whether selection for some collective function as described in our model can account for these and other facilitation patterns observed in microbial communities [52] is an interesting open question.

Dedicated laboratory experiments may allow to test this hypothesis and at the same time explore the possibility of retrieving the salient features of evolving communities without measuring every single interaction coefficient. The application of the maximum entropy method [44] to infer the emergence of the isolated eigenvalue, similar to what we have done on our simulations, seems to be a particularly promising avenue.

We chose to analyze a very idealized model in order to achieve analytical tractability. This model has two basic ingredients: there are populations at two levels of organization (species and communities), and the ecological dynamics of a community composed of a large number of species is described as a Lotka-Volterra system with random interactions. Although it can be argued that disordered models are an oversimplification of real communities, they provide null expectations for both community-level, collective properties such as total diversity or total biomass, and for the demography of individual species from which such properties arise. Moreover, the actual strength of ecological interactions is unknown in most microbial communities, though important efforts are underway towards quantification [50]. Given this state of the affairs, and that not all detailed properties of the interactions are expected to matter in shaping the general behaviors of ecosystem, the statistical approach based on random interactions appears a valuable method to formulate null, predictive models [53].

This model could nonetheless be made more realistic in several meaningful ways. Instead of modelling species interactions through direct effects, the equations for the community ecology could include explicitly the resources that are consumed or exchanged [9, 11, 54]. Given the equivalence of the Lotka-Volterra and MacArthur models when resource dynamics is much faster than the ecological one, we do not expect this extension to qualitatively affect the main results of our work as long as time scales remain sufficiently separated. However, a formulation in terms of resource consumption would connect more directly theoretical results to experiments exploring the metabolic foundations of ecological interactions in microbial communities [51, 52, 55–57]. Especially, this may guide the choice of more realistic interaction matrices, such as for instance sparse networks [58, 59], or networks with empirical biases [51].

Even maintaining the convenient simplicity of random direct interactions, the model we considered could be further explored in regimes where the perturbative approach we introduced is expected to break down. First of all, it might be that the ecological dynamics of the community is not at equilibrium. This could be due to transients pushing selection to act in a direction that does not maximize the asymptotic function [11], to stochastic demographic fluctuations [32] or to chaotic population dynamics [22, 24]. All these processes may reduce the heritability of the community function and thus alter the evolutionary trajectory.

Finally, consistent with the idea of studying evolution of communities [37], we chose mutations that would provide community-level variation in the function that is under selection. These mutations can be thought of as the result of multiple changes in species interactions that occurred during

the lifetime of a community. More detailed descriptions of how sequential species-level mutations give rise to variation of the interaction matrix at the time of community reproduction would provide additional constraints, such as trade-offs different from symmetry or bounds to the achievable parameter values. Alternative to mutations, variation between communities could be enhanced by ecological means, like migration. It has been suggested that migration, despite producing only a limited range of community variants, contributes to speeding up evolution [11]. At the same time, however, it risks of opposing the evolution of more mutualistic variants [60]. Modifications of our model [28] may address what would be the dominant process when both ecological and evolutionary variation are present.

Communities are increasingly conceived as coherent units that provide collective-level functions, to the point to be attributed the status of 'organisms' [61–64]. If this view can reflect the way ecological interactions produce a given population structure [65], it can also extend as far as identifying communities as full-fledged evolutionary units, despite the conflicts generated by their inherent heterogeneity. In the latter case, how they are 'scaffolded' by physical compartmentalization and the establishment of community-level lineages, is all-important in determining the action of natural selection at the level of communities [66, 67]. We have modelled here the protocol commonly used in experiments of artificial selection [4, 5]. Considering that the collective level is the true center of interest for this process, moreover, we described mutations only for their effect on the community-level function under selection. Nested levels of reproducing units are widespread in the hierarchy of living beings. Our results might thus be relevant whenever selection on high-level functions bestows a structure on the interaction among heterogeneous constituent units, and contribute to understanding how integration across levels of organization is achieved.

MATERIALS AND METHODS

Description of the code

Numerical simulations were performed in python using the code accessible at https://github.com/jules-fbl/LV_community_selection. All the figures of the paper were obtained with a number of species $S = 100$, $m = 1$ selected community out of $n = 10$, a mutation strength $\varepsilon = 0.02$, an initial interaction matrix drawn from a Gaussian distribution of parameters $\mu = 3$, $\sigma = 0.3$ and $\gamma = 0$ and random carrying capacities drawn uniformly between 0.5 and 1.5. The collective generation time was chosen to be $T = 500$ (with the exception of the first generation where a time $T = 5000$ was used in order to avoid the propagation of transient effects). This time is long enough for the mutated communities to approach their equilibrium abundances. To integrate the Lotka-Volterra equations, we used an integration scheme described in section 11 of the SI. We also imposed an abundance cut-off $N_{\min} = 10^{-20}$ below which species are deemed extinct.

Derivation of recursive equation for the interaction matrix

We here explain the derivation of Eq. 7 in the case $\gamma = 0$. The complete derivation can be found in the SI.

Let α_{ij} be the interaction matrix of a community at generation τ, of empirical mean and variance μ/S and $\sigma/\sqrt{S}$ and let $\mathbf{N}$ be the associated abundances at equilibrium. After a mutational step, the interaction matrix becomes, at first order in ε :

$$\hat{\alpha}_{ij} = \alpha_{ij} + \frac{\varepsilon\sigma}{\sqrt{S}}\eta_{ij} \tag{12}$$

with η a Gaussian matrix of expected value zero and variance 1. The mutation of α is equivalent to altering the carrying capacities by $\delta\mathbf{K} = -\varepsilon\frac{\sigma}{\sqrt{S}}\eta\mathbf{N}$.

We define the perturbation matrix as $\chi_{ij} = \frac{\partial N_i}{\partial K_j}$. This matrix measures the effect of a small change in the carrying capacities on the abundances at equilibrium and is related to the interaction matrix through the identity $\chi = (\mathbb{I}^\star + \alpha^\star)^{-1}$ (see SI section 3). Then if $\hat{\mathbf{N}}$ are the equilibrium abundances associated to the mutated matrix, the variation $\delta\mathbf{N} = \hat{\mathbf{N}} - \mathbf{N}$ can be seen as a first order perturbation:

$$\delta\mathbf{N} = \chi\delta\mathbf{K} = -\varepsilon\frac{\sigma}{\sqrt{S}}\chi\eta\mathbf{N}. \tag{13}$$

The induced variation δf of the total abundance is:

$$\delta f = \mathbf{1}\cdot\delta\mathbf{N} = -\varepsilon\frac{\sigma}{\sqrt{S}}\sum_{ij}(\chi^\top\mathbf{1})_i N_j\eta_{ij} \tag{14}$$

For n realizations of matrices $(\eta_1, \ldots, \eta_n)$ (one for each newborn community), we can find the distribution of the matrix that maximizes δf by using extreme value statistics (see SI section 4):

$$\eta_{ij}^{\max} = -M_n\frac{(\chi^\top\mathbf{1})_i}{\|\chi^\top\mathbf{1}\|}\frac{N_j}{\|\mathbf{N}\|} + B_{ij} \tag{15}$$

with M_n a random variable following the statistic of the maximum value of n Gaussian variables and B a Gaussian matrix.

Then, the selected interaction matrix at generation $\tau + 1$ is obtained by putting the expression of $\eta^{\max}$ in Eq. 12. Defining $\mathbf{v} = \chi^\top\mathbf{1}$, we get Eq. 7.

ACKNOWLEDGMENTS

We thank Matthieu Barbier for insightful discussions. This research was partially supported by a grant from the Simons Foundation (N. 454935 Giulio Biroli). SDM was supported by the French Government under the program Investissements d'Avenir (ANR-10-LABX-54 MEMOLIFE and ANR-11-IDEX- 0001-02PSL).

[1] D. U. Hooper, F. S. Chapin, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, D. M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Setälä, A. J. Symstad, J. Vandermeer, and D. A. Wardle, Effects of Biodiversity on Ecosystem

Functioning: A Consensus of Current Knowledge, Ecological Monographs **75**, 3 (2005).

[2] K. E. Fujimura, N. A. Slusher, M. D. Cabana, and S. V. Lynch, Role of the gut microbiota in defining human health, Expert Review of Anti-infective Therapy **8**, 435 (2010).

[3] M. E. Katz, Z. V. Finkel, D. Grzebyk, A. H. Knoll, and P. G. Falkowski, Evolutionary Trajectories and Biogeochemical Impacts of Marine Eukaryotic Phytoplankton, Annual Review of Ecology, Evolution, and Systematics **35**, 523 (2004).

[4] F. I. Arias-Sánchez, B. Vessman, and S. Mitri, Artificially selecting microbial communities: If we can breed dogs, why not microbiomes?, PLOS Biology **17**, e3000356 (2019).

[5] Á. Sánchez, J. C. C. Vila, C.-Y. Chang, J. Diaz-Colunga, S. Estrela, and M. Rebolleda-Gomez, Directed Evolution of Microbial Communities, Annual Review of Biophysics **50**, 323 (2021).

[6] W. Swenson, D. S. Wilson, and R. Elias, Artificial ecosystem selection, Proceedings of the National Academy of Sciences **97**, 9110 (2000).

[7] S. Jacquiod, A. Spor, S. Wei, V. Munkager, D. Bru, S. J. Sørensen, C. Salon, L. Philippot, and M. Blouin, Artificial selection of stable rhizosphere microbiota leads to heritable plant phenotype changes, Ecology Letters 10.1111/ele.13916 (2021).

[8] U. G. Mueller, T. E. Juenger, M. R. Kardish, A. L. Carlson, K. Burns, C. C. Smith, and D. L. D. Marais, Artificial Microbiome-Selection to Engineer Microbiomes That Confer Salt-Tolerance to Plants, bioRxiv , 081521 (2016).

[9] H. T. P. Williams and T. M. Lenton, Artificial selection of simulated microbial ecosystems, Proceedings of the National Academy of Sciences **104**, 8918 (2007).

[10] A. Penn, Modelling Artificial Ecosystem Selection: A Preliminary Investigation, in *Advances in Artificial Life*, Lecture Notes in Computer Science, edited by W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, and J. T. Kim (Springer, 2003) pp. 659–666.

[11] C.-Y. Chang, J. C. C. Vila, M. Bender, R. Li, M. C. Mankowski, M. Bassette, J. Borden, S. Golfier, P. G. L. Sanchez, R. Waymack, X. Zhu, J. Diaz-Colunga, S. Estrela, M. Rebolleda-Gomez, and A. Sanchez, Engineering complex communities by directed evolution, Nature Ecology & Evolution **5**, 1011 (2021).

[12] L. Xie and W. Shou, Steering ecological-evolutionary dynamics to improve artificial selection of microbial communities, Nature Communications **12**, 6799 (2021).

[13] S. De Monte, Ecological recipes for selecting community function, Nature Ecology & Evolution **5**, 894 (2021).

[14] L. Xie, A. E. Yuan, and W. Shou, Simulations reveal challenges to artificial community selection and possible strategies for success, PLOS Biology **17**, e3000295 (2019).

[15] S. van Vliet and M. Doebeli, The role of multilevel selection in host microbiome evolution, Proceedings of the National Academy of Sciences **116**, 20591 (2019).

[16] G. Doulcier, A. Lambert, S. De Monte, and P. B. Rainey, Eco-evolutionary dynamics of nested Darwinian populations and the emergence of community-level heredity, eLife **9**, e53433 (2020).

[17] E. P. Wigner, Random matrices in physics, SIAM review **9**, 1 (1967).

[18] R. M. May, Stability in multispecies community models, Mathematical Biosciences **12**, 59 (1971).

[19] R. M. May, Will a Large Complex System be Stable?, Nature **238**, 413 (1972).

[20] D. A. Kessler and N. M. Shnerb, Generalized model of island biodiversity, Physical Review E **91**, 042705 (2015).

[21] C. K. Fisher and P. Mehta, The transition between the niche and neutral regimes in ecology, Proceedings of the National Academy of Sciences **111**, 13111 (2014).

[22] G. Bunin, Ecological communities with Lotka-Volterra dynamics, Physical Review E **95**, 042414 (2017).

[23] C. A. Serván, J. A. Capitán, J. Grilli, K. E. Morrison, and S. Allesina, Coexistence of many species in random ecosystems, Nature ecology & evolution **2**, 1237 (2018).

[24] G. Biroli, G. Bunin, and C. Cammarota, Marginally stable equilibria in critical ecosystems, New Journal of Physics **20**, 083051 (2018).

[25] M. Tikhonov and R. Monasson, Collective phase in resource competition in a highly diverse ecosystem, Physical review letters **118**, 048103 (2017).

[26] A. Altieri and S. Franz, Constraint satisfaction mechanisms for marginal stability and criticality in large ecosystems, Physical Review E **99**, 010401 (R) (2019).

[27] M. T. Pearce, A. Agarwala, and D. S. Fisher, Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos, Proceedings of the National Academy of Sciences (2020).

[28] F. Roy, M. Barbier, G. Biroli, and G. Bunin, Complex interactions can create persistent fluctuations in high-diversity ecosystems, PLOS Computational Biology **16**, 1 (2020).

[29] R. Marsland, W. Cui, and P. Mehta, A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns, Scientific reports **10**, 1 (2020).

[30] L. Sidhom and T. Galla, Ecological communities from random generalized lotka-volterra dynamics with nonlinear feedback, Physical Review E **101**, 032101 (2020).

[31] I. Dalmedigos and G. Bunin, Dynamical persistence in high-diversity resource-consumer communities, PLOS Computational Biology **16**, e1008189 (2020).

[32] A. Altieri, F. Roy, C. Cammarota, and G. Biroli, Properties of equilibria and glassy phases of the random lotka-volterra model with demographic noise, Physical Review Letters **126**, 258301 (2021).

[33] A. Altieri and G. Biroli, Effects of intraspecific cooperative interactions in large ecosystems, arXiv:2105.04519 [cond-mat] (2021), arXiv:2105.04519 [cond-mat].

[34] G. Garcia Lorenzana and A. Altieri, Well-mixed lotka-volterra model with random strongly competitive interactions, arXiv e-prints , arXiv (2021).

[35] A. Traulsen and M. A. Nowak, Evolution of cooperation by multilevel selection, Proceedings of the National Academy of Sciences **103**, 10952 (2006).

[36] V. Bucci and J. B. Xavier, Towards predictive models of the human gut microbiome, Journal of molecular biology **426**, 3907 (2014).

[37] R. C. Lewontin, The Units of Selection, Annual Review of Ecology and Systematics **1**, 1 (1970).

[38] A. Penn and I. Harvey, The Role of Non-Genetic Change in the Heritability, Variation, and Response to Selection of Artificially Selected Ecosystems, in *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems* (The MIT Press, 2004) https://direct.mit.edu/book/chapter-pdf/268269/9780262291392_ccf.pdf.

[39] T. Ikegami and K. Hashimoto, Dynamical Systems Approach to Higher-level Heritability, Journal of Biological Physics **28**, 799 (2002).

[40] J. Ginibre, Statistical Ensembles of Complex, Quaternion, and Real Matrices, Journal of Mathematical Physics **6**, 440 (1965).

[41] J. Baik, G. B. Arous, and S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Annals of Probability **33**, 1643 (2005).

[42] S. F. Edwards and R. C. Jones, The eigenvalue spectrum of a large symmetric random matrix, Journal of Physics A: Mathematical and General **9**, 1595 (1976).

[43] E. J. Gumbel, *Statistics of Extremes* (Courier Corporation, 2004).

[44] M. Barbier, C. de Mazancourt, M. Loreau, and G. Bunin, Fingerprints of High-Dimensional Coexistence in Complex Ecosystems, Physical Review X **11**, 011009 (2021).

[45] Y. Kasai and A. Okiji, Hidden mattis phase in annealed system, Progress of Theoretical Physics **69**, 20 (1983).

[46] R. Penney, A. Coolen, and D. Sherrington, Coupled dynamics of fast spins and slow interactions in neural networks and spin systems, Journal of Physics A: Mathematical and General **26**, 3681 (1993).

[47] L. Foini and J. Kurchan, Annealed averages in spin and matrix models, arXiv preprint arXiv:2104.04363 (2021).

[48] A. M. Saxe, J. L. McClelland, and S. Ganguli, A mathematical theory of semantic development in deep neural networks, Proceedings of the National Academy of Sciences **116**, 11537 (2019).

[49] F. Schuessler, F. Mastrogiuseppe, A. Dubreuil, S. Ostojic, and O. Barak, The interplay between randomness and structure during learning in RNNs, arXiv:2006.11036 [q-bio] (2021), 2006.11036 [q-bio].

[50] Venturelli, Deciphering microbial interactions in synthetic human gut microbiome communities, Molecular Systems Biology **14**, e8157 (2018).

[51] D. Machado, O. M. Maistrenko, S. Andrejev, Y. Kim, P. Bork, K. R. Patil, and K. R. Patil, Polarization of microbial communities between competitive and cooperative metabolism, Nature Ecology & Evolution **5**, 195 (2021).

[52] J. E. Goldford, N. Lu, D. Bajić, S. Estrela, M. Tikhonov, A. Sanchez-Gorostiaga, D. Segrè, P. Mehta, and A. Sanchez, Emergent simplicity in microbial community assembly, Science **361**, 469 (2018).

[53] M. Barbier, J.-F. Arnoldi, G. Bunin, and M. Loreau, Generic assembly patterns in complex ecological communities, Proceedings of the National Academy of Sciences **115**, 2156 (2018).

[54] W. Cui, R. Marsland, and P. Mehta, Diverse communities behave like typical random ecosystems, Physical Review E **104**, 034416 (2021).

[55] K. Faust and J. Raes, Microbial interactions: From networks to models, Nature Reviews Microbiology **10**, 538 (2012).

[56] C. Ratzke, J. Barrere, and J. Gore, Strength of species interactions determines biodiversity and stability in microbial communities, Nature Ecology & Evolution **4**, 376 (2020).

[57] S. Estrela, Á. Sánchez, and M. Rebolleda-Gómez, Multi-Replicated Enrichment Communities as a Model System in Microbial Ecology, Frontiers in Microbiology **12**, 760 (2021).

[58] S. Marcus, A. M. Turner, and G. Bunin, Local and collective transitions in sparsely-interacting ecological communities, arXiv:2110.13603 [cond-mat, physics:physics, q-bio] (2021), arXiv:2110.13603 [cond-mat, physics:physics, q-bio].

[59] G. Bunin, Directionality and community-level selection, Oikos **130**, 489 (2021).

[60] P. Márquez-Zacarías, P. L. Conlin, K. Tong, J. T. Pentz, and W. C. Ratcliff, Why have aggregative multicellular organisms stayed simple?, Current Genetics 10.1007/s00294-021-01193-0 (2021).

[61] D. S. Wilson and E. Sober, Reviving the superorganism, Journal of Theoretical Biology **136**, 337 (1989).

[62] M. Loreau, The Ecosystem: Superorganism, or Collection of Individuals?, in *The Ecosystem: Superorganism, or Collection of Individuals?* (Princeton University Press, 2020) pp. 218–224.

[63] J.-C. Simon, J. R. Marchesi, C. Mougel, and M.-A. Selosse, Host-microbiota interactions: From holobiont theory to analysis, Microbiome **7**, 5 (2019).

[64] C. Bang, T. Dagan, P. Deines, N. Dubilier, W. J. Duschl, S. Fraune, U. Hentschel, H. Hirt, N. Hülter, T. Lachnit, D. Picazo, L. Pita, C. Pogoreutz, N. Rädecker, M. M. Saad, R. A. Schmitz, H. Schulenburg, C. R. Voolstra, N. Weiland-Bräuer, M. Ziegler, and T. C. G. Bosch, Metaorganisms in extreme environments: Do microbes play a role in organismal adaptation?, Zoology **127**, 1 (2018).

[65] K. Liautaud, E. H. van Nes, M. Barbier, M. Scheffer, and M. Loreau, Superorganisms or loose collections of species? a unifying theory of community patterns along environmental gradients, Ecology Letters **22**, 1243 (2019).

[66] S. De Monte and P. B. Rainey, Nascent multicellular life and the emergence of individuality, Journal of Biosciences **39**, 237 (2014).

[67] A. J. Black, P. Bourrat, and P. B. Rainey, Ecological scaffolding and the evolution of individuality, Nature Ecology & Evolution **4**, 426 (2020).