Impact of colonization history on the composition of ecological systems

Nannan Zhao, Serguei Saavedra, and Yang-Yu Liu

Phys. Rev. E **103**, 052403 — Published 5 May 2021

DOI: 10.1103/PhysRevE.103.052403

The impact of colonization history on the composition of ecological systems

Nannan Zhao,[1, 2] Serguei Saavedra,[3] and Yang-Yu Liu[1, 4, *]
[1] *Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA*
[2] *Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, 710129, China*
[3] *Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA*
[4] *Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, MA, 02115, USA*

(Dated: April 23, 2021)

Observational studies of ecological systems have shown that different species compositions can arise from distinct species arrival orders during community assembly—also known as colonization history. The presence of multiple interior equilibria in the positive orthant of the state space of the population dynamics will naturally lead to history dependency of the final state. However, it is still unclear whether and under which conditions colonization history will dominate community composition in the absence of multiple interior equilibria. Here, by considering that only one species can invade at a time and there are no recurrent invasions, we show clear evidence that the colonization history can have a big impact on the composition of ecological systems even in the absence of multiple interior equilibria. In particular, we first derive two simple rules to determine whether the composition of a community will depend on its colonization history in the absence of multiple interior equilibria and recurrent invasions. Then we apply them to communities governed by generalized Lotka-Volterra (gLV) dynamics and propose a numerical scheme to measure the probability of colonization history dependence. Finally, we show, via numerical simulations, that for gLV dynamics with a single interior equilibrium the probability that community composition is dominated by colonization history increases monotonically with community size, network connectivity, and the variation of intrinsic growth rates across species. These results reveal that in the absence of multiple interior equilibria and recurrent invasions community composition is a probabilistic process mediated by ecological dynamics via the interspecific variation and the size of regional pools.

I. INTRODUCTION

Ecological communities are formed by co-occurring and interacting species in a given place and time [1–3]. It has been shown that within these communities, the specific composition of species is a function of several ecological, evolutionary, and stochastic processes [3–6]. Importantly, one of the main factors affecting community composition is the order of species arrival—also known as colonization history [7–11]. That is, colonization history can introduce priority effects, where the persistence of species depends on the order at which they join a given community.

Many mathematical or physical tools have been used to investigate the impact of colonization history on community composition. For example, Ref. [12] introduced a toy model of ecosystem assembly to map out all assembly pathways generated by external invasions. The colonization process was characterized as a finite Markov chain, and proved to exhibit a unique set of recurrent states (the end state of the process) that are resistant to invasions. This also shows that the end state is independent of the assembly history. For replicator systems and the gLV model, a phase transition was found from a phase where a unique globally attractive fixed point exists to a phase where multiple dynamical attractors exist [13–20]. The latter phase naturally leads to history-dependent community properties.

Although many other studies have also shown history dependent or independent communities in different theoretical models [21–27], these studies do not allow us to find general conditions under which colonization history can have the highest (or lowest) chance to affect community composition. In fact, it is still unclear whether and under which conditions colonization history will dominate community composition in the absence of multiple interior equilibria and recurrent invasions. Note that in the presence of a single interior equilibrium, if we allow for recurrent invasion, then sooner or later this interior equilibrium will be reached, and the assembly process (in the long run) will be trivially history independent. Therefore, in this work we are interested in the history dependency by considering that only one species can invade at a time, and there are no recurrent invasions.

The complexity of factors affecting community assembly has undercut our ability to anticipate whether a given regional pool of species can be more susceptible to colonization history than another. Yet, knowing this can advance our understanding about the probabilistic nature and predictability of ecological communities. In this paper, we try to address the following key questions: In the absence of multiple interior equilibria and recurrent invasions, are there any conditions under which colonization history will completely dominate community composition? Does the type of interspecific interactions affect the probability that community composition depends on colonization history? How do the intrinsic properties of species affect the impact of colonization history on community composition? In the face of an accelerating rate of species turnover, answering these questions is important in order to understand and anticipate key biodiversity changes in ecological communities.

The rest of the paper is organized as follows: Section II provides two motivating examples, and then introduces two simple rules to determine the relationship between community composition and colonization history. Section III defines a community model with gLV dynamics and proposes a numerical scheme to measure the probability of colonization history dependence. Section IV discusses the effects of both community and intrinsic properties on the history dependence. Section V is devoted to a discussion on the limitations of our current work and some potential extensions.

II. MOTIVATING EXAMPLES

To illustrate the scope and assumptions behind our study, we start our analysis by considering a small pool of 3 species that can coexist at a unique interior equilibrium, as shown in Fig. 1. It is worth mentioning that in order to increase the tractability of the problem, here we assume that only one species can invade at a time and the ecological dynamics is fast enough to reach a boundary equilibrium (i.e., with some of the S species having zero abundance) before the next species invasion. Thus, in this example, there are 6 possible colonization histories (or assembly paths), one unique interior equilibrium, and 6 different boundary equilibria from which an ecological community of 3 species can be assembled by introducing one species at a time, i.e., via successive invasions to an empty community $\varnothing$ (see Fig. 1b and Fig. 1e).

As shown in Fig. 1b, species 2 and 3 cannot coexist on their own (i.e., the boundary equilibrium $\{2,3\}$ is infeasible), the assembly paths (of length 3): $\varnothing \to 3 \to 2 \to 1$ and $\varnothing \to 2 \to 3 \to 1$ cannot achieve the final community $\{1,2,3\}$, while other assembly paths (of length 3) can. Note that if we allow for recurrent invasion, then the assembly path (of length 4): $\varnothing \to 2 \to 3 \to 1 \to 2$ can eventually achieve $\{1,2,3\}$. Hence, the history dependence we are interested in is equivalent to the existence of non-optimal assembly paths (of length $> S$) to assemble a community formed by S species, that is, there is at least one infeasible boundary equilibrium. And history independence means that starting from the empty community all the $S!$ assembly paths (of optimal length S) can lead to the final state with S species stably coexist, that is, all boundary equilibria are feasible. For example, for the system shown in Fig. 1e, any subset of the 3 species can coexist at their equilibria. In this case, the final community composition is independent of the colonization history, and any optimal assembly paths (of length S) can assemble a community formed by S species.

Based on the above observations, in the absence of multiple interior equilibria and recurrent invasions, we have two simple rules to determine the relationship of community composition and colonization history:

First, if there exists a set of species $\{\mathcal{S}\}$ that can coexist at a unique equilibrium, but a smaller subset of species $\{\mathcal{T}\}\,(\subset \{\mathcal{S}\})$ cannot, then the final community composition formed by the S $(= |\{\mathcal{S}\}|)$ species depends on the colonization history. This rule can be understood as follows. For a given regional pool $\{\mathcal{S}\} = \{1, ..., S\}$, without loss of generality, let us assume there is only one subcommunity $\{\mathcal{T}\} = \{1, ..., T\}$ (with $T < S$) that does not have a feasible boundary equilibrium, then the following two assembly paths (of the same length S): (1) $\varnothing \to 1 \to \cdots \to T-1 \to T \to T+1 \to \cdots \to S$; (2) $\varnothing \to 1 \to \cdots \to T-1 \to T+1 \to T \to \cdots \to S$ will yield different states. The former leads to $\{1, \cdots, T-1, T+1, \cdots, S\}$ with $S-1$ species present, while the latter leads to $\{1, \cdots, S\}$ with all the S species present. Similar arguments can be applied to the case of more infeasible subcommunities. Basically, the assembly paths (of length S) that avoid those roadblocks (infeasible subcommunities) will naturally lead to $\{1, \cdots, S\}$ with all the S species present, while those assembly paths (of length S) that encounter those roadblocks (infeasible subcommunities) will lead to different final states, in the absence of recurrent invasions.

Second, if any subcommunities of $\{\mathcal{S}\}$ can coexist at a unique boundary equilibrium, then the final community composition formed by the S species does not depend on the colonization history. This result can be explained via the following example: when a new species is added to a species pool to build a new community $\{\mathcal{I}\} = \{1, 2, ..., I\}$, the community with I $(< S)$ species can stably coexist in a unique boundary equilibrium, which is stable and feasible. Following these steps to add more and more species one at a time, we will always reach the unique interior equilibrium with all the S species present, regardless of the detailed assembly path. Thus, the final state is independent of the colonization history.

We emphasize that the above two simple rules hold for a wide class of population dynamics models where all the S species can coexist at a unique interior equilibrium and recurrent invasions are not allowed. However, we still lack a numerical scheme to quantitatively study the impact of colonization history on community assembly. Directly constructing the assembly graph for large systems is computationally intractable. Indeed, for a species pool of size S, suppose the S species can coexist at a stable interior equilibrium, then starting from the empty community, with one species can invade at a time, there are $S!$ potential assembly paths to reach the final state where all the S species present, and there are $(2^S - 2)$ possible subcommunities (with at least one and at most $S-1$ species). If there are some subcommunities

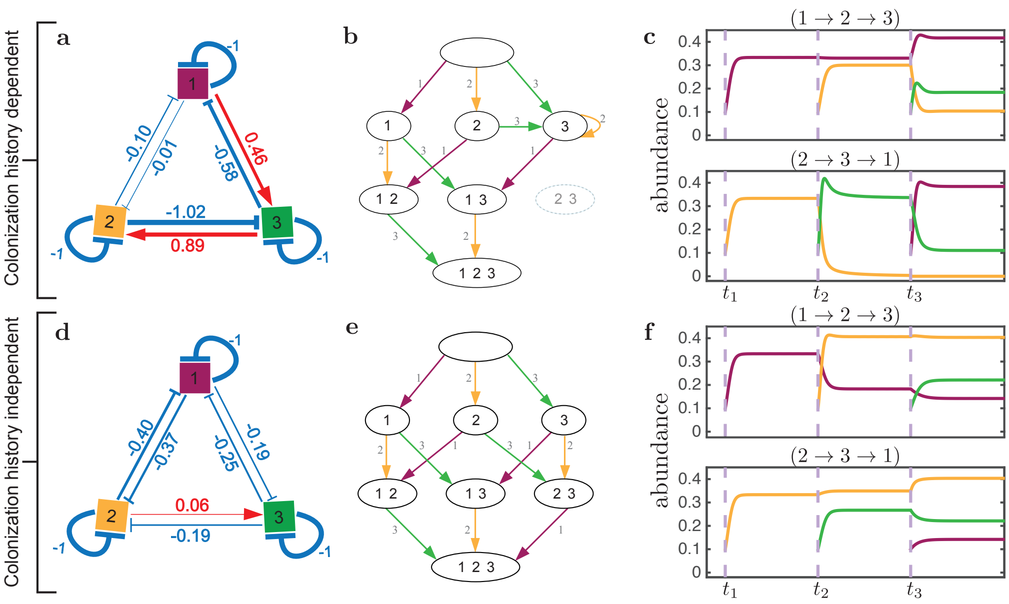

FIG. 1. **Ecological communities can display different dependencies on colonization history.** For illustration purposes, we show the assembly of a 3-species community $\{1, 2, 3\}$ by the invasion or colonization of one species at a time, following the gLV dynamics. There are in total $3! = 6$ different colonization trajectories. (a) The ecological network depicts the pairwise interactions among the three species (which are also encoded in the interaction matrix $\mathbf{A}$). The feasible intrinsic growth rate vector $\mathbf{r}$ is set to be $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^{\top}$. (b) Starting from an empty ecological community $\varnothing$ (top node), the three species are added successively into the community via different orders. Since species-2 and 3 cannot coexist (gray node), the community composition will be dependent on colonization history. That is, the final state of the three species together cannot be assembled if we follow the trajectory $\varnothing \to 3 \to 2 \to 1$ or $\varnothing \to 2 \to 3 \to 1$, while the other four trajectories will lead to the desired final state. (c) As an example, we show two different trajectories and their final community compositions. Panels (d-f) show a similar case as the previous example, but with different interaction matrix $\mathbf{A}$. In this case, the community composition is independent on colonization history.

that do not have feasible boundary equilibria, they will serve as roadblocks in some assembly paths (or isolated nodes in the assembly graph, e.g., $\{2, 3\}$ in Fig. 1b). For large S, we know that both $S!$ and $(2^S - 2)$ are notoriously large, rendering the construction of the assembly graph computationally intractable. Without the assembly graph at hand, it is very hard, if not impossible, to identify which assembly paths (of optimal length S) will lead to the final state with all S species present. Moreover, once we encounter a roadblock in a particular assembly path, it is very challenging to predict exactly how many recurrent invasions will be needed to eventually reach the final state with all S species present, if we do not have the assembly graph. Therefore, for ecosystems with only one attracting interior equilibrium, even though the community assembly with recurrent invasions allowed is trivially history independent in the long run, it becomes highly nontrivial without recurrent invasions.

III. MODEL DEFINITION

In this work we choose the classical gLV model to quantitatively study the impact of colonization history on community composition. This model includes parameters that govern the intrinsic growth rates of different species and pairwise interactions among different species, and it is tractable enough to allow us to investigate the conditions under which community composition depends on colonization history.

The gLV model can be written as follows

$$\frac{\mathrm{d}N_i(t)}{\mathrm{d}t} = N_i(t)\left(r_i + \sum_{j=1}^{S} A_{ij}\, N_j(t)\right), i = 1, \cdots, S, \quad (1)$$

where N_i is the abundance (or biomass) of species-i, S corresponds to the number of species in the community, $\mathbf{A} = [A_{ij}]_{S \times S}$ is the interaction matrix whose elements denote the per capita effect of one species on the per capita growth rate of another species, and r_i is the intrinsic growth rate of species-i.

To ensure that the S species can coexist at a unique interior equilibrium, following previous studies [28, 29], we focus on diagonally stable interaction matrices $\mathbf{A}$ (i.e., there is a positive definite diagonal matrix $\mathbf{D}$ such that $\mathbf{D}\mathbf{A} + \mathbf{A}^{\top}\mathbf{D}$ is a negative definite symmetric matrix [30]). A diagonally stable interaction matrix $\mathbf{A}$ guarantees that the gLV model has a single, globally, attractive equilibrium [31]. We emphasize that the assumption of a diagonally stable interaction matrix $\mathbf{A}$ is deeply driven by the complexity of this problem and allows us to focus on the feasibility of the system—the necessary condition for species coexistence [32, 33].

To construct the $\mathbf{A}$ matrix, we first capture its binary structure by constructing an Erdős-Rényi (ER) random graph. We begin with S isolated nodes (species). For each of the $S(S-1)/2$ node pairs, we construct an undirected edge between the two nodes with probability C. It is worth noting that C represents the connectance of the community (i.e., the ratio between actual and potential interactions in the ecological network). Once the ER graph is constructed, we assign the interspecific interaction strengths to the edges (here we treat edges as bidirectional). The interaction strengths $A_{ij} (i \neq j)$ are drawn from a normal distribution $\mathcal{N}\left(0, \sigma^2\right)$, where $\sigma = 1/\sqrt{S(2+\epsilon)}$ denotes the characteristic interspecific interaction strength, and ϵ is a constant [34]. The diagonal elements are set to be $A_{ii} = -d$ with d representing the intrinsic damping time scale of each species.

We consider three interaction types: (1) random (no sign structure). A_{ij} and A_{ji} are independently sampled from $\mathcal{N}\left(0, \sigma^2\right)$. (2) predator-prey $(+,-)$: We generate a random number p from a uniform distribution $\mathcal{U}[0,1]$. If $p \leq 0.5$, we draw A_{ij} from a half-normal distribution $\left|\mathcal{N}\left(0, \sigma^2\right)\right|$, and A_{ji} from $-\left|\mathcal{N}\left(0, \sigma^2\right)\right|$. If $p > 0.5$, we do the opposite. (3) mixture of competition $(-,-)$ and mutualism $(+,+)$: we generate a random number p from $\mathcal{U}[0,1]$. If $p \leq 0.5$, we draw A_{ij} and A_{ji} independently from $\left|\mathcal{N}\left(0, \sigma^2\right)\right|$. If $p > 0.5$, we draw A_{ij} and A_{ji} independently from $-\left|\mathcal{N}\left(0, \sigma^2\right)\right|$.

To ensure the randomly-generated interaction matrix $\mathbf{A}$ is diagonally stable, we set $\epsilon = 0.01$ and $d = 1$ [34]. Furthermore, to ensure the coexistence of the whole community with S species, we use the feasibility domain to construct feasible vector of intrinsic growth rates [35]. When the interaction matrix $\mathbf{A}$ is given, the feasibility domain can be determined as an algebraic cone:

$$D_F(\mathbf{A}) = \{\mathbf{r} = N_1^* \mathbf{v}_1 + \cdots + N_S^* \mathbf{v}_S\}, \tag{2}$$

where $N_i^* > 0$ is the equilibrium abundance of species i, and $\mathbf{v}_i$ is the spanning vector of the algebraic cone, whose j-th component is given by $v_{ij} = \frac{-A_{ji}}{\sqrt{\sum_{k=1}^{S} A_{ki}^2}}$. If the vector of intrinsic growth rates $\mathbf{r}$ is chosen inside the feasibility domain $D_F(\mathbf{A})$, the community with S species will always be feasible. This feasible vector can be defined as follows: $\mathbf{r} = \sum_{i=1}^{S} n_i^* \mathbf{v}_i$, where $n_i^* \in (0,1)$ and $\sum_{i=1}^{S} n_i^* = 1$. This procedure guarantees that there is at least one assembly path that can given rise to the whole community formed by S-coexisting species.

IV. ANALYSIS AND RESULTS

As noted in the simple rules mentioned in Sec. II, whether all subsets of species can coexist at their boundary equilibria will determine if the community composition depends on colonization history. Under the gLV dynamics, this coexistence is guaranteed if the equilibria of system (1) are feasible (i.e., all present species have positive abundance) and globally stable for all subcommunities. It has been proved that if the interaction matrix $\mathbf{A}$ is diagonally stable, then all sub-matrices $(\hat{\mathbf{A}})$ are diagonally stable as well [30], and the non-trivial positive equilibrium will be globally asymptotically stable (that is, species can stably coexist) [31]. These matrix properties imply that we only need to guarantee the feasibility of the boundary equilibria for all subcommunities. The unique boundary equilibrium of every subcommunity with k species $(k < S)$ under gLV dynamics can be calculated as $\hat{\mathbf{N}}^* = (\hat{N}_1^*, \cdots, \hat{N}_k^*) = -\hat{\mathbf{A}}^{-1}\hat{\mathbf{r}}$, where $\hat{\mathbf{A}}$ and $\hat{\mathbf{r}}$ are the reduced interaction matrix $(k \times k)$ and intrinsic growth rate vector $(k \times 1)$ of the corresponding subcommunity, and when $\hat{N}_i^* > 0\,(i = 1, \cdots, k)$, this subcommunity is feasible, otherwise, it is infeasible. Thus, we can obtain a numerical scheme to determine for the gLV model whether the community composition depends on colonization history: For a community $\{\mathcal{S}\}$ that follows the gLV dynamics characterized by a diagonally stable interaction matrix $\mathbf{A}$ and a feasible intrinsic growth rate vector $\mathbf{r}$, in the absence of recurrent invasions, if there exists an infeasible subcommunity $\{\mathcal{T}\}$, then the final community composition of the community $\{\mathcal{S}\}$ depends on colonization history. Otherwise, it will be colonization-history independent.

A. Examples of small communities

To illustrate the application of the above numerical scheme, we consider the two 3-species communities shown in Fig. 1. The community shown in Fig. 1a is characterized by a feasible intrinsic growth rate vector $\mathbf{r} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^\top$, and a diagonally stable interaction matrix.

$$\mathbf{A} = \begin{bmatrix} -1.00 & -0.01 & 0.46 \\ -0.10 & -1.00 & -1.02 \\ -0.58 & 0.89 & -1.00 \end{bmatrix}.$$

In this case, we can verify that there exists a feasible interior equilibrium $\mathbf{N}^* = (0.417, 0.104, 0.184)^\top$ for the three species, but there is no feasible boundary equilibrium for the species-pair $\{2, 3\}$. Thus, based on the above result, the final community composition will depend on the colonization history. Indeed, Figure 1c shows two different final states obtained by two assembly paths. Figure 1d shows a community characterized by the same feasible intrinsic growth rate vector $\mathbf{r}$, but with a different diagonally stable interaction matrix

$$\mathbf{A} = \begin{bmatrix} -1.00 & -0.37 & -0.19 \\ -0.40 & -1.00 & 0.06 \\ -0.25 & -0.19 & -1.00 \end{bmatrix}.$$

In this case, all subsets of species have the unique and feasible boundary equilibria. Thus, the final community composition does not depend on the colonization history: any assembly path will eventually yield the same final community composition of the three species (see Fig. 1f for examples of assembly paths).

B. Effects of community properties

A big advantage of our numerical scheme is that it allows us to perform extensive numerical simulations to investigate which properties of the community and individual species can affect the probability that community composition depend on colonization history. In particular, for a given community $\{\mathcal{S}\}$ with gLV dynamics, we can numerically check all subsets (subcommunities) to search whether there exists any infeasible subset $\{\mathcal{T}\}$, and if so, we conclude that the community is history dependent. Furthermore, to study the impact of community properties on the history dependence, we systematically generate diagonally stable interaction matrices $\mathbf{A}$ with different community size (S), network connectance (C), and interaction types. In particular, for each interaction type and given (C, S) values, we first randomly construct an interaction matrix $\mathbf{A}$, and then generate 2,000 feasible intrinsic growth rate vectors $\mathbf{r} \in D_F(\mathbf{A})$. From the 2,000 communities $(\mathbf{A}, \mathbf{r})$, we count how many communities having infeasible subcommunities to calculate the probability that community composition depends on colonization history, denoted as P, for a given interaction type and (C, S) values. We repeat this process for an ensemble of 50 different realizations of $\mathbf{A}$, and then calculate the standard error of the mean (SEM) of P for a given interaction type and (C, S) values.

We find that the probability P always increases with the community size S (Fig. 2-top) or network connectance C (Fig. 2-bottom), regardless of the interaction type. This indicates that the community composition will almost surely dependent on colonization history when an ecological system is composed of a large number of species or when species are highly connected.

To check how this result holds beyond expectations, we consider a simple null model as follows. We assume that a random subset of species has a fixed probability p to have a feasible boundary equilibrium. Then, the probability to have at least one subset of species that does not have a feasible boundary equilibrium is given by $1-p^n$, where $n = 2^S-2$ is the number of possible subsets with at least one and at most $(S-1)$ species. According to the above result, we conclude that the probability that community composition depends on colonization history is $P_{\text{null}} = 1-p^n$, where the subscript 'null' stands for the null model.

In the top panel of Fig. 2, for a given community size S, we plot P_{null} based on different values of p (horizontal lines). Clearly, lower p yields higher P_{null}, regardless of the connectance C. For example, for $S = 10$, we have $P_{\text{null}} \sim 0.1$ for $p = 0.9999$ (green line), and $P_{\text{null}} \sim 1$ for $p = 0.99$ (yellow line). However, our calculation based on the gLV model indicates that P increases monotonically with increasing C, and for $S = 10$ we have $P \to 1$ only if C is above 0.6, regardless of the interaction types. In the bottom panel of Fig. 2, for a given value of p, we plot P_{null} as a function of the community size S, finding that P_{null} increases monotonically with S. Note that the S-dependency of P_{null} is heavily driven by the value of p. For example, for $p = 0.999$ (or 0.9), P_{null} will always underestimate (or overestimate, respectively) P for $S < 12$, regardless of the interaction types. The difference between P_{null} and P shown here suggests that ecological dynamics (as simple as they can be) can fundamentally alter the dependency of the community composition on colonization history. That is, this probability cannot be precisely predicted from the probability of feasibility of each individual subset (as assumed in the null model).

C. Effects of intrinsic properties

To investigate the extent to which the variation of intrinsic properties across species affect the history dependence, we systematically generate feasible vectors of intrinsic growth rates $\mathbf{r}$ with different levels of variability across the elements. In particular, for different interaction types, we sample $2,000$ feasible vectors $\mathbf{r} \in D_F(\mathbf{A})$ for each randomly-generated interaction matrix $\mathbf{A}$ with community size $S = 8$ and network connectance $C = 0.4$, and then in each case we can calculate different interspecific variations of intrinsic growth rates $\xi = \frac{<r^2>}{<r>^2}$. We categorize them (with bin width $= 0.05$) according to the variation (ξ) across their elements, and use the curves to replace the histogram to show the expected value. Similarly, here we sample 50 different realizations of the matrix $\mathbf{A}$ to calculate the SEM of P for each ξ bin.

From Fig. 3, we can observe that P increases monotonically with increasing ξ, regardless of the interaction type. In other words, the higher the interspecific variation within a community, the higher the probability that community composition depends on colonization history. An intuitive explanation of this phenomenon is as follows. As we increase ξ, different species tend to have quite different intrinsic growth rates, and hence play different "roles" in the community assembly, rendering higher probability that community composition depends on colonization history.

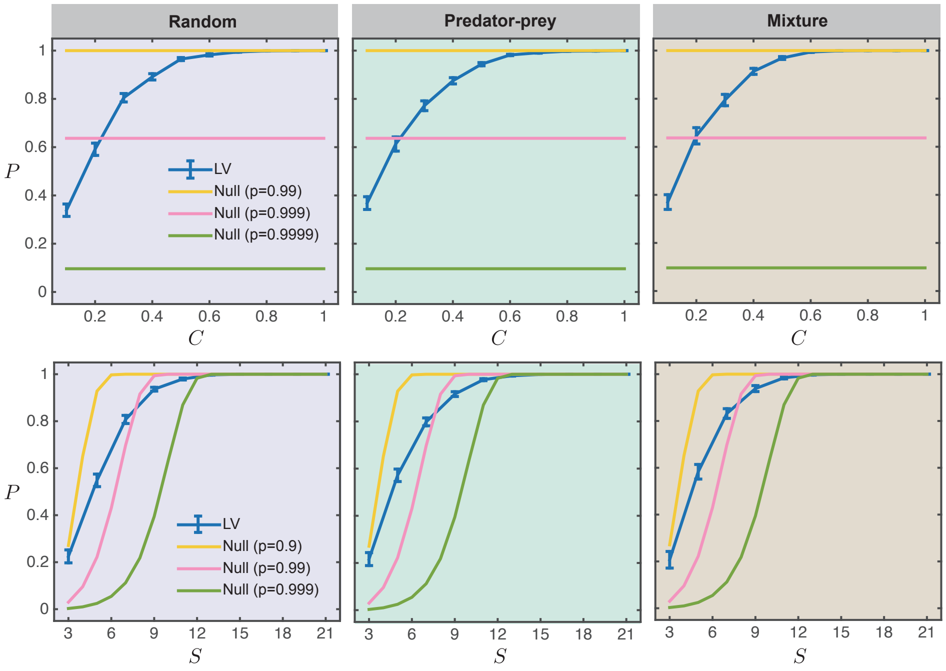

FIG. 2. **The probability P that community composition depends on colonization history as a function of community properties.** P is calculated as a function of community properties: community size S, connectance C, and interaction types, using the gLV model (line plots with error bars) or the null model (lines). (Top) P as a function of connectance C, with fixed community size $S = 10$. (Bottom) P as a function of community size S, with fixed connectance $C = 0.5$. Each column corresponds to a particular interaction type: random interactions, predator-prey, and mixture of competition and mutualism.

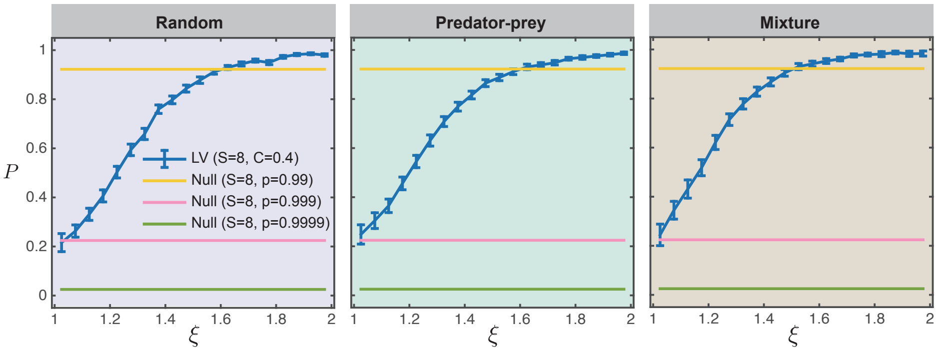

FIG. 3. **The probability P that community composition depends on colonization history as a function of intrinsic properties of species.** The probability is calculated for different levels of interspecific variation ξ (intrinsic growth rate) and interaction types, using the gLV model (line plots with error bars) or the null model (lines). We fix community size $S = 8$ and network connectance $C = 0.4$. Each column corresponds to a particular interaction type: random interactions, predator-prey, and mixture of competition and mutualism.

Here, for the given community size S, we also plot P_{null} based on different values of p. As shown in Fig. 3 (horizontal lines), lower p yields higher P_{null}, regardless of the interspecific variation ξ and interaction types. For example, for $S = 8$, we have $P_{\text{null}} \sim 0.01$ for $p = 0.9999$ (green line), and $P_{\text{null}} \sim 0.9$ for $p = 0.99$ (yellow line). However, our calculation based on the gLV model indicates that P increases monotonically with increasing ξ, and for $S = 8$ we have $P \to 1$ if ξ is above 2, regardless of the interaction types. The difference between P_{null} and P underscores the impact of interspecific variation on the probability that community composition depends on colonization history, which cannot be predicted from the null model. This is acceptable because the simple null model only contains two parameters p and S, while the gLV model of N species contains $N + N^2$ parameters (stored in $\mathbf{r}$ and $\mathbf{A}$). Although the null model cannot accurately fit the simulation results of the gLV model, it can still offer a theoretical guide to predict the qualitative trend so that we can better appreciate why the colonization history matters for large and complex ecological communities.

V. DISCUSSION

In summary, here we offered simple rules linking colonization history and community composition in the absence of multiple interior equilibria and recurrent invasions. Moreover, we applied those rules to communities that are governed by gLV dynamics and proposed a numerical scheme to measure the probability that community composition depends on colonization history. Through extensive simulations, we demonstrated that this probability increases monotonically with community size, network connectance, and variation of intrinsic growth rates across species. Moreover, we proposed a simple null model to fit the above numerical results. However, due to the complexity of the gLV model, especially the various parameter settings of species number, network connectance, or growth rate heterogeneity, the effects cannot be precisely predicted from the null model that only considers the probability of feasibility of each each sub-community of species.

It is worth noting that our current framework has several limitations. First, it focuses on the coexistence of species at a globally stable interior equilibrium[36] based on the assumption of a diagonally stable interaction matrix. But the coexistence of species could be driven by not only an equilibrium state, but also different dynamical behavior, such as limit cycles or chaos, those more complicated scenarios deserve a more dedicated research effort [37]. Second, the simulation framework is applicable to gLV dynamics with linear functional responses. Extending the calculations to population dynamics models with more complicated functional response will also be an interesting future direction [33, 38]. Third, our explanation of the probability P increasing with the variation ξ of species intrinsic growth rates (Fig. 3) is very conceptual. We call for more quantitative explanations of this very interesting phenomenon. Despite those limitations, the simplicity of our work allows us to provide a first-order classification of the conditions modulating the impact of colonization history. This work can serve as a basis for future work aiming to study the extent to which it is possible to reconstruct (or to partially reconstruct) the species arrival order in a community.

ACKNOWLEDGMENT

N.Z. was supported by grant No.201806290054 from China Scholarship Council (CSC). Y.-Y.L. was supported by grants R01AI141529, R01HD093761, R01AG067744, UH3OD023268, U19AI095219, and U01HL089856 from National Institutes of Health (NIH). S.S. was supported by grant No. DEB-2024349 from National Science Foundation (NSF).

---

[*] Corresponding author: yyl@channing.harvard.edu

[1] Jared M Diamond. Assembly of species communities. In *Ecology and Evolution of Communities*, pages 342–444. Harvard University Press, Cambridge, MA, 1975.

[2] Eugene P Odum. The strategy of ecosystem development. In *The Ecological Design and Planning Reader*, pages 203–216. Springer, 2014.

[3] Mark Vellend. *The theory of ecological communities (MPB-57)*. Princeton University Press, 2016.

[4] Robert M May. How many species are there on earth? *Science*, 241(4872):1441–1449, 1988.

[5] Tadashi Fukami. Community assembly dynamics in space. In *Community Ecology: Processes, Models, and Applications*, pages 45–54. Oxford University Press, Oxford, 2010.

[6] Tadashi Fukami. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.*, 46:1–23, 2015.

[7] David Tilman. *Resource competition and community structure*. Princeton university press, 1982.

[8] Peter Jay Morin. Odonate guild composition: experiments with colonization history and fish predation. *Ecology*, 65(6):1866–1873, 1984.

[9] Maarit Jaarola, Hakan Tegelström, and Karl Fredga. Colonization history in fennoscandian rodents. *Biol. J. Linnean Soc.*, 68(1-2):113–127, 1999.

[10] Dina M Fonseca and David D Hart. Colonization history masks habitat preferences in local distributions of stream insects. *Ecology*, 82(10):2897–2910, 2001.

[11] Inés Martínez, Maria X Maldonado-Gomez, João Carlos Gomes-Neto, Hatem Kittana, Hua Ding, Robert Schmaltz, Payal Joglekar, Roberto Jiménez Cardona, Nathan L Marsteller, and Steven W Kembel. Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly. *Elife*, 7:e36521, 2018.

[12] José A Capitán, José A Cuesta, and Jordi Bascompte. Statistical mechanics of ecosystem assembly. *Phys. Rev. Lett.*, 103(16):168101, 2009.

[13] Sigurd Diederich and Manfred Opper. Replicators with random interactions: A solvable model. *Phys. Rev. A*, 39(8):4333, 1989.

[14] Manfred Opper and Sigurd Diederich. Phase transition and 1/f noise in a game dynamical model. *Phys. Rev. Lett.*, 69(10):1616, 1992.

[15] Manfred Opper and Sigurd Diederich. Replicator dynamics. *Comput. Phys. Commun.*, 121:141–144, 1999.

[16] Johannes Berg and Martin Weigt. Entropy and typical properties of nash equilibria in two-player games. *EPL*, 48(2):129, 1999.

[17] Tobias Galla and J Doyne Farmer. Complex dynamics in learning complicated games. *PNAS*, 110(4):1232–1236, 2013.

[18] Tobias Galla. Two-population replicator dynamics and number of nash equilibria in matrix games. *EPL*, 78(2):20005, 2007.

[19] Laura Sidhom and Tobias Galla. Ecological communities from random generalized lotka-volterra dynamics with nonlinear feedback. *Phys. Rev. E*, 101(3):032101, 2020.

[20] Guy Bunin. Ecological communities with lotka-volterra dynamics. *Phys. Rev. E*, 95(4):042414, 2017.

[21] Wilfred M Post and Stuart L Pimm. Community assembly and food web stability. *Math. Biosci.*, 64(2):169–192, 1983.

[22] James A Drake. The mechanics of community assembly and succession. *J. Theor. Biol*, 147(2):213–233, 1990.

[23] Richard Law and R Daniel Morton. Alternative permanent states of ecological communities. *Ecology*, 74(5):1347–1361, 1993.

[24] Julie L Lockwood, Robert D Powell, M Philip Nott, and Stuart L Pimm. Assembling ecological communities in time and space. *Oikos*, 80(3):549–553, 1997.

[25] Yohsuke Murase, Takashi Shimada, Nobuyasu Ito, and Per Arne Rikvold. Effects of demographic stochasticity on biological community assembly on evolutionary time scales. *Phys. Rev. E*, 81(4):041908, 2010.

[26] Shay Be'er, Michael Assaf, and Baruch Meerson. Colonization of a territory by a stochastic population under a strong allee effect and a low immigration pressure. *Phys. Rev. E*, 91(6):062126, 2015.

[27] Ignacio A Rodriguez-Brenes, Dominik Wodarz, and Natalia L Komarova. Beyond the pair approximation: Modeling colonization population dynamics. *Phys. Rev. E*, 101(3):032404, 2020.

[28] Serguei Saavedra, Rudolf P Rohr, Jordi Bascompte, Oscar Godoy, Nathan JB Kraft, and Jonathan M Levine. A structural approach for understanding multispecies coexistence. *Ecol. Monogr.*, 87:470–486, 2017.

[29] Carlos A Serván, José A Capitán, Jacopo Grilli, Kent E Morrison, and Stefano Allesina. Coexistence of many species in random ecosystems. *Nat. Ecol. Evol.*, 2:1237, 2018.

[30] Eugenius Kaszkurewicz and Amit Bhaya. *Matrix diagonal stability in systems and computation*. Springer Science and Business Media, 2012.

[31] Bo S Goh. Global stability in many-species systems. *Am. Nat.*, 111(977):135–143, 1977.

[32] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998.

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

[34] Travis E Gibson, Amir Bashan, Hong-Tai Cao, Scott T Weiss, and Yang-Yu Liu. On the origins and control of community types in the human microbiome. *PLoS Comput. Biol.*, 12(2):e1004688, 2016.

[35] Chuliang Song and Serguei Saavedra. Will a small randomly assembled community be feasible and stable? *Ecology*, 99(3):743–751, 2018.

[36] R. C. Lewontin. The meaning of stability. *Brookhaven Symp. Biol.*, 22:13–24, 1969.

[37] Sebastian J Schreiber, Masato Yamamichi, and Sharon Y Strauss. When rarity has costs: coexistence under positive frequency-dependence and environmental stochasticity. *Ecology*, 100:e02664, 2019.

[38] Mohammad AlAdwani and Serguei Saavedra. Is the addition of higher-order interactions in ecological models increasing the understanding of ecological dynamics? *Math. Biosci.*, 315:108222, 2019.