MIT Open Access Articles

Untangling the complexity of priority effects in multispecies communities

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Song, Chuliang, Fukami, Tadashi and Saavedra, Serguei. 2021. "Untangling the complexity of priority effects in multispecies communities." Ecology Letters, 24 (11).

Published Version: 10.1111/ELE.13870

Publisher: Wiley

Permanent Link: https://hdl.handle.net/1721.1/148636

Version: Original manuscript: author’s manuscript prior to formal peer review

Terms of use: http://creativecommons.org/licenses/by-nc-sa/4.0/

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Untangling the complexity of priority effects in multi-species communities

Chuliang Song[1,2,3], Tadashi Fukami[4], Serguei Saavedra[1]

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

[2]Department of Biology, McGill University,
1205 Dr. Penfield Avenue, Montreal, H3A 1B1 Canada

[3] Department of Ecology and Evolutionary Biology, University of Toronto,
25 Willcocks Street, Toronto, Ontario M5S 3B2 Canada

[4]Biology Department, Stanford University, Stanford,
California 94305 USA

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Priority effects arise when the history of species arrival influences local species interactions, thereby affecting the composition of ecological communities. The outcome of some priority effects may be more difficult to predict than others, but this possibility remains to be fully investigated. Here, we provide a graph-based, non-parametric, theoretical framework to understand the classification of priority effects and the predictability of multi-species communities. We show that we can classify priority effects by decomposing them into four basic dynamical sources: the number of alternative stable states, the number of alternative transient paths, the length of composition cycles, and the interaction between alternative stable states and composition cycles. Although the number of alternative stable states has received most of the attention, we show that the other three sources can contribute more to the predictability of community assembly, especially in small communities. We discuss how this theoretical framework can guide new experimental studies.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The history of community assembly, especially the order of species arrival, can be highly stochastic (Gould, 1990; Sprockett *et al.*, 2018), making it hard to predict which species persist (Fukami, 2015). For example, arrival order dictates one of the three well-studied outcomes of 2-species competition in the Lotka-Volterra model (Figure 1A) (Case, 2000): (I) coexistence (two species coexist regardless of who arrives first); (II) deterministic exclusion (the same species always excludes the other regardless of who arrives first); and (III) history-dependent exclusion (the species that arrives first excludes the other). Though not as widely recognized, three other outcomes are also possible in 2-species competition: (IV) two species eventually coexist even though it requires a specific arrival order; (V) only one species survives or two species coexist depending on who arrives first; and (VI) only the species that arrives late survives (Warren *et al.*, 2003; Amor *et al.*, 2020; Angulo *et al.*, 2020). Of these six scenarios, (III), (V) and (VI) all represent history-dependent priority effects. The common omission of cases (V) and (VI) in theoretical studies is a consequence of assuming that model parameters are fixed (Rudolf, 2019; Zou & Rudolf, 2020). In reality, interaction strengths can be history-dependent (Rasmussen *et al.*, 2014; Poulos & McCormick, 2014; Vannette & Fukami, 2017; Carter & Rudolf, 2019; Sniegula *et al.*, 2019), increasing the scope of priority effects.

The number of possible outcomes quickly increases as more species are considered. Even with just three species, 4,1979 outcomes are theoretically possible, the majority of which are priority effects (Figure 1D). The super-exponential increase in the number of possible outcomes with the increasing number of species (Figure 1D) makes the prospect of predicting community assembly daunting. It may still be possible, however, to understand the degree of predictability of community assembly by considering different types of priority effects (Margalef, 1973). Some priority effects may reduce community predictability more than others. Such understanding of predictability is not only of basic interest to ecologists, but can also aid applications of community ecology for ecosystem management, including ecological restoration, biological control, and the medical treatment of the gut microbiome (Rohr *et al.*, 2020). Yet, a theoretical framework for quantifying the predictability of community assembly in this context is largely lacking.

Here, we propose an information-based metric to measure the predictability of outcomes generated by possible assembly dynamics, with a focus on the influence of priority effects. We characterize similarities among priority effects in terms of their contribution to community predictability. Then, we show that this classification can be explained based on four basic dynamical sources: the number of alternative stable states, the number of alternative transient paths, the length of composition cycles, and the interaction between alternative stable states and composition

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cycles. Finally, we discuss how our results can be used to guide experimental studies and provide new testable hypotheses about the importance of priority effects for the assembly of ecological communities.

A non-parametric graph-based approach

To capture the full diversity of priority effects, we introduce a non-parametric graph-based approach. This approach maps any assembly dynamics uniquely onto an assembly graph, where nodes represent combinations of coexisting species and directed links represent how species combinations change when a new species invades. We illustrate this approach with a generic 2-species community, where species are denoted as A and B (Fig. 1B). This illustrative assembly graph has $2^2 - 1 = 3$ nodes: $\{A\}$, $\{B\}$, and $\{A, B\}$—which represent all possible species combinations. The link starting from community (node) $\{A\}$ is generated by the invasion of species B, which can lead to all three possible species combinations (similarly for community $\{B\}$). In turn, community $\{A, B\}$ does not have any directed link since all species are present in this community. For example, case (II) in Figure 1B shows the graph representation of deterministic exclusion—invasion from species B into community $\{A\}$ leads to community $\{B\}$, whereas invasion from species A into community $\{B\}$ also leads to community $\{B\}$. Note that there are in total $3 \times 3 = 9$ assembly dynamics without considering which species is named as $\{A\}$ or $\{B\}$ (which is arbitrary). Formally, this only considers *topologically unique* assembly dynamics (graphs), where an assembly graph is unique up to the ordering of species labels. Thus, in a 2-species community, there are only 6 topologically unique assembly dynamics (Fig. 2B).

We now generalize our approach to multi-species communities. For simplicity, we present this extension with a generic 3-species community, where species are denoted as A, B, and C. The assembly graph has $2^3 - 1 = 8$ nodes, representing all possible species combinations. In a community with a single species ($\{A\}$, $\{B\}$, or $\{C\}$), there are two outgoing links representing invasions by the other two species, respectively. For instance, in community $\{A\}$, one of the two outgoing links represents the invasion by species B, which can lead to all three possible combinations of species A and B ($\{A\}$, $\{B\}$, or $\{A,B\}$). Then, in a community with two species ($\{A,B\}$, $\{A,C\}$, or $\{B,C\}$), there is one outgoing link representing the invasion by the only species not present in the community. That is, in community $\{A,B\}$, the only outgoing link represents the invasion by species C, which can lead to all eight possible combinations of species A, B, and C. Note that community size may decrease after an invasion. Finally, in the community with all three species ($\{A,B,C\}$), there is no outgoing link since all species are present. In total, there are 41,979 topologically different assembly dynamics in a 3-species communities.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The same procedure above applies to a community with an arbitrary number of species. The assembly graph of a generic S-species community has $(2^S - 1)$ nodes. The nodes representing communities with n species have $(S - n)$ outgoing links. Each of the associated outgoing links from the node with n species can possibly lead to $(2^{(n+1)} - 1)$ nodes. Then, out of the $(2^{(n+1)} - 1)$ nodes, only one node contains $(n + 1)$ species, $\binom{S}{n+1}$ nodes contain n species, while all the other nodes contain less than n species. As a first-order of approximation, the diversity (number) of topologically different assembly dynamics can be calculated as (see Appendix A for derivation):

$$\frac{\prod_{n=1}^{S} \left(2^{n+1} - 1\right)^{(S-n)\binom{S}{n}}}{S!} \sim \frac{2^{2^{S-2}S^2} e^S}{\sqrt{2\pi} S^{S+\frac{1}{2}}}. \tag{1}$$

Figure 1D shows how the diversity of priority effects scales with the number of species. Note that already with 6 species, the diversity is significantly greater than the total number of atoms in the entire universe (Kragh, 2003).

Priority effects and the predictability of community assembly

Priority effects are assembly dynamics that are not fully predictable. To quantity this lack of predictability owing to priority effects, it is necessary to define both the pool of possible assembly histories for a given community and the type of uncertainty to analyze (Fig. 2). Focusing on the pool of assembly histories, if an infinite number of invasions is possible, the pool of assembly histories is also infinite. While this assumption is typically applied to allow statistical convergence, it is a rather strong assumption that is often not met (Hubbell, 1997; Capitán *et al.*, 2009; Serván & Allesina, 2020). Here, we assume that each species invades m times. For example, in a 2-species community, if species can only invade $m = 1$ time, the pool of possible assembly histories consists of only two assembly histories: $\underline{AB}$ (meaning species A invades first and then species B invades) and $\underline{BA}$ (Fig. 2B). But if species can invade $m = 2$ times in a 3-species community, the pool of possible assembly histories consists of $\frac{6!}{2!2!2!} = 90$ assembly histories. For example, $\underline{ACBCAB}$ is a possible assembly history, meaning that species A invades first, then species C invades, and so on until species B invades last. By changing the number of invasion attempts (m), it is also possible to answer how large m has to be to effectively generate the same effects as infinite invasions. Similarly, our framework can be easily adapted to incorporate more ecological complexity, such as that some species arrive with higher frequency than others (see Appendix B for details).

Shifting our focus to the type of uncertainty to analyze, given a pool of assembly histories, it is possible to measure two types of uncertainties related to community composition. The first type is the uncertainty associated with the final community composition across all assembly

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

histories. The second type is the uncertainty associated with transient community compositions along all assembly histories. Here, to quantify the predictability of assembly dynamics, we adopt a normalized, information entropy metric (Rohr *et al.*, 2016). Although there are many alternative uncertainty metrics (Vellend, 2016), information entropy has been useful to quantify and explain different ecological processes (O'Connor *et al.*, 2019; Marleau *et al.*, 2020; Zu *et al.*, 2020; Margalef, 1973). We define the predictability of an assembly dynamics as

$$\text{predictability} := 1 - \text{entropy}\,(\text{uncertainties}) = 1 + \frac{\sum_i P(x_i)\log(P(x_i))}{\log(2^S - 1)}, \qquad (2)$$

where x_i is a species combination and $P(x_i)$ is the probability that combination x_i occurs (Fig. 2A). The entropy is normalized to $[0, 1]$ to ensure interpretability across different community sizes. Thus, a predictability of one (resp. zero) implies that that there is no (resp. full) uncertainty about the assembly dynamics.

To illustrate this measure, let us consider a 2-species community with assembly dynamics defined by case (IV) shown in Figure 1B, where each species invades only once. In this example (Fig. 2B), if the assembly history $\underline{AB}$ takes place, the trajectory of the community is from $\{A\} \to \{B\}$. Instead, if assembly history $\underline{BA}$ takes place, the trajectory is from $\{B\} \to \{A, B\}$. Therefore, the predictability of the final community composition is given by $\text{predictability}(\{B\}, \{A, B\}) = 1 + \frac{.5\log(.5) + .5\log(.5)}{\log(3)} = 0.37$. Similarly, the predictability of the assembly trajectory (the temporary community compositions) are $\text{predictability}(\{A\}, \{B\}) = .37$ and $\text{predictability}(\{B\}, \{A, B\}) = .37$. Figure 2B illustrates the process of quantifying this predictability. The agreement between the two types of predictability is not a coincidence. Figure 3A shows the strong correlation between the two types of predictability across all possible assembly dynamics in 3-species communities. The level of correlation is invariant across different pools of potential assembly histories. Thus, without loss of generality, hereafter our results are based on the first type of predictability (i.e., on the final community composition). Communities become more predictable when species are allowed to invade multiple times. Figure 3B shows how the distribution of predictability across all possible assembly dynamics (in 3-species communities) increases as a function of the number of invasion attempts.

Classifying priority effects

As shown above, there are significant differences in the predictability of priority effects. Here, we show that we can decompose community predictability into four basic sources: the number of alternative stable states, the number of alternative transient paths, the length of composition

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cycles, and the interaction between alternative stable states and composition cycles.

Alternative stable states occur when a community has more than one stable composition (left column of Figure 4A) (Gilpin & Case, 1976; Schröder *et al.*, 2005; Schooler *et al.*, 2011). Alternative transient states occur when there are more than one assembly histories (or trajectories) from the founding species to the stable states (middle column of Fig. 4A) (Fukami & Nakajima, 2011; Sarneel *et al.*, 2019). Composition cycles occur when the assembly histories involve cyclic seques of community compositions (right column of Fig. 4A) (Schreiber & Rittenhouse, 2004; Fox, 2008). Our graph-based approach can identify these dynamical sources as topological features. That is, alternative stable states arise when the assembly graph has more than one sink (nodes that have incoming links but no outgoing link); alternative transient states arise when more than one directed path exist from single species to a sink; while composition cycles occur when directed cycles exist in the assembly graph. These three dynamical sources have already been hypothesized to be major drivers of priority effects (Fukami, 2015).

The three dynamical sources are not mutually exclusive, implying that assembly dynamics can potentially exhibit all of these sources. For example, the assembly graph of a 3-species community shown in Figure 4B has two alternative stable states ($\{A, C\}$ and $\{C\}$), one transient path for each stable state ($\{A\} \to \{A, C\}$ and $\{A\} \to \{B\} \to \{A, C\}$), and one composition cycle of length 3 ($\{B\} \to \{A, B\} \to \{B, C\} \to \{B\}$). However, the picture is not complete with only these three sources. Figure 4C shows another example of an assembly graph in a 3-species community with the same types and number of dynamical sources as shown in Figure 4B. Nevertheless, the assembly dynamics in Figure 4C has lower predictability than the one shown in Figure 4B. The difference between these two cases is given by the fact that only in Figure 4B, the assembly dynamics exhibits the possibility to escape from a cycle to a stable state (i.e., the trajectory can escape the composition cycle ($\{B\} \to \{A, B\} \to \{B, C\} \to \{B\}$) into a stable state ($\{C\}$). Because cycles are less predictable than stable states in general, this possibility to escape can equally increase the predictability of assembly dynamics.

Therefore, to classify priority effects by their predictability, we propose to use the four dynamical sources (topological features): alternative stable states, alternative transient states, composition cycles, and the interaction between cycles and stable states. To carry out this classification, we used a neural network with three hidden layers of ReLU activation (see Appendix C for detailed description). Focusing on 3-species communities as a case study, we found that the classification works better for assembly dynamics with multiple invasions. Specifically, the classification displayed a correlation of 0.57 and 0.98 for single ($m = 1$; Fig 5A) and multiple ($m = 15$; Fig 5B) invasions, respectively. The explanatory (classification) power of our four topological features combined increases with the number of invasions and reaches a plateau around $m = 8$ invasions

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Figure 5C). These results are qualitatively the same for larger communities (Appendix D).

Finally, we used a regression-based scheme (Grömping *et al.*, 2006) to quantify the relative importance (contribution) of each topological feature (dynamical source) to the classification power of priority effects. Figure 5D shows that the relative importance of topological features changes with the number of species invasions. Additionally, this figure shows that the relative importance of the number of stable states remains constant across invasion times, while the relative importance increases for both the number of transient paths and the number of interactions between cycles and stable states, but it decreases for the length of cycles. Moreover, this result shows that the relative importance of the number of stable states can be smaller than the combined importance of the other three sources (similar results are found for larger communities, see Appendix D), revealing the need to account for these other topological features to better understand the existence of priority effects in ecological communities.

From theory to testable hypotheses

While theoretical analysis like what we have done in this paper so far can reveal the full landscape of assembly dynamics, only empirical evidence can discern which assembly dynamics are possible and which ones are not given the internal and external constraints acting on ecological communities (Medeiros *et al.*, 2021). For example, Warren *et al.* (2003) empirically mapped the assembly graph of a 6-species protist community and Drake (1991) recorded all the necessary data to empirically map the assembly graph of a 3-species algae community. In Appendix E, we provide a full analysis of these two assembly graphs following our methodology. Unfortunately, there has been little additional evidence about these assembly dynamics (Fukami, 2015). The lack of empirical studies regarding these dynamics is not surprising given the challenges involved in performing detailed experiments and their under-appreciated potential to answer central questions in community ecology (Vellend, 2016). In this regard, we discuss below possible empirical designs that can facilitate the inference of assembly graphs in ecological communities.

Our theoretical framework illustrates that priority effects can take a large set of possibilities, most of which are outside the realm of traditional theoretical predictions. We hypothesize that such a larger diversity of priority effects can be found in empirical communities without the need to map every single dynamics. To test this hypothesis, a direct approach to infer assembly graphs should involve knowing which species combinations can persist and then mapping how these combinations change after the introduction of a new species. Although this direct approach may seem like an exhaustive combinatorial design, the experimental workload can be generally much

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

lower. For example, let us focus on the persistent species combinations (i.e., knowing the nodes in assembly graphs). While a community with S species has $2^S - 1$ potential species combinations, it has been shown that only a small fraction of such combinations can persist (Angulo *et al.*, 2020). Because of the sparsity of these persistent combinations, it can be possible to infer all persistent combinations using a minimum of $S + 1$ experiments in a community with S species (Maynard *et al.*, 2020). Moreover, if one is only interested in bottom-up assembly, then persistent combinations that are unreachable via introduction of single species do not need to be mapped to construct the corresponding assembly graph. In Appendix F, we analyzed all 3-species assembly dynamics and found that more than 99% of assembly dynamics contain unreachable combinations, which make the inference problem more feasible.

While alternative stable states have been the most studied consequence of priority effects, our framework indicates that the three other consequences of priority effects can have a stronger contribution to community predictability, especially in small communities. This possibility can be tested by exploiting the strong constraints between dynamical sources and predictability of priority effects. Specifically, to study these additional dynamical sources, we can use alternative computational approaches based on pairwise interaction strengths inferred from experiments on 2-species communities (Case, 2000). Such empirical data are increasingly available spanning a wide range of study systems, such as annual plants (Godoy *et al.*, 2014; Kraft *et al.*, 2015), perennial plants (Uricchio *et al.*, 2019; Song *et al.*, 2020), and microbial systems (Xiao *et al.*, 2017; Kehe *et al.*, 2020). Thus, the empirical assembly graphs can be computationally mapped with empirically parameterized population dynamics models. However, it is worth remembering that a defining feature of priority effects is that interaction strengths are likely to be dependent on assembly history (Fukami, 2015; Song *et al.*, 2018), questioning the validity of inferences based on fixed interaction strengths. Because we know little about how variable these interaction strengths are (Park, 1954), it would be best to combine experimental and computational approaches. That is, the differences between the observed assembly graph in the experimental approach and the inferred graph from the computational approach may provide clues as to how variable interaction strengths are due to the assembly history.

Discussion

In the story “The Garden of Forking Paths,” Jorge Luis Borges envisioned a labyrinth where divergence takes place in time rather than space, and where different paths sometimes lead to the same conclusion. Similarly, an assembly history on an assembly graph can be thought of as a forking path since it also creates a temporal trajectory of species compositions that can

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

lead to the same final composition. By borrowing tools from graph theory, we have provided a non-parametric framework to survey the full landscape of the labyrinth of priority effects. This framework has allowed us to enumerate all possible assembly dynamics operating in a community, quantify the predictability of these dynamics, and classify priority effects.

Taking this approach, we have shown that priority effects can differ in terms of their contribution to community predictability. Following previous work (Fukami, 2015), we have shown that four dynamical sources can be expressed as topological features within our graph-based approach to know predictability of assembly dynamics. We have found that both the frequency and combination of the four topological features are good predictors of how many outcomes to expect. While the number of alternative stable states has received most of the attention (Schröder *et al.*, 2005; Serván & Allesina, 2020), we show that the other three sources can contribute more to community predictability, especially in small communities. We also have found that the explanatory power of these four topological features increases when species attempt to invade multiple times. Laboratory experiments that only allow single invasions may underestimate community predictability (Drake, 1991; Lawler & Morin, 1993; McGrady-Steed *et al.*, 1997).

Our non-parametric graph-based approach is not intended to replace the parametric model-based approach. Parametric models are irreplaceable tools to understand priority effects, but non-parametric approaches are more flexible for accommodating different theoretical tools (Barabás *et al.*, 2018; Arnoldi *et al.*, 2019; Pande *et al.*, 2020; Spaak & De Laender, 2020). For example, while our approach uncovers three other types of priority effects in 2-species communities (Figure 1) that are not covered in classic parametric Lotka-Volterra approaches (Fukami *et al.*, 2016; Ke & Letten, 2018), all priority effects can occur in parametric models by integrating processes where the assembly history affects parameters.

Acknowledgement: We thank Lucas P. Medeiros, Hengxing Zou, and Pengjuan Zu for discussions that improved this work.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

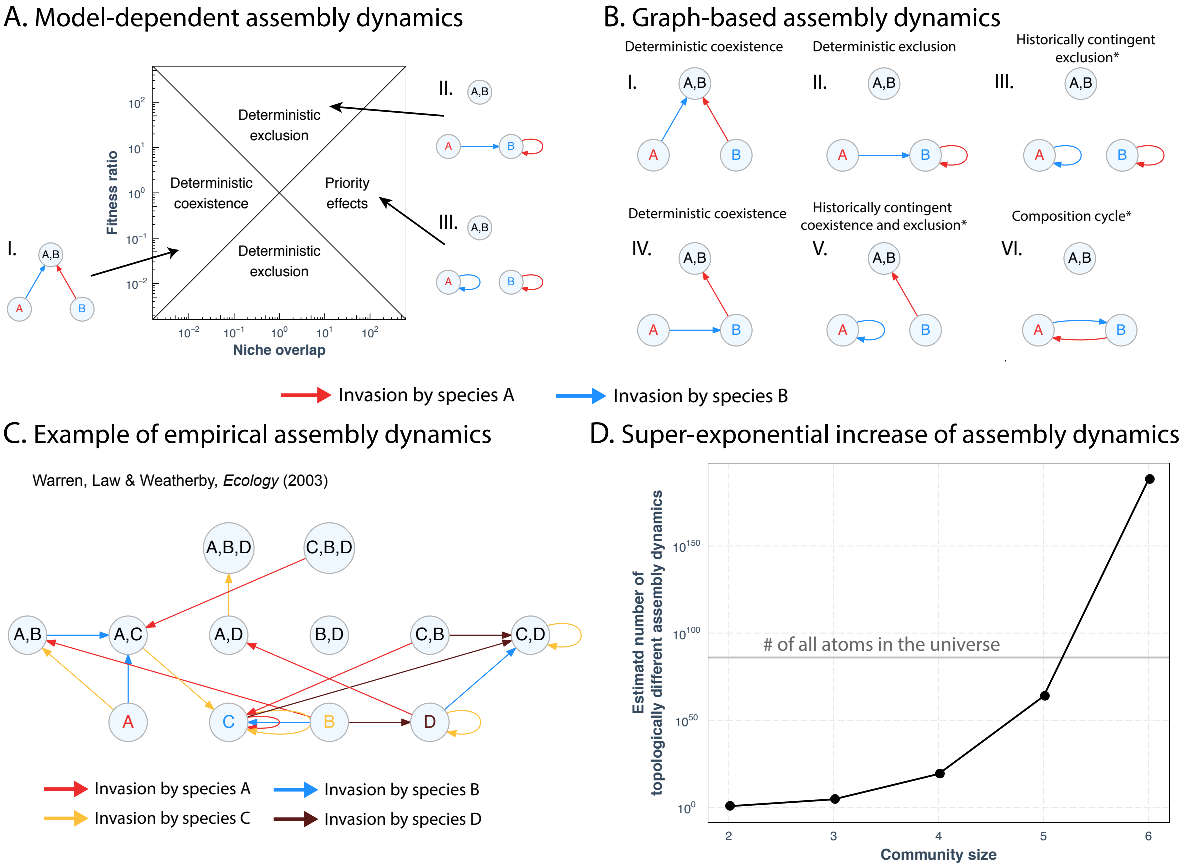

Figure 1: **Non-parametric graph-based approach.** Panel (**A**) presents the traditional model-based phase diagram of assembly dynamics in 2-species communities using coexistence theory. The balance between niche overlap (x axis) and fitness ratio (y axis) determines the types of assembly dynamics: (I) deterministic existence, (II) deterministic exclusion, or (III) priority effects. Panel (**B**) presents all six types of possible assembly dynamics (model-free) using the graph-based approach. In an assembly graph, the nodes represent the species combinations, while the links represent how the species combination change into another combination after a non-resident species invades. Cases (III), (V), and (VI) are all priority effects (marked with $*$ symbol) as at least two different assembly histories can lead to different compositions. Panel (C) shows an empirical assembly graph in a 4-species protist community extracted from Warren *et al.* (2003). Species A stands for *Blepharisma japonicum*, species B for *Colpidium striatum*, species C for *Paramecium caudatum*, , and species D for *Tetrahymena pyriformis*. Panel (D) plots how the diversity of priority effects increases with larger community size. We find a super-exponential increase of diversity of topologically different assembly dynamics (Eqn. 1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

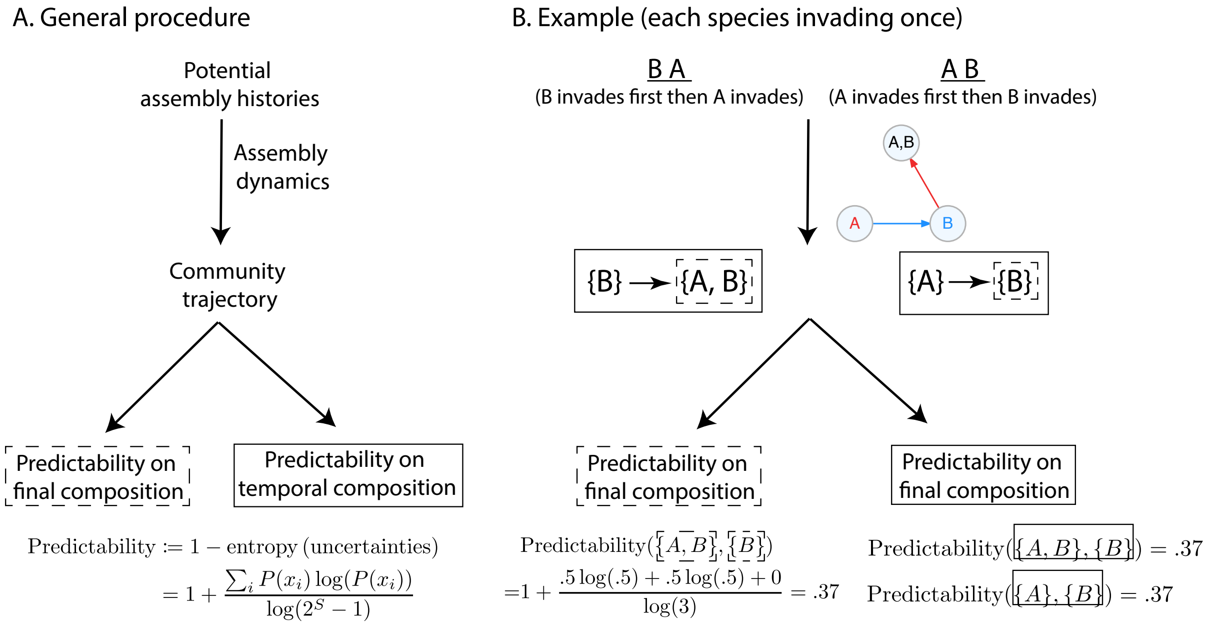

Figure 2: **Quantifying the predictability of assembly dynamics.** Panel (**A**) presents the general procedure used for the calculation: First, based on a given uncertainty (either final composition or temporary compositions), we define a pool of potential assembly histories. Second, we need to know the assembly dynamics of the community. Third, each assembly history produces its community trajectory (how community composition changes with an invading species) based on the assembly dynamics. Last, we compute predictability on the final composition across the community trajectory (boxed with dashed lines) and on the temporal compositions along the communities trajectories (boxed with solid lines). The definition of predictability is based on normalized entropy (see the mathematical definition in Eqn. 2 or in the figure). Panel (**B**) illustrates this procedure with an example of a 2-species community. First, we define the pool of potential assembly histories as each species invades only once. This gives us two potential assembly histories: $\underline{BA}$ and $\underline{AB}$. Second, we choose the type (IV) assembly dynamics in 1B for illustration. Third, the community trajectories are $\{B\} \to \{A, B\}$ for the assembly history $\underline{BA}$ and $\{A\} \to \{B\}$ for the assembly history $\underline{AB}$. Last, the predictability on final composition is the predictability of the states $(\{A, B\}, \{B\})$—the final compositions in each community trajectory (boxed in dashed lines); and the predictability on temporal composition is the predictability of the states $(\{B\}, \{A, B\})$ for the assembly history $\underline{BA}$ and of the states $(\{A\}, \{B\})$ for the assembly history $\underline{AB}$ (boxed in solid lines). All calculations (predictability) equal to .37 in this example.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

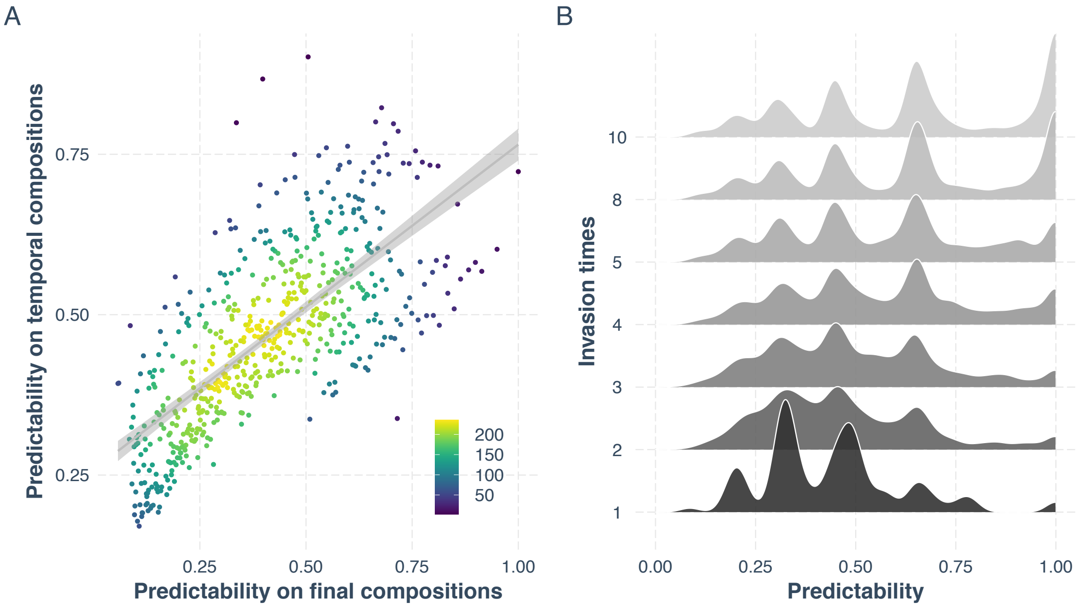

Figure 3: **Regularities of the predictability of priority effects**. Panel (**A**) shows the correlation (Pearson correlation = .72 with 95% confidence interval [.68, .75]) between the predictability of assembly dynamics on the final compositions (x axis) and on the temporary compositions (y axis). Recall that all assembly dynamics with predictability less than 1 correspond to priority effects. The color map represents the number of data points. Panel (**B**) plots the predictability distribution of assembly dynamics as a function of invasion times. The predictability (x axis) corresponds to the predictability on the final composition. Note that the distribution converges to a stationary distribution with increasing invasion times.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

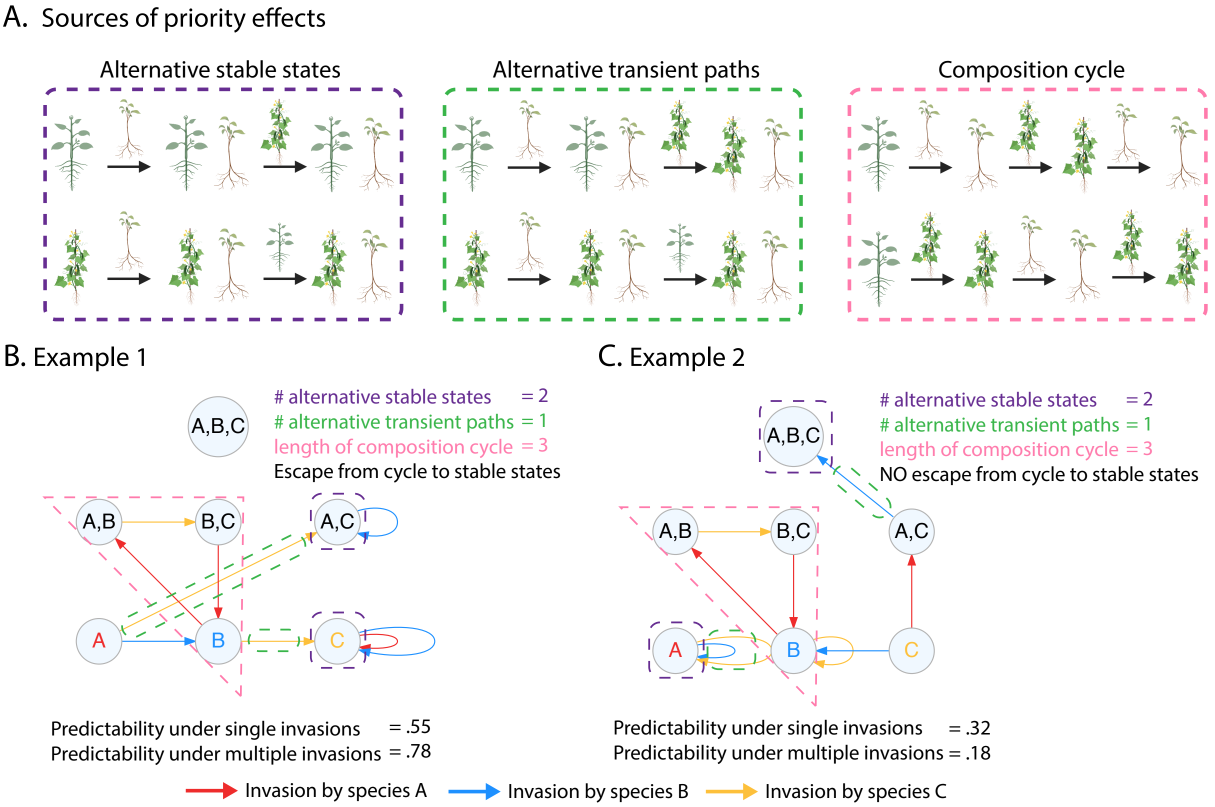

Figure 4: **Dynamical sources acting on priority effects.** Panel (**A**) presents the three known dynamical sources acting on priority effects (Fukami, 2015): alternative stable states, alternative transient paths, and composition cycles. We illustrate these sources with two plants (cartoons). Panels (**B**) and (**C**) illustrate how these dynamical sources can be represented as topological features of assembly graphs. Panel (**B**) shows an assembly graph in a 3-species community. It has two alternative stable states ($\{A, C\}$ and $\{C\}$; boxed in purple lines), one transient path for each stable state ($\{A\} \rightarrow \{A, C\}$ and $\{A\} \rightarrow \{B\} \rightarrow \{A, C\}$; boxed in green lines), and one composition cycle of length 3 ($\{B\} \rightarrow \{A, B\} \rightarrow \{B, C\} \rightarrow \{B\}$; boxed in purple lies). Panel (**C**) shows another assembly graph with identical topological features as the ones presented in Panel (**B**). It also has two alternative stable states ($\{A\}$ and $\{A, B, C\}$; boxed in purple lines), one transient path for each stable state ($\{A\} \rightarrow \{A\}$ and $\{C\} \rightarrow \{A, C\} \rightarrow \{A, B, C\}$; boxed in green lines), and one composition cycle of length 3 ($\{B\} \rightarrow \{A, B\} \rightarrow \{B, C\} \rightarrow \{B\}$; boxed in purple lies). However, the assembly graphs in Panels (**B**) and (**C**) have drastically different predictability. The cause of this difference is a missing fourth topological feature: escape from composition cycle to stable states. In Panel (**B**), the system can escape from the composition cycles to stable states (i.e., from $\{B\}$ to $\{C\}$). In Panel (**C**), the system cannot escape.

.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

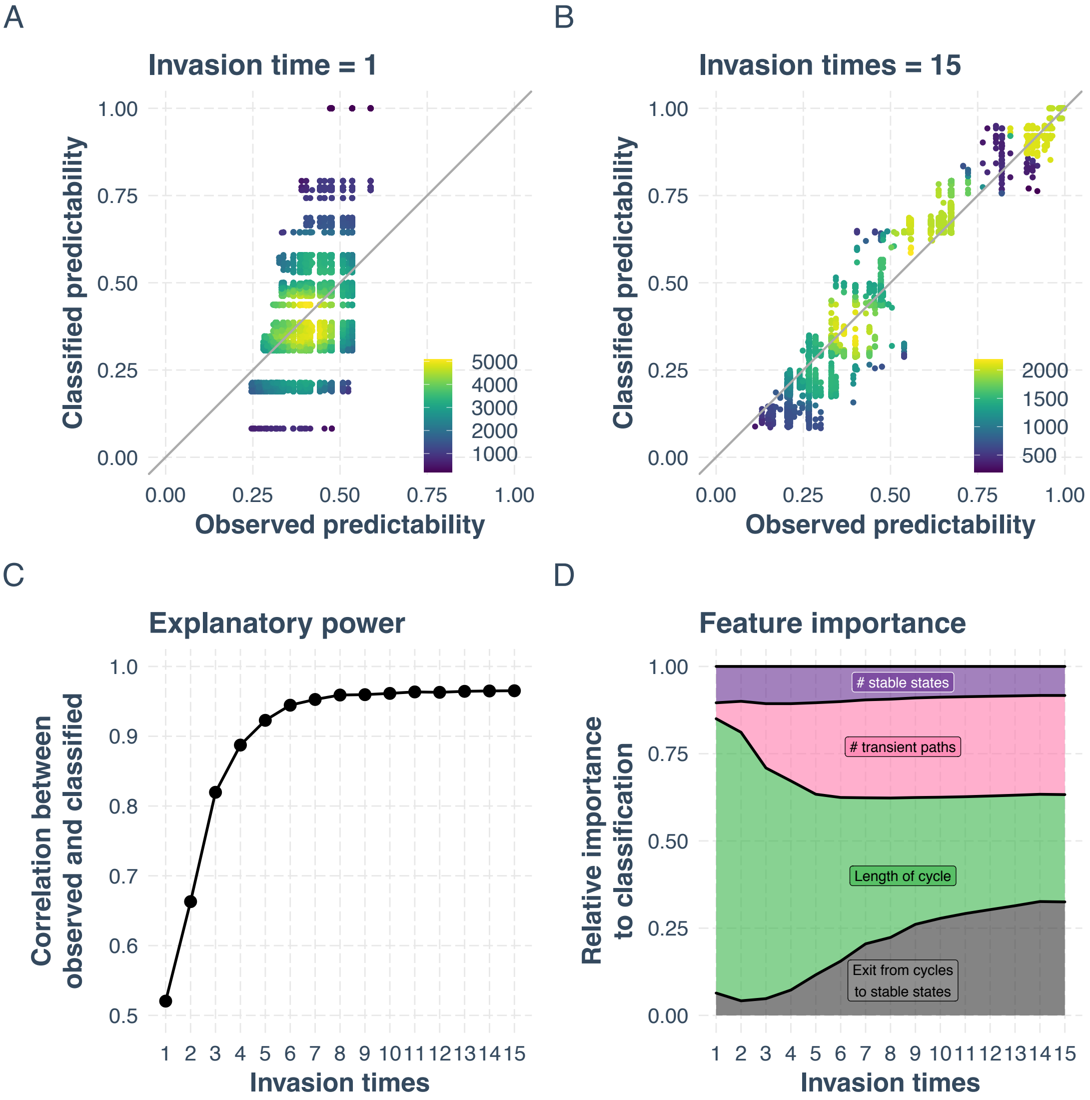

Figure 5: **Explaining the predictability of priority effects using topological features.** This figure illustrates the out-of-sample explanation power of the four dynamical sources (topological features) in classifying the predictability on the final composition of 3-species communities. We used 80% out of 41,979 assembly dynamics as the training set, while the remaining 20% is used as the test set (the classification). We used a neural network to achieve the maximum classification power (see Appendix C for details). Panel (**A**) shows the explanation power for the pool of assembly histories where each species invades only once. The x axis represents the observed predictability while the y axis represents the classified predictability (from the test set) using the four topological features. The color of the point indicates how many assembly graphs are represented by the point. The gray line is the 45-degree line, representing that the observed predictability equals the classified predictability. Panel (**B**) shows the explanation power for the pool of assembly histories where each species invades 15 times. Panel (**C**) shows how the explanatory power increases with invasion times and reaches a plateau fast. We indicate the explanatory power using the correlation between the observed and classified predictability (using different pools of potential assembly histories). Panel (**D**) shows the relative importance of each topological feature (dynamical source) to the explanation power.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Amor, D. R., Ratzke, C. & Gore, J. (2020). Transient invaders can induce shifts between alternative stable states of microbial communities. *Science Advances*, 6, eaay8676.

Angulo, M. T., Kelley, A., Montejano, L., Song, C. & Saavedra, S. (2020). Coexistence holes characterize the assembly and disassembly of multispecies systems. *BioRxiv*.

Arnoldi, J.-F., Barbier, M., Kelly, R., Barabás, G. & Jackson, A. L. (2019). Fitness and community feedbacks: the two axes that drive long-term invasion impacts. *bioRxiv*, 705756.

Barabás, G., D'Andrea, R. & Stump, S. M. (2018). Chesson's coexistence theory. *Ecological Monographs*, 88, 277–303.

Capitán, J. A., Cuesta, J. A. & Bascompte, J. (2009). Statistical mechanics of ecosystem assembly. *Physical Review Letters*, 103, 168101.

Carter, S. K. & Rudolf, V. H. (2019). Shifts in phenological mean and synchrony interact to shape competitive outcomes. *Ecology*, 100, e02826.

Case, T. J. (2000). *An Illustrated Guide to Theoretical Ecology*. Oxford University Press.

Drake, J. A. (1991). Community-assembly mechanics and the structure of an experimental species ensemble. *The American Naturalist*, 137, 1–26.

Fox, J. W. (2008). Testing whether productivity mediates the occurrence of alternate stable states and assembly cycles in a model microcosm system. *Oikos*, 117, 1153–1164.

Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, 46, 1–23.

Fukami, T., Mordecai, E. A. & Ostling, A. (2016). A framework for priority effects. *Journal of Vegetation Science*, 27, 655–657.

Fukami, T. & Nakajima, M. (2011). Community assembly: alternative stable states or alternative transient states? *Ecology Letters*, 14, 973–984.

Gilpin, M. E. & Case, T. J. (1976). Multiple domains of attraction in competition communities. *Nature*, 261, 40–42.

Godoy, O., Kraft, N. J. & Levine, J. M. (2014). Phylogenetic relatedness and the determinants of competitive outcomes. *Ecology Letters*, 17, 836–844.

Gould, S. J. (1990). *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company.

Grömping, U. *et al.* (2006). Relative importance for linear regression in r: the package relaimpo. *Journal of Statistical Software*, 17, 1–27.

Hubbell, S. P. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral reefs*, 16, S9–S21.

Ke, P.-J. & Letten, A. D. (2018). Coexistence theory and the frequency-dependence of priority effects. *Nature Ecology & Evolution*, 2, 1691.

Kehe, J., Ortiz, A., Kulesa, A., Gore, J., Blainey, P. C. & Friedman, J. (2020). Positive interactions are common among culturable bacteria. *BioRxiv*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Kraft, N. J., Godoy, O. & Levine, J. M. (2015). Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, 112, 797–802.

Kragh, H. (2003). Magic number: A partial history of the fine-structure constant. *Archive for History of Exact Sciences*, 57, 395–431.

Lawler, S. P. & Morin, P. J. (1993). Food web architecture and population dynamics in laboratory microcosms of protists. *The American Naturalist*, 141, 675–686.

Margalef, R. (1973). *Information theory in ecology*. Dept. of the Environment, Fisheries Research Board of Canada, Great Lakes Biolimnology Laboratory.

Marleau, J. N., Peller, T., Guichard, F. & Gonzalez, A. (2020). Converting ecological currencies: Energy, material, and information flows. *Trends in Ecology & Evolution*.

Maynard, D. S., Miller, Z. R. & Allesina, S. (2020). Predicting coexistence in experimental ecological communities. *Nature Ecology & Evolution*, 4, 91–100.

McGrady-Steed, J., Harris, P. M. & Morin, P. J. (1997). Biodiversity regulates ecosystem predictability. *Nature*, 390, 162–165.

Medeiros, L. P., Boege, K., del Val, E., Zaldivar-Riverón, A. & Saavedra, S. (2021). Observed ecological communities are formed by species combinations that are among the most likely to persist under changing environments. *The American Naturalist*, 197, E17–E29.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

O'Connor, M. I., Pennell, M., Altermatt, F., Matthews, B., Melian, C. & Gonzalez, A. (2019). Principles of ecology revisited: Integrating information and ecological theories for a more unified science. *Frontiers in Ecology and Evolution*, 7, 219.

Pande, J., Fung, T., Chisholm, R. & Shnerb, N. M. (2020). Mean growth rate when rare is not a reliable metric for persistence of species. *Ecology letters*, 23, 274–282.

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

Poulos, D. E. & McCormick, M. I. (2014). Who wins in the battle for space? the importance of priority, behavioural history and size. *Animal Behaviour*, 90, 305–314.

Rasmussen, N. L., Van Allen, B. G. & Rudolf, V. H. (2014). Linking phenological shifts to species interactions through size-mediated priority effects. *Journal of Animal Ecology*, 83, 1206–1215.

Rohr, J. R., Civitello, D. J., Halliday, F. W., Hudson, P. J., Lafferty, K. D., Wood, C. L. & Mordecai, E. A. (2020). Towards common ground in the biodiversity – disease debate. *Nature Ecology & Evolution*, 4, 24–33.

Rohr, R. P., Saavedra, S., Peralta, G., Frost, C. M., Bersier, L.-F., Bascompte, J. & Tylianakis, J. M. (2016). Persist or produce: A community trade-off tuned by species evenness. *The American Naturalist*, 188, 411–422.

Rudolf, V. H. (2019). The role of seasonal timing and phenological shifts for species coexistence. *Ecology letters*, 22, 1324–1338.

Sarneel, J. M., Hefting, M. M., Kowalchuk, G. A., Nilsson, C., Van der Velden, M., Visser, E. J., Voesenek, L. A. & Jansson, R. (2019). Alternative transient states and slow plant community responses after changed flooding regimes. *Global Change Biology*, 25, 1358–1367.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.29.437541; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Schooler, S. S., Salau, B., Julien, M. H. & Ives, A. R. (2011). Alternative stable states explain unpredictable biological control of salvinia molesta in kakadu. *Nature*, 470, 86–89.

Schreiber, S. J. & Rittenhouse, S. (2004). From simple rules to cycling in community assembly. *Oikos*, 105, 349–358.

Schröder, A., Persson, L. & De Roos, A. M. (2005). Direct experimental evidence for alternative stable states: a review. *Oikos*, 110, 3–19.

Serván, C. A. & Allesina, S. (2020). Tractable models of ecological assembly. *bioRxiv*.

Sniegula, S., Golab, M. J. & Johansson, F. (2019). Size-mediated priority and temperature effects on intra-cohort competition and cannibalism in a damselfly. *Journal of Animal Ecology*, 88, 637–648.

Song, C., Altermatt, F., Pearse, I. & Saavedra, S. (2018). Structural changes within trophic levels are constrained by within-family assembly rules at lower trophic levels. *Ecology Letters*, 21, 1221–1228.

Song, C., Uricchio, L. H., Mordecai, E. A. & Saavedra, S. (2020). Understanding the emergence of contingent and deterministic exclusion in multispecies communities. *BioRxiv*.

Spaak, J. W. & De Laender, F. (2020). Intuitive and broadly applicable definitions of niche and fitness differences. *Ecology Letters*, 23, 1117–1128.

Sprockett, D., Fukami, T. & Relman, D. A. (2018). Role of priority effects in the early-life assembly of the gut microbiota. *Nature Reviews Gastroenterology & Hepatology*, 15, 197–205.

Uricchio, L. H., Daws, S. C., Spear, E. R. & Mordecai, E. A. (2019). Priority effects and nonhierarchical competition shape species composition in a complex grassland community. *The American Naturalist*, 193, 213–226.

Vannette, R. L. & Fukami, T. (2017). Dispersal enhances beta diversity in nectar microbes. *Ecology Letters*, 20, 901–910.

Vellend, M. (2016). *The theory of ecological communities (MPB-57)*. Princeton University Press.

Warren, P. H., Law, R. & Weatherby, A. J. (2003). Mapping the assembly of protist communities in microcosms. *Ecology*, 84, 1001–1011.

Xiao, Y., Angulo, M. T., Friedman, J., Waldor, M. K., Weiss, S. T. & Liu, Y.-Y. (2017). Mapping the ecological networks of microbial communities. *Nature communications*, 8, 1–12.

Zou, H.-X. & Rudolf, V. H. (2020). Priority effects and season length shape long-term competition dynamics. *bioRxiv*.

Zu, P., Boege, K., Del-Val, E., Schuman, M. C., Stevenson, P. C., Zaldivar-Riverón, A. & Saavedra, S. (2020). Information arms race explains plant-herbivore chemical communication in ecological communities. *Science*, 368, 1377–1381.