

# Current Biology

## The development of ecological systems along paths of least resistance

### Highlights

- We define development as the appearance of taxa, as seen in ecological succession
- Development is more likely under a sequenced process than simultaneously
- Development is more likely under an ordered process than disorderly
- Biodiversity enhances the likelihood of fast-growing taxa appearing earlier

### Authors

Jie Deng, Otto X. Cordero, Tadashi Fukami, Simon A. Levin, Robert M. Pringle, Ricard Solé, Serguei Saavedra

### Correspondence

jiedeng@mit.edu (J.D.),  
rpringle@princeton.edu (R.M.P.),  
sersaa@mit.edu (S.S.)

### In brief

Why do ecological systems develop following a path from fast- to slow-growing species? Deng et al. theoretically show that this path is the most likely to emerge when species minimize environmental resistance for their survival. This work corroborates these findings using empirical data across a wide diversity of organisms and taxonomic levels.



## Report

# The development of ecological systems along paths of least resistance

Jie Deng,<sup>1,7,8,\*</sup> Otto X. Cordero,<sup>1</sup> Tadashi Fukami,<sup>2</sup> Simon A. Levin,<sup>3</sup> Robert M. Pringle,<sup>3,\*</sup> Ricard Solé,<sup>4,5,6</sup> and Serguei Saavedra<sup>1,6,\*</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

<sup>2</sup>Departments of Biology and Earth System Science, Stanford University, 371 Jane Stanford Way, Stanford, CA 94305, USA

<sup>3</sup>Department of Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

<sup>4</sup>Complex Systems Laboratory, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain

<sup>5</sup>Institució Catalana de Recerca i Estudis Avançats, Lluís Companys 23, 08010 Barcelona, Spain

<sup>6</sup>Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

<sup>7</sup>X (formerly Twitter): @jie\_eco

<sup>8</sup>Lead contact

\*Correspondence: [jiedeng@mit.edu](mailto:jiedeng@mit.edu) (J.D.), [rpringle@princeton.edu](mailto:rpringle@princeton.edu) (R.M.P.), [sersaa@mit.edu](mailto:sersaa@mit.edu) (S.S.)

<https://doi.org/10.1016/j.cub.2024.08.050>

## SUMMARY

A long-standing question in biology is whether there are common principles that characterize the development of ecological systems (the appearance of a group of taxa), regardless of organismal diversity and environmental context.<sup>1–11</sup> Classic ecological theory holds that these systems develop following a sequenced, orderly process that generally proceeds from fast-growing to slow-growing taxa and depends on life-history trade-offs.<sup>2,12,13</sup> However, it is also possible that this developmental order is simply the path with the least environmental resistance for survival of the component species and hence favored by probability alone. Here, we use theory and data to show that the order from fast- to slow-growing taxa is the most likely developmental path for diverse systems when local taxon interactions self-organize in light of environmental resistance. First, we demonstrate theoretically that a sequenced development is more likely than a simultaneous one, at least until the number of iterations becomes so large as to be ecologically implausible. We then show that greater diversity of taxa and life histories improves the likelihood of a sequenced order from fast- to slow-growing taxa. Using data from bacterial and metazoan systems,<sup>14–19</sup> we present empirical evidence that the developmental order of ecological systems moves along the paths of least environmental resistance. The capacity of simple principles to explain the trend in the developmental order of diverse ecological systems paves the way to an enhanced understanding of collective features of life.

## RESULTS AND DISCUSSION

### Ecological development

One of the most iconic thought experiments in the life sciences is what would happen if the tape of life on Earth were rewound and played again.<sup>1</sup> Would the same species return in the same or similar order? Or do historical accidents lead to unpredictable outcomes? This question is not just philosophical or counterfactual because developmental processes (the appearance of a group of taxa) undergone by complex living systems are continuously replicating the experiment.<sup>2</sup> Ecological systems ranging from microbes to forests are formed by self-organizing networks of interacting taxa, whose emergent behavior is crucially influenced by their history, changing biotic and abiotic environment, and chance.<sup>20</sup> From the primary development of microbiota in newborns to the secondary development of macrobiota in grasslands due to environmental perturbations or seasonal patterns, local tapes of life are constantly starting or being rewound. Such development over ecological timescales is analogous to evolutionary outcomes derived from similar developmental

processes, increasing the relevance to understand general basic processes.<sup>21–24</sup>

The ubiquitous collective pattern of orderly and sequenced development of ecological systems (also known as succession<sup>2</sup>) is widely considered to be one of the foundational problems of ecology,<sup>3–5</sup> and it remains central to contemporary research on biodiversity and environmental change.<sup>6–8</sup> Many mechanisms and trade-offs promoting development have been documented, predominantly in plants,<sup>2,12,13</sup> but these emerge at the interface of specific organismal traits (e.g., dispersal ability, resource requirements, and relative growth rates) and local systemic contexts (e.g., environmental stressors, prevalence of facilitative versus inhibitory interactions). As a result, fundamental questions remain unanswered about the generality of these mechanisms and the inevitability of collective patterns or particular developmental pathways.<sup>9–11</sup> A generalized theoretical framework capable of positing testable predictions about developmental trajectories across the tree of life with minimal assumptions about specific organismal traits and environmental contexts would aid in answering these questions and advance



the understanding of one of ecology's oldest and still-paramount challenges.<sup>25–27</sup>

One key insight is that the developmental order displayed by ecological systems is generally a process from fast-growing to slow-growing taxa.<sup>2,28,29</sup> Yet it remains unclear whether this collective pattern depends critically on the complex combination of specific life-history traits and trade-offs that are often strongly linked to growth rate (e.g., dispersal ability) or whether it might simply be the result of local taxon interactions following the path with the least environmental resistance for species survival (requiring fewer adaptations) and, hence, the configuration upon which systems will probabilistically converge.<sup>30–32</sup> If there is path convergence, it becomes critical to understand the role of individual traits in modulating such local taxon interactions and paths. Indeed, discovering minimal processes that can explain the development of diverse ecological systems can shed new light on the conditions under which the tape of life might have a relatively predictable sequence of events.<sup>5,33</sup> This is increasingly a practical concern, too, as conservationists seek to restore functional biodiversity in areas where species have been driven locally extinct<sup>34–36</sup> and as the essential roles of microbiome development in sustaining both human and environmental health become clearer.<sup>37</sup>

Formally, we define development as the self-organized process of moving from an empty set (a novel environment with no taxa) to a community of  $n$  taxa (a modified environment with a particular set of taxa) through a collection of different taxon subcommunities. Previous work in experimental and theoretical ecology has established foundational knowledge linking community composition with historical events, assembly and disassembly dynamics, and invasibility.<sup>38–47</sup> This body of work has assumed that the environments remain relatively constant throughout the development of ecological systems (i.e., no chance events).<sup>48</sup> However, the survival of species depends on historical accidents as well as on their capacity to modify and withstand changes in the abiotic and biotic environments (e.g., resource availability, competition, and temperature variation).<sup>49</sup> At intermediate levels of biological organization,<sup>50</sup> the possible solutions to these environmental challenges can be represented by the possible sets of interacting populations (i.e., communities) that partition or modify the resources available at a given place and time.<sup>29,49</sup> Consequently, it becomes imperative not only to understand which environments give rise to developmental processes but also which constructed environments enabled the persistence of such processes.<sup>23,51</sup>

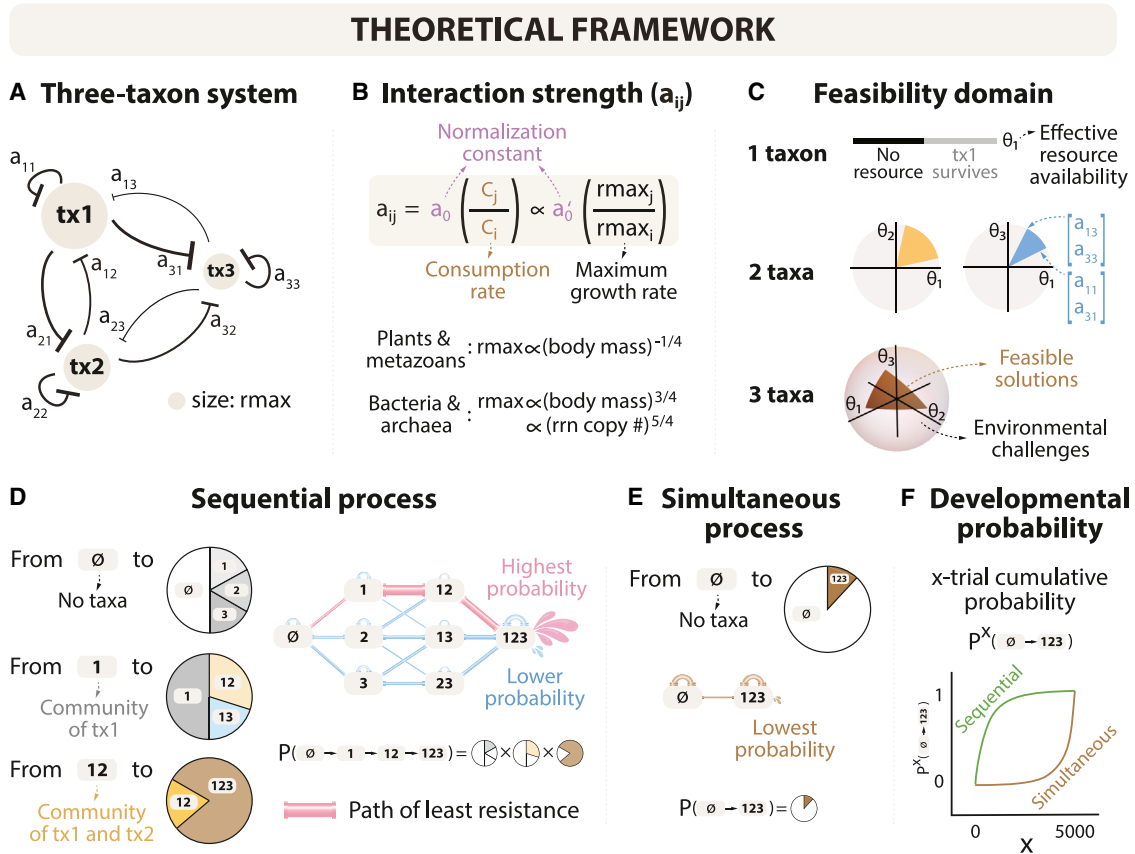
Here, we focus on the developmental order of ecological systems (i.e., knowing which taxa should be expected to appear before others regardless of whether the former taxa can survive across the entire developmental process). That is, we are not focusing on the actual step-by-step assembly and disassembly of an ecological system, but simply on the expected order of taxon appearance. Specifically, our research goal is to investigate the extent to which a minimal model rooted in local taxon interactions, but without invoking the complexity of multiple life-history traits, can explain the developmental order from fast-growing to slow-growing taxa generally observed in nature.

We operationalize development using a general Markov chain process moving from an empty set  $\emptyset$  to a given set of  $n$  taxa (Figure 1; see STAR Methods for more details). Conceptually, we assume that development is possible by the compatibility

(feasibility) of environmental conditions with current and future taxon communities. In our theory, development takes the path of least environmental resistance (i.e., higher compatibility of environmental conditions across a sequence of taxon communities). As a metaphor, one can think about the compatibility of environmental conditions with current and future communities as the flow of water between reservoirs. The thicker the pipeline between two adjacent reservoirs, the faster the water flows (the higher the compatibility). The thickness of a pipeline is a function of the water capacity of such reservoirs. The water capacity is constrained by the nature of building elements and the interactions between these elements. Paths (sequence of pipelines) with higher thickness represent lower environmental resistance. Water would tend to flow through the thickest pipelines, imposing less friction (Figures 1D and 1E). The probability of reaching a target reservoir from an initial reservoir after  $x$  time steps is given by the product of probabilities (proportional to the thickness of pipelines) of using all possible path combinations across  $x$  pipelines. In the case of a simultaneous developmental process, there is just one pipeline. We explore how the probability of reaching a target reservoir changes under sequential and simultaneous processes (Figure 1F).

The environmental resistance of developing to a new possible community (conditional probability of moving to state  $j$  given current state  $i$ ) is a function of the feasibility of such communities (Figure 1C). The higher the feasibility of the current community and the lower the feasibility of the future community, the higher the environmental resistance of developing to a new community (see STAR Methods for a formal definition of feasibility). Feasibility is a function of local taxon interactions (Figures 1A and 1B) and is formalized following five key assumptions: (1) all taxa have equal dispersal rates. This means that the order of appearance is unbiased, but this order can be stochastic. (2) Taxon interactions are a function of taxon consumption rates. Taxa with higher consumption rates have higher competitive effects on other taxa (Figure S11). (3) All taxa have the same carrying capacity. This means that the higher the per capita growth rate of a taxon, the higher its consumption rate and consequently the higher its competitive advantage over available resources. (4) For modeling purposes and for application to particular datasets, the maximum growth rate ( $r_{max}$ ) of a taxon is proportional to its average body size using metabolic scaling laws. (5) The average body size is not expected to change over time. This implies that competitive effects and the feasibility of a community are also time-invariant. We present the formal feasibility analysis in the STAR Methods, which combines ecological dynamics, metabolic scaling theory, and Markov processes (Figure 1).

Using our proposed framework, we theoretically investigate the probability of a sequenced developmental process (introducing one taxon at a time) as opposed to a simultaneous developmental process (introducing all taxa at the same time). We then evaluate the conditions under which the sequential developmental process can follow an orderly path from fast-growing to slow-growing taxa. Specifically, we address whether this orderly path corresponds to the theoretical path with the highest probability, referred to as the path of least resistance. We focus our analysis on the special case of directional assembly (once a taxon is successfully introduced, it cannot go extinct). Effectively, this analysis aims to explain the probability that a group



**Figure 1. An illustration of developmental processes in ecological systems**

(A) A three-taxon competition system (taxon pool: tx1, tx2, and tx3), where node size represents the maximum growth rates ( $r_{max}$ ) of each taxon. The thickness of arrows represents the strength of pairwise interactions.

(B) Estimation of pairwise interactions  $a_{ij}$  based on  $r_{max}$ . Following metabolic scaling theory,  $r_{max}$  scales with body mass in plants and metazoans with a power of  $-1/4$  and in bacteria and archaea with a power of  $3/4$ . For bacteria and archaea,  $r_{max}$  also scales with the copy number of ribosomal RNA operons ( $rrn$ ) with a power of  $5/4$  (STAR Methods; Figure S10).

(C) Feasibility domains for communities that consist of one taxon (e.g., community of tx1 in gray), two taxa (e.g., systems of tx1 and tx2 in yellow or tx1 and tx3 in blue), and three taxa (in brown). The feasibility domain of a single taxon is assumed to be 0.5 (when its effective resource availability  $\theta_1 > 0$ ). For a multi-taxon community, the feasibility domain is governed by the interaction strengths  $a_{ij}$  between taxa (STAR Methods).

(D) Sequential development can be probabilistically described by a Markov chain, where the feasibility-based transition probabilities  $P$  are shown as pie charts for communities starting from none  $\emptyset$  to all three taxa  $\{123\}$ . These transition probabilities are determined by the feasibility of all possible communities in the sequence (match with colors in C; STAR Methods). Each community is assumed to either maintain the current taxon composition or transition to a larger community with an additional taxon. The results remain robust considering disassembly processes (STAR Methods). The overall probability of developing the three-taxon system along the path  $\emptyset \rightarrow \{1\} \rightarrow \{12\} \rightarrow \{123\}$  can be calculated by multiplying the three transition probabilities along the path. The path with the highest probability is called the path of least resistance (represented by pink water pipes).

(E) Simultaneous development can also be probabilistically described by a Markov chain and is equivalent to a binomial process (STAR Methods). The transition probabilities are determined by the feasibility of  $\{123\}$  only (match with colors in C; STAR Methods).

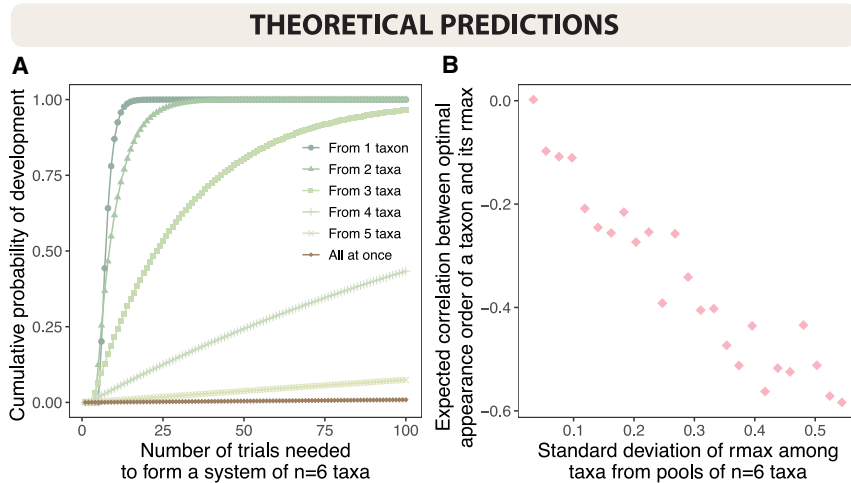
(F) When comparing the probability of developing from the empty set  $\emptyset$  (standardized baseline) to the largest community  $\{123\}$  with all taxa, both sequential and simultaneous processes are guaranteed to succeed ( $\gamma = 1$ ) after a certain number of trials ( $x$  axis). However, sequential development can achieve this with fewer trials.

of  $n$  species can be found together at a given point in time. While this analysis is a simplification of the developmental process undergone by natural systems (i.e., no extinctions), and under our assumptions, this analysis generates the same results as if additionally considering disassembly processes (Figures S1D–S1F; STAR Methods). That is, under assembly only, the question is which path is the system most likely to take. Under assembly and disassembly, the question would be which path is most frequently visited by the system if a system is observed with  $n$  taxa. We test our theoretical predictions using empirical

developmental data across a wide range of taxonomic levels, including infant gut and marine microbiota,<sup>14,15</sup> tree and bird systems,<sup>16,17</sup> and multiyear and seasonal mammalian herbivore systems.<sup>18,19</sup> Using these empirical data, we investigate the extent to which the observed developmental order follows the theoretical path of least resistance.

#### Theoretical predictions

We started our theoretical analysis by investigating whether a sequenced assembly (a special case of development) is more



**Figure 2. Probabilistic explanation for the sequential order of development going from fast-growing to slow-growing taxa**

(A) Comparison of cumulative probability of development (y axis) between sequential development (green) with varying numbers of initial taxa and a simultaneous development (brown) within a limited number of trials (or attempts; x axis). Notably, in sequential development, the process starting from one taxon reaches probability 1 more quickly compared to those starting with larger initial communities. Additionally, the probability of development through the simultaneous process is negligibly small, though it may reach probability 1 if the number of attempts is sufficiently large (STAR Methods).

(B) Relationships between the path of least resistance and  $r_{max}$  of sequential development starting from 1 taxon. Firstly, the negative expected correlations between the most likely developmental order of a taxon and its  $r_{max}$  ( $\gamma < 0$ ; Spearman's rank

correlation) indicate that, in order to maximize the probability of development, fast-growing taxa with larger  $r_{max}$  are expected to appear earlier. Also, as the standard deviation of  $r_{max}$  among taxa (x axis) increases, these negative correlations tend to be stronger (approaching  $\gamma = -1$ ). This finding suggests that when the diversity in the  $r_{max}$  among taxa within a system is higher, the appearance order starting from fast-growing taxa to slow-growing taxa becomes more distinguishable. Note that in both panels, we choose  $n = 6$  taxa only for illustrative purposes. The results remain consistent across different choices of  $n$  and when additionally considering disassembly (Figure S1).

See also Figures S11 and S12.

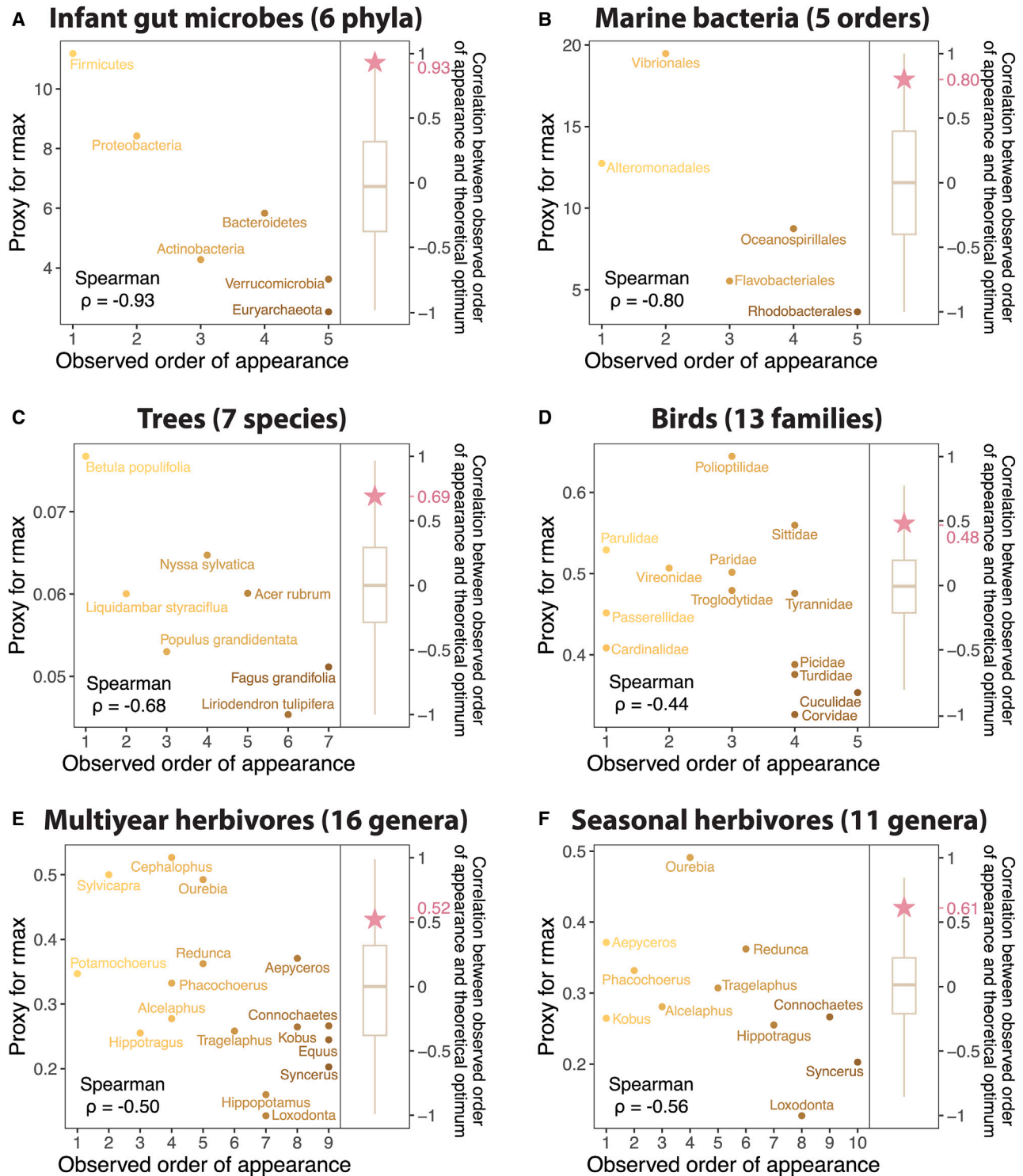
likely than having all taxa introduced at once. In other words, is there an inherent tendency toward sequenced assemblies even without invoking extensive details about individual characteristics (only local taxon interactions)? Is development (and ecological restoration) more likely to succeed when proceeding one species at a time rather than in batches? To test these ideas, we calculated the probabilities for sequential and simultaneous development using model-generated systems following our proposed feasibility analysis (STAR Methods).

Figure 2A shows the cumulative probabilities of reaching full development as a function of the number of trials, comparing sequenced versus simultaneous development in a hypothetical system of  $n = 6$  taxa, starting from different initial numbers of taxa ( $1 \leq m \leq n$ ). Note that  $m = n$  is equivalent to the simultaneous development (i.e., all taxa appear simultaneously). Results are qualitatively consistent using other dimensions  $n$  (Figures S1A and S1B). By construction, these systems can be assembled starting from any initial community (STAR Methods). We found that the smaller the initial number of taxa (after the empty set  $\emptyset$ ), the larger the cumulative probability of development. Only after an extremely large number of attempts does the cumulative probability of simultaneous development asymptotically converge to one. Together, these results suggest that, barring a vast number of introduction attempts, it is far less likely to observe a system developed from the simultaneous appearance of many taxa than one arising from a sequenced process.

Having established the inherent probability of a sequenced development of ecological systems, we theoretically investigated the conditions under which it follows a sequenced, orderly process as a function of biodiversity. To do so, we used the same systems of  $n = 6$  taxa generated before (as an illustrative example). Then, we divided these model-generated systems into groups, ensuring equal intervals of the standard deviations of  $r_{max}$  among taxa. For each system, we identified the path

with the highest probability, which is the path of least resistance. Next, we calculated Spearman's rank correlation between the taxon order in each estimated path and the generated taxon  $r_{max}$ . Figure 2B shows that the most likely appearance order of a taxon is, in general, strongly and negatively correlated with its  $r_{max}$ , indicating that the path of least resistance occurs when fast-growing taxa appear first. However, note that positive correlations between appearance and  $r_{max}$  can be observed, especially in unlikely paths. Furthermore, we found that higher variations in growth rates are associated with more strongly negative correlations between a taxon's most likely appearance order and its  $r_{max}$ . By extension, pools with a greater diversity of taxa accentuate the positive association between developmental success and a successional process going from fast- to slow-growing taxa.

Because our feasibility analysis is based on local taxon interactions, our results can be explained by the fact that similar competitive advantages (similar  $r_{max}$ ) promote symmetric interactions and in turn increase the feasibility of a community (Figure S11). Asymmetries get amplified between taxa with small competitive advantages (Figure S12). Similarly, because the probability of a path is given by the multiplication of conditional probabilities, this product is maximized by having taxa with similar  $r_{max}$  together in an ordered assembly rather than a disorderly one. Theoretically, the path of least resistance solves a Pareto front, maximizing the probability of staying and leaving. For example, this is achieved at  $P(\text{staying}) = P(\text{leaving}) = 0.5$ . This means that the path cannot select the maximum probability at each step, only the maximum product of probabilities, implying a combination of forming and breaking symmetries. By starting from fast-growing taxa, the probability of developing the system (the largest community) under a Markov process increases, but it is difficult to find a perfectly ordered process unless there is a high variability of  $r_{max}$  among taxa. While all these



**Figure 3. Observed developmental order in empirical systems and comparisons with theoretical and random orders**

- (A) Primary development involving 6 phyla of gut microbes in a healthy male infant.<sup>14</sup>  
 (B) Primary development involving 5 orders of marine bacteria in a laboratory study.<sup>15</sup>  
 (C) Secondary development involving 7 species of trees in the northeastern United States.<sup>16</sup>  
 (D) Secondary development involving 13 families of birds in the southeastern United States (Figure S2).<sup>17</sup>

(legend continued on next page)

previous theoretical results are generated assuming a directional assembly (i.e., no extinctions), and under our assumptions, these results do not change by following more complex scenarios involving disassembly dynamics (Figures S1D–S1F; STAR Methods). These results imply that biodiversity can turn into a heterogeneity of local taxon interactions, which can then amplify the emergence of a sequenced and orderly development.

### Empirical corroboration

To empirically evaluate our theoretical predictions, we compiled publicly available data from developmental processes involving taxonomic levels ranging from species to phyla, time-scales spanning days to decades, and organisms spanning ~15 orders of magnitude in size from bacteria to elephants (STAR Methods). The first two datasets represent primary development in microbes: the development of an infant gut microbiota comprising six main phyla across 2.5 years using fecal samples and 16S rRNA and metagenomic sequence data,<sup>14</sup> and the development of marine bacteria from a laboratory model system comprising five main orders over a span of 144 h.<sup>15</sup> The third and fourth datasets represent secondary development in trees and birds: the development of a community of trees comprising 7 species in Princeton, NJ, that was disturbed by a blight in 1920 and hurricanes in 1938 and 1944,<sup>16</sup> and the development of an assemblage of birds comprising 13 main families from abandoned agricultural fields.<sup>17</sup> The fifth and sixth datasets represent secondary and seasonal development, respectively, of large mammalian herbivores in Gorongosa National Park: the development of the entire herbivore system comprising 16 genera over 28 years after the Mozambican Civil War<sup>19</sup> and the development of a local herbivore system comprising 11 genera throughout the dry season in 2016.<sup>18</sup>  $r_{max}$  were approximated using metabolic scaling relationships based on average ribosomal RNA operons copy numbers for bacteria and archaea,<sup>52</sup> diameter at breast height for trees,<sup>53</sup> and body masses for metazoans.<sup>18,54</sup>

Figure 3 shows that the development of these empirical systems tends to proceed from fast-growing to slow-growing taxa. As predicted, we found strong negative correlations (Spearman's rank correlation  $\rho \in [-0.93, -0.44]$ ) between the observed order of taxon first appearance and the  $r_{max}$  estimated for each taxon. Next, for each empirical system, we determined the theoretical path of least resistance using  $r_{max}$  following our proposed feasibility analysis (STAR Methods). In line with our theoretical results (Figure 2), we found strong positive correlations (Spearman's rank correlation  $\rho \in [0.48, 0.93]$ ; pink stars) between the observed order of taxon first appearance and the theoretical optimum, which were always larger than those expected from random appearance orders (boxplots). These results are robust to various assumptions and criteria used in data processing (Figures S2 and S3; STAR Methods).

Additionally, Figure 4 shows that development from fast- to slow-growing taxa becomes stronger with increasing diversity (variation) in their growth rates, confirming our theoretical predictions. For each empirical system, we drew all possible subsets (combinations) consisting of 60% taxa. Then, for each subset, we calculated Spearman's rank correlation between the observed order of first appearance of a taxon and its  $r_{max}$ , as well as the standard deviation of  $r_{max}$  among all taxa within the subset. Next, we divided all subsets into three diversity groups based on the standard deviations of  $r_{max}$  and plotted the distribution of correlations between observed first appearance order and  $r_{max}$  within each group. In line with our theoretical predictions (Figure 2B), we found that average correlations become more negative with greater diversity in  $r_{max}$ . These results are robust to the choice of taxon pool size and the number of diversity groups (Figures S4–S9), suggesting that diversity plays a central role in shaping the development of ecological systems.<sup>55</sup>

### Conclusions

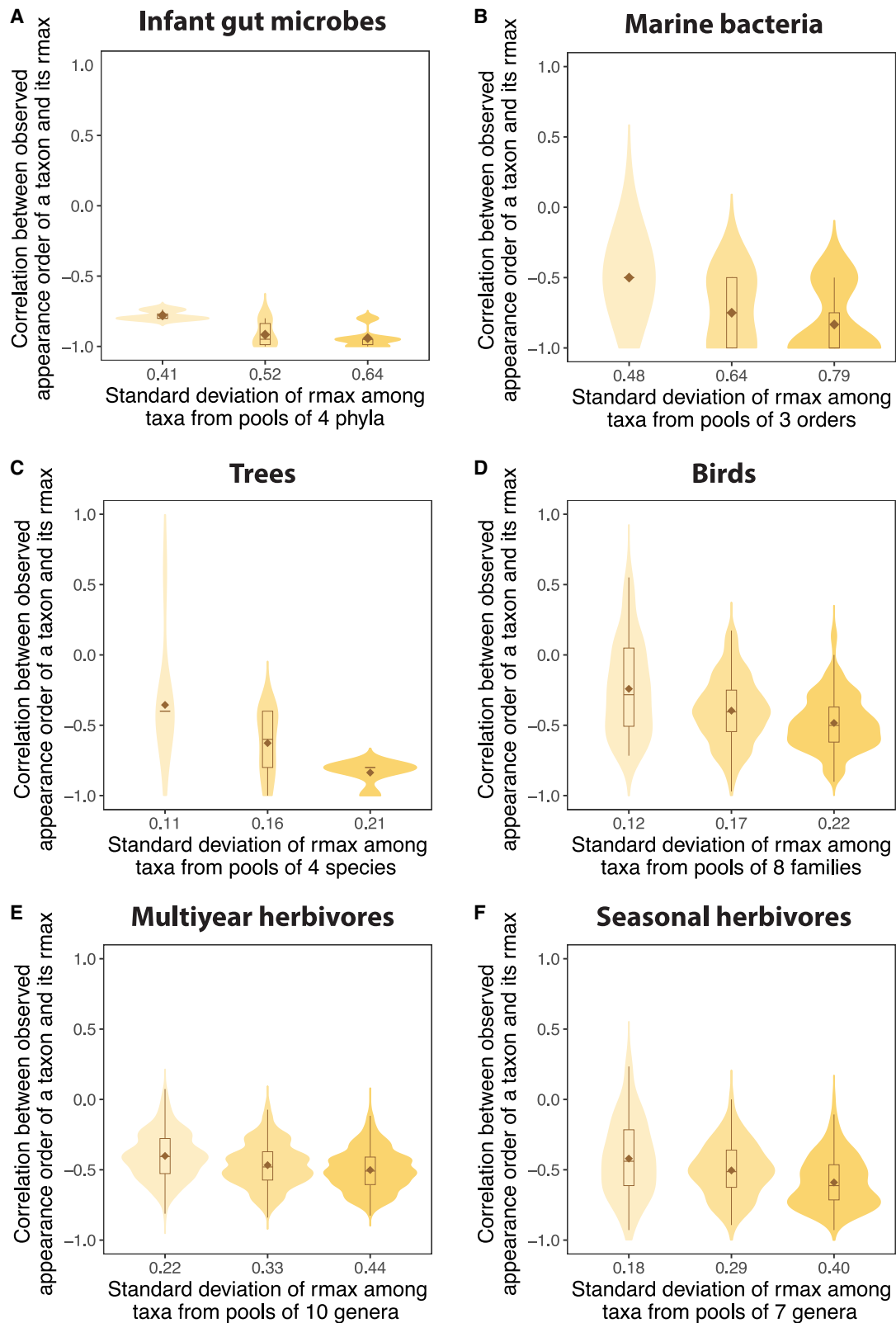
We have presented a general theoretical framework for development based on basic feasibility assumptions rooted in local taxon interactions to better understand the conditions underlying the tendency of taxa to appear from fast-growing to slow-growing taxa. More than a century of evidence shows that the development of ecological systems is characterized by ordered processes and that systems often converge on similar trajectories.<sup>2,6,12,56</sup> Can this phenomenon be explained only with reference to a large combination of individual traits and circumstances, or can it be understood as a self-organized collective outcome emerging from basic premises? We have shown that if one assumes a given developmental order as a local solution via taxon interactions to changing environmental challenges, then the solution that maximizes the number of surmountable environmental challenges is typically the one that proceeds from fast-growing to slow-growing taxa. In other words, development follows paths of least environmental resistance. Because biodiversity promotes an orderly sequential development and natural selection acts at the local level,<sup>50</sup> it can be speculated that eco-evolutionary dynamics via life-history traits and environmental compatibility may play a reinforcement role in preserving this basic developmental process at the system level.<sup>23</sup>

Supporting the statements above, we have shown theoretically that a sequential development is more likely than a simultaneous development, at least on the scales of time and iteration that plausibly resemble those of natural events. Indeed, it requires an enormous amount of tape repetitions (attempts) for a simultaneous assembly to become comparable to a sequential one. Another prediction derived from our theory is that the higher the heterogeneity of intrinsic growth rates within a pool of taxa,

(E) Secondary development (corresponding to the postwar reassembly over the period 1994–2022) involving 16 genera of herbivores in Mozambique (Figure S3).<sup>19</sup>

(F) Seasonal development (over the dry season in 2016) involving 11 genera of herbivores in Mozambique.<sup>18</sup>

In each of these empirical systems, the correlation between the observed order of appearance (x axis) and the proxy for  $r_{max}$  (y axis) is negative (Spearman's rank correlation  $\rho = -0.93, -0.80, -0.68, -0.44, -0.50,$  and  $-0.56$  in A–F, respectively), in line with our theoretical prediction (Figure 2B). On the right side of each panel, the correlation between the observed order of appearance and the theoretical order of path of least resistance (pink stars;  $\rho = 0.93, 0.80, 0.69, 0.48, 0.52,$  and  $0.61,$  respectively) is significantly higher than expected under randomly generated orders (boxplots; based on  $10^3$  random paths).



(legend on next page)



the more pronounced the ability of an orderly process can characterize the development of ecological systems—perhaps helping to explain why changes in biodiversity can drastically alter the development of ecological systems.<sup>55</sup> Importantly, we have focused on paths that can minimize environmental resistance and become more likely. While there is a tendency in these paths to go from fast-growing to slow-growing taxa, small differences in growth rates can introduce deviations from this development structure. For example, variations in life-history traits or life-history trajectories can help to explain deviations from the expectation.<sup>57</sup> It is worth mentioning that these results do not provide the specific step-by-step order of assembly (and disassembly), but simply they provide a minimal model rooted in local taxon interactions to explain the expected order of taxon appearance in natural systems without the need to invoke a high level of information about life-history traits. That being said, this minimal model supports the existence of least-action principles that can apply to the development of life on Earth.

Using empirical data from a wide range of developmental processes and taxonomic groups, we have found support for our theoretical results and predictions. The qualitative consistency of these results from very small ecological systems (bacteria in 1 L bottles under controlled laboratory conditions over 6 days) to some of the world's largest (ungulates and elephants in a 4,000 km<sup>2</sup> national park) suggests a potentially high level of generality. We emphasize, however, that our results do not rule out alternative hypotheses of developmental processes, nor are they incompatible with explanations that invoke more nuanced biological mechanisms to explain specific trajectories. Rather, our results illustrate that basic assumptions can be sufficient to explain coarse-grained properties of these dynamics and establish an expectation for their common tendencies. Future studies might use our work as a platform to explore general paths and processes for taxonomic and interaction turnover during development. Indeed, the interplay between stochasticity and determinism continues to be an active line of research.<sup>8,9,58–60</sup> On the one hand, historical contingency (environmental conditions) seems to be an underlying factor modulating the short-term dynamics of ecological systems.<sup>1,29</sup> On the other hand, convergent evolution seems like a plausible explanation for finding a finite set of long-term solutions to an environmental challenge.<sup>56,61</sup> Our work suggests that accounting for least-action principles (e.g., least environmental resistance) can be a step toward integrating different schools of thought.

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jie Deng ([jiedeng@mit.edu](mailto:jiedeng@mit.edu)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

- This paper analyzes existing, publicly available data. The references for datasets are listed in the [key resources table](#).
- All original code has been deposited at Github: [https://github.com/MITEcology/Deng\\_Assembly](https://github.com/MITEcology/Deng_Assembly). This link is also listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## ACKNOWLEDGMENTS

J.D. thanks the financial support from MathWorks. S.S. thanks the support by the National Science Foundation under grant no. DEB-2436069. R.S. thanks the support of the Generalitat de Catalunya under grant AGAUR 2021 SGR 00751. R.M.P. thanks the National Science Foundation BoCP DEB-2225088 and DEB-2303835, the Carr and Cameron Schrier Foundations, the Biodiversity Grand Challenge of the High Meadows Environmental Institute, and Gorongosa National Park's Department of Scientific Services. S.A.L. thanks NSF grant DMS-1951358 and ARO grant W911NF2410126. The authors thank the contributors and maintainers of databases rrnDB, BIEN, and AVONET for their valuable work. Furthermore, the authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing high-performance computing resources that have contributed to the research results reported within this paper.

## AUTHOR CONTRIBUTIONS

Conceptualization, all authors; methodology, J.D., R.M.P., and S.S.; software, J.D.; validation, all authors; formal analysis, J.D., R.M.P., and S.S.; investigation, all authors; resources, O.X.C. and R.M.P.; data curation, J.D.; writing – original draft, J.D., R.M.P., and S.S.; writing – review and editing, all authors; supervision, S.S.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
  - Details of empirical data
- [METHOD DETAILS](#)
  - Feasibility analysis
  - Lotka-Volterra model and feasibility
  - Estimating competitive interactions ( $a_{ij}$ ) using consumption rates ( $C$ )
  - Trade-offs induced by increasing growth rates ( $r$ )
  - Inferring consumption rates ( $C$ ) using maximum growth rates ( $r_{max}$ )
  - Quantification of the transition matrix ( $P$ ) in the Markov chain
  - Understanding simultaneous development as a binomial process
  - Identification of the path of least resistance
  - Theoretical predictions are not influenced by disassembly

## Figure 4. Sequential development from fast-growing to slow-growing taxa becomes more distinguishable as the diversity in $r_{max}$ among taxa increases

(A–F) For each of the empirical systems in [Figure 3](#), violin plots and embedded boxplots show the distribution of correlations between the observed appearance order of a taxon and its  $r_{max}$  (y axis; Spearman's rank correlation) as a function of the standard deviation of  $r_{max}$  among taxa (x axis). Brown diamonds inside each boxplot show the expected value (mean) of the correlations. As the standard deviation of  $r_{max}$  increases, the correlation between the observed order of appearance and  $r_{max}$  becomes stronger (approaching  $y = -1$ ), in line with our theoretical prediction in [Figure 2B](#). Here, we choose the pool size as 60% of the entire system size and set the number of diversity groups to three, but results are consistent across different choices of parameters (see [Figures S4–S9](#)). See also [Figures S2](#) and [S3](#).

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Proxy for maximum growth rate (r<sub>max</sub>)

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2024.08.050>.

Received: December 10, 2023

Revised: July 25, 2024

Accepted: August 28, 2024

Published: September 26, 2024

## REFERENCES

1. Gould, S.J. (1989). *Wonderful Life* (W. W. Norton & Company).
2. Odum, E.P. (1969). The strategy of ecosystem development. *Science* *164*, 262–270.
3. Cowles, H.C. (1899). The ecological relations of the vegetation on the sand dunes of Lake Michigan. Part I.—Geographical relations of the dune floras. *Botanical Gazette* *27*, 95–117.
4. Clements, F.E. (1916). *Plant Succession: An Analysis of the Development of Vegetation* (Carnegie Institution of Washington).
5. Levin, S.A. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* *1*, 431–436.
6. Walker, L.R., and Del Moral, R. (2003). *Primary Succession and Ecosystem Rehabilitation* (Cambridge University Press).
7. Prach, K., and Walker, L.R. (2011). Four opportunities for studies of ecological succession. *Trends Ecol. Evol.* *26*, 119–123.
8. Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* *81*, 10.
9. Dini-Andreote, F., Stegen, J.C., Van Elsas, J.D., and Salles, J.F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci. USA* *112*, E1326–E1332.
10. Chang, C.C., and Turner, B.L. (2019). Ecological succession in a changing world. *Journal of Ecology* *107*, 503–509.
11. Chase, J.M. (2003). Community assembly: when should history matter? *Oecologia* *136*, 489–498.
12. Connell, J.H., and Slatyer, R.O. (1977). Mechanisms of succession in natural communities and their role in community stability and organization. *Am. Nat.* *111*, 1119–1144.
13. Chapin, F.S., Walker, L.R., Fastie, C.L., and Sharman, L.C. (1994). Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. *Ecol. Monogr.* *64*, 149–175.
14. Koenig, J.E., Spor, A., Scafone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. USA* *108*, 4578–4585.
15. Datta, M.S., Sliwerska, E., Gore, J., Polz, M.F., and Cordero, O.X. (2016). Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat. Commun.* *7*, 11965.
16. Horn, H.S. (1975). Forest succession. *Sci. Am.* *232*, 90–98.
17. Johnston, D.W., and Odum, E.P. (1956). Breeding bird populations in relation to plant succession on the Piedmont of Georgia. *Ecology* *37*, 50–62.
18. Pansu, J., Hutchinson, M.C., Anderson, T.M., Te Beest, M., Begg, C.M., Begg, K.S., Bonin, A., Chama, L., Chamailé-Jammes, S., Coissac, E., et al. (2022). The generality of cryptic dietary niche differences in diverse large-herbivore assemblages. *Proc. Natl. Acad. Sci. USA* *119*, e2204400119.
19. Stalmans, M.E., Massad, T.J., Peel, M.J.S., Tarnita, C.E., and Pringle, R.M. (2019). War-induced collapse and asymmetric recovery of large-mammal populations in Gorongosa National Park, Mozambique. *PLoS One* *14*, e0212864.
20. Solé, R.V., and Bascompte, J. (2005). *Self-Organization in Complex Ecosystems* (MPB-42)58 (Princeton University Press).
21. Margalef, R. (1968). *Perspectives in Ecological Theory* (University of Chicago Press).
22. Jablonski, D., and Sepkoski, J.J., Jr. (1996). Paleobiology, community ecology, and scales of ecological pattern. *Ecology* *77*, 1367–1378.
23. Kalambokidis, M., and Travisano, M. (2024). The eco-evolutionary origins of life. *Evolution* *78*, 1–12.
24. Solé, R.V., Montoya, J.M., and Erwin, D.H. (2002). Recovery after mass extinction: evolutionary assembly in large-scale biosphere dynamics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *357*, 697–707.
25. Palmer, M.A., Ambrose, R.F., and Poff, N.L. (1997). Ecological theory and community restoration ecology. *Restor. Ecol.* *5*, 291–300.
26. Wainwright, C.E., Staples, T.L., Charles, L.S., Flanagan, T.C., Lai, H.R., Loy, X., Reynolds, V.A., and Mayfield, M.M. (2018). Links between community ecology theory and ecological restoration are on the rise. *J. Appl. Ecol.* *55*, 570–581.
27. Bunin, G. (2021). Directionality and community-level selection. *Oikos* *130*, 489–500.
28. Vellend, M. (2016). *The Theory of Ecological Communities* (Princeton University Press).
29. Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.* *46*, 1–23.
30. Boltzmann, L. (1964). *Lectures on Gas Theories* (Dover Publications).
31. Gibbs, J.W. (1960). *Elementary Principles in Statistical Mechanics* (Dover Publications).
32. Parisi, G. (1999). Complex systems: a physicist's viewpoint. *Phys. A* *263*, 557–564.
33. Doolittle, W.F., and Inkpen, S.A. (2018). Processes and patterns of interaction as units of selection: an introduction to ITSNTS thinking. *Proc. Natl. Acad. Sci. USA* *115*, 4006–4014.
34. Pringle, R.M. (2017). Upgrading protected areas to conserve wild biodiversity. *Nature* *546*, 91–99.
35. Perino, A., Pereira, H.M., Navarro, L.M., Fernández, N., Bullock, J.M., Ceaușu, S., Cortés-Avizanda, A., van Klink, R., Kuemmerle, T., Lomba, A., et al. (2019). Rewilding complex ecosystems. *Science* *364*, eaav5570.
36. Guyton, J.A., Pansu, J., Hutchinson, M.C., Kartzinel, T.R., Potter, A.B., Coverdale, T.C., Daskin, J.H., da Conceição, A.G., Peel, M.J.S., Stalmans, M.E., et al. (2020). Trophic rewilding revives biotic resistance to shrub invasion. *Nat. Ecol. Evol.* *4*, 712–724.
37. McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. USA* *110*, 3229–3236.
38. Lawton, J.H., and Brown, V.K. (1994). *Redundancy in Ecosystems* (Springer).
39. Drake, J.A. (1991). Community assembly mechanics and the structure of an experimental species ensemble. *Am. Nat.* *137*, 1–26.
40. Warren, P.H., and Spencer, M. (1996). Community and food-web responses to the manipulation of energy input and disturbance in small ponds. *Oikos* *75*, 407–418.
41. Gilpin, M.E., Carpenter, M.P., and Pomerantz, M.J. (1986). The assembly of a laboratory community: multispecies competition in *Drosophila*. In *Community Ecology*, J.M. Diamond, and T.J. Case, eds. (Harper and Row), pp. 23–40.
42. Friedman, J., Higgins, L.M., and Gore, J. (2017). Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* *1*, 109.

43. Serván, C.A., Capitán, J.A., Grilli, J., Morrison, K.E., and Allesina, S. (2018). Coexistence of many species in random ecosystems. *Nat. Ecol. Evol.* **2**, 1237–1242.
44. Zhao, N., Saavedra, S., and Liu, Y.Y. (2021). Impact of colonization history on the composition of ecological systems. *Phys. Rev. E* **103**, 052403.
45. Logofet, D.O. (1993). *Matrices and Graphs. Stability Problems in Mathematical Ecology* (CRC).
46. Hofbauer, J., and Schreiber, S.J. (2022). Permanence via invasion graphs: incorporating community assembly into modern coexistence theory. *J. Math. Biol.* **85**, 54. <https://doi.org/10.1007/s00285-022-01815-2>.
47. Serván, C.A., and Allesina, S. (2021). Tractable models of ecological assembly. *Ecol. Lett.* **24**, 1029–1037.
48. Deng, J., Taylor, W., Levin, S.A., and Saavedra, S. (2024). On the limits to invasion prediction using coexistence outcomes. *J. Theor. Biol.* **577**, 111674.
49. Odum, E.P., and Barrett, G.W. (2005). *Fundamentals of Ecology, Fifth Edition* (Brooks-Cole).
50. Levin, S.A. (2005). Self-organization and the emergence of complexity in ecological systems. *BioScience* **55**, 1075–1079.
51. Long, C., Deng, J., Nguyen, J., Liu, Y.Y., Alm, E.J., Solé, R., and Saavedra, S. (2024). Structured community transitions explain the switching capacity of microbial systems. *Proc. Natl. Acad. Sci. USA* **121**, e2312521121.
52. Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K., and Schmidt, T.M. (2015). rrn DB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **43**, D593–D598.
53. Marbà, N., Duarte, C.M., and Agustí, S. (2007). Allometric scaling of plant life history. *Proc. Natl. Acad. Sci. USA* **104**, 15777–15780.
54. Tobias, J.A., Sheard, C., Pigot, A.L., Devenish, A.J.M., Yang, J., Sayol, F., Neate-Clegg, M.H.C., Alioravainen, N., Weeks, T.L., Barber, R.A., et al. (2022). AVONET: morphological, ecological and geographical data for all birds. *Ecol. Lett.* **25**, 581–597.
55. Enquist, B.J., Abraham, A.J., Harfoot, M.B.J., Malhi, Y., and Doughty, C.E. (2020). The megabiota are disproportionately important for biosphere functioning. *Nat. Commun.* **11**, 699.
56. Chuang, J.S., Frenzt, Z., and Leibler, S. (2019). Homeorhesis and ecological succession quantified in synthetic microbial ecosystems. *Proc. Natl. Acad. Sci. USA* **116**, 14852–14861.
57. Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., and Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2403–2413.
58. Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J.D., Yang, Y., He, Z., Wu, L., Stahl, D.A., et al. (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl. Acad. Sci. USA* **111**, E836–E845.
59. Tripathi, B.M., Stegen, J.C., Kim, M., Dong, K., Adams, J.M., and Lee, Y.K. (2018). Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* **12**, 1072–1083.
60. Ning, D., Deng, Y., Tiedje, J.M., and Zhou, J. (2019). A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl. Acad. Sci. USA* **116**, 16892–16898.
61. Morris, S.C. (2005). *Life's Solution: Inevitable Humans in a Lonely Universe* (Cambridge University Press).
62. Maitner, B.S., Boyle, B., Casler, N., Condit, R., Donoghue, J., Durán, S.M., Guaderrama, D., Hinchliff, C.E., Jørgensen, P.M., Kraft, N.J.B., et al. (2018). The bien R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* **9**, 373–379.
63. Peet, R.K., Lee, M.T., Boyle, F., Wentworth, T.R., Schafale, M.P., and Weakley, A.S. (2012). Vegetation-plot database of the Carolina Vegetation Survey. *Biodivers. Ecol.* **4**, 243–253.
64. Aussenac, R., Bergeron, Y., Ghotsa Mekontchou, C., Gravel, D., Pilch, K., and Drobyshev, I. (2016). Data from: Intraspecific variability in growth response to environmental fluctuations modulates the stabilizing effect of species diversity on forest growth. <https://doi.org/10.5061/dryad.55gb7>.
65. Anderson-Teixeira, K., McGarvey, J., Muller-Landau, H., Park, J., Gonzalez-Akre, E., Herrmann, V., Bennett, A., So, C., Bourg, N., Thompson, J., et al. (2015). Data from: Size-related scaling of tree form and function in a mixed-age forest. <https://doi.org/10.5061/dryad.6nc8c>.
66. Herberstein, M.E., McLean, D.J., Lowe, E., Wolff, J.O., Khan, M.K., Smith, K., Allen, A.P., Bulbert, M., Buzatto, B.A., Eldridge, M.D.B., et al. (2022). Animal traits—a curated animal trait database for body mass, metabolic rate and brain size. *Sci. Data* **9**, 265.
67. Gorongosa National Park Stalmans, M., and Peel, M. (2020). Aerial wildlife count of the Gorongosa National Park, Mozambique, November 2020. <https://gorongosa.org/wp-content/uploads/2023/10/GorongosaAerialWildlifeCount2020.pdf>.
68. Stalmans, M., Peel, M., and Gonçalves, D. (2022). Aerial wildlife count of the Gorongosa National Park, Mozambique, October 2022. Gorongosa National Park. <https://gorongosa.org/wp-content/uploads/2023/10/GorongosaAerialWildlifeCount-October-2022.pdf>.
69. Walker, R.H., Hutchinson, M.C., Becker, J.A., Daskin, J.H., Gaynor, K.M., Palmer, M.S., Gonçalves, D.D., Stalmans, M.E., Denlinger, J., Bouley, P., et al. (2023). Trait-based sensitivity of large mammals to a catastrophic tropical cyclone. *Nature* **623**, 757–764.
70. Daskin, J.H., Becker, J.A., Kartzinel, T.R., Potter, A.B., Walker, R.H., Eriksson, F.A.A., Buoncore, C., Getraer, A., Long, R.A., and Pringle, R.M. (2023). Allometry of behavior and niche differentiation among congeneric African antelopes. *Ecol. Monogr.* **93**, e1549.
71. Forest Inventory and Analysis National Program (2013). <http://www.fia.fed.us/>.
72. Saavedra, S., Rohr, R.P., Bascompte, J., Godoy, O., Kraft, N.J.B., and Levine, J.M. (2017). A structural approach for understanding multispecies coexistence. *Ecol. Monogr.* **87**, 470–486.
73. Deng, J., Angulo, M.T., and Saavedra, S. (2021). Generalizing game-changing species across microbial communities. *ISME Commun.* **1**, 22.
74. Deng, J., Taylor, W., and Saavedra, S. (2022). Understanding the impact of third-party species on pairwise coexistence. *PLoS Comp. Biol.* **18**, e1010630.
75. Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., and West, G.B. (2004a). Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789.
76. DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M., and Brown, J.H. (2010). Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc. Natl. Acad. Sci. USA* **107**, 12941–12945.
77. Enquist, B.J., Tiffney, B.H., and Niklas, K.J. (2007). Metabolic scaling and the evolutionary dynamics of plant size, form, and diversity: toward a synthesis of ecology, evolution, and paleontology. *Int. J. Plant Sci.* **168**, 729–749.
78. Roller, B.R.K., Stoddard, S.F., and Schmidt, T.M. (2016). Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* **1**, 16160.
79. AIAdwani, M., and Saavedra, S. (2020). Ecological models: higher complexity in, higher feasibility out. *J. R. Soc. Interface* **17**, 20200607.
80. Arnoldi, J.-F., Barbier, M., Kelly, R., Barabás, G., and Jackson, A.L. (2022). Invasions of ecological communities: hints of impacts in the invader's growth rate. *Methods Ecol. Evol.* **13**, 167–182.
81. Medeiros, L.P., Boege, K., del-Val, E., Zaldivar-Riverón, A., and Saavedra, S. (2021). Observed ecological communities are formed by species combinations that are among the most likely to persist under changing environments. *Am. Nat.* **197**, E17–E29.
82. Alberch, P. (1989). The logic of monsters: evidence for internal constraint in development and evolution. *Geobios* **22**, 21–57.

83. Rohr, R.P., Saavedra, S., and Bascompte, J. (2014). Ecological networks. On the structural stability of mutualistic systems. *Science* 345, 1253497.
84. Hofbauer, J., and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics* (Cambridge University Press).
85. Vandermeer, J.H., and Goldberg, D.E. (2013). *Population Ecology: First Principles* (Princeton University Press).
86. Case, T.J. (2000). *An Illustrated Guide to Theoretical Ecology* (Oxford University Press).
87. Marshall, D.J., Cameron, H.E., and Loreau, M. (2023). Relationships between intrinsic population growth rate, carrying capacity and metabolism in microbial populations. *ISME J.* 17, 2140–2143.
88. Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. *Numer. Math.* 1, 269–271.
89. Reuther, A., Kepner, J., Byun, C., Samsi, S., Arcand, W., Bestor, D., Bergeron, B., Gadepally, V., Houle, M., Hubbell, M., et al. (2018). Interactive supercomputing on 40,000 cores for machine learning and data analysis. In *IEEE High Performance extreme Computing Conference (HPEC)*, pp. 1–6.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Computational analysis	This paper	<a href="https://github.com/MITEcology/Deng_Assembly">https://github.com/MITEcology/Deng_Assembly</a>
High-Performance Computing	MIT SuperCloud	<a href="https://supercloud.mit.edu/">https://supercloud.mit.edu/</a>
R 4.3.2	The R Foundation for Statistical Computing	<a href="https://www.r-project.org">https://www.r-project.org</a>
Other		
Empirical data of infant gut microbes	Koenig et al. <sup>14</sup>	N/A
Empirical data of marine bacteria	Datta et al. <sup>15</sup>	N/A
Empirical data of trees	Horn et al. <sup>16</sup>	N/A
Empirical data of birds	Johnston et al. <sup>17</sup>	N/A
Empirical data of multiyear herbivores	Pansu et al. <sup>18</sup>	<a href="https://doi.org/10.5061/dryad.br15dvcj">https://doi.org/10.5061/dryad.br15dvcj</a>
Empirical data of seasonal herbivores	Stalmans et al. <sup>19</sup>	N/A
rrnDB database	Stoddard et al. <sup>52</sup>	<a href="https://rrnDB.umms.med.umich.edu/">https://rrnDB.umms.med.umich.edu/</a>
BIEN database	Maitner et al. <sup>62</sup> ; Peet et al. <sup>63</sup> ; Aussenac et al. <sup>64</sup> ; Anderson-Teixeira et al. <sup>65</sup> ; Forest Inventory and Analysis National Program	<a href="https://bien.nceas.ucsb.edu/bien">https://bien.nceas.ucsb.edu/bien</a>
AVONET database	Tobias et al. <sup>54</sup>	N/A

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

## Details of empirical data

In this paper, we investigate six published empirical datasets of varying ecological systems, including infant gut microbes,<sup>14</sup> marine bacteria,<sup>15</sup> trees,<sup>16</sup> birds,<sup>17</sup> and herbivores.<sup>18,19</sup>

The first empirical system of infant gut microbes (presented in Figures 3A and 4A;<sup>14</sup>) was drawn from a case study that tracked the gut microbial composition of a healthy male infant over 2.5 years, beginning on day 3. From a time series of more than 60 fecal samples, this dataset represents primary succession (*de novo* system development) and consists of six main phyla of bacteria and archaea (observed order of appearance): Firmicutes (#1), Proteobacteria (#2), Actinobacteria (#3), Bacteroidetes (#4), Verrucomicrobia (#5), and Euryarchaeota (#6). This observed order of appearance is shown in Figure 6A and Table S6 of Koenig et al.<sup>14</sup>

The second empirical system of marine bacteria (presented in Figures 3B and 4B;<sup>15</sup>) was drawn from a laboratory microcosm study. The authors immersed chemically defined, nutrient-rich microparticles in coastal seawater over a span of 144 hours. Then they recorded a collection of bacterial strains isolated from different phases of colonization and generated 16S rRNA gene sequences. This dataset also represents primary succession and consists of five main orders of bacteria (observed order of appearance): Alteromonadales (#1), Vibrionales (#2), Flavobacteriales (#3), Oceanospirillales (#4), and Rhodobacterales (#5). This observed order of appearance is shown in Figure 3A of Datta et al.<sup>15</sup>

The third empirical system of trees (presented in Figures 3C and 4C;<sup>16</sup>) was from a 500-acre woods in Princeton, NJ, owned by the Institute for Advanced Study. The forest was disturbed by a blight in 1920 and hurricanes in 1938 and 1944. Given the slow pace of forest succession, which makes direct observation impractical, the author inferred the successional patterns, including the order of taxon appearance, using the ages of all trees on various plots of land with unknown histories. The author has also confirmed these inferred successional patterns with further observations. This dataset represents secondary succession (local reassembly of an established species pool following an environmental change) and consists of 7 species (order of appearance): *Betula populifolia* (#1), *Liquidambar styraciflua* (#2), *Populus grandidentata* (#3), *Nyssa sylvatica* (#4), *Acer rubrum* (#5), *Liriodendron tulipifera* (#6), and *Fagus grandifolia* (#7). This order of appearance is shown in Figure SUCCESSIONAL CHANGES IN PRODUCTIVITY on Page 95 of Horn.<sup>16</sup>

The fourth empirical system of birds (presented in Figures 3D and 4D;<sup>17</sup>) was from abandoned agricultural fields in the Piedmont region of Georgia, United States. Between 1947 and 1951, the authors counted and identified birds across ten areas at various stages of natural vegetative succession, ranging from one-year abandoned fields to young mature forests aged over 150 years. This dataset represents secondary succession and consists of 14 main families (observed order of appearance): Parulidae (#1), Passerellidae (#1), Cardinalidae (#1), Vireonidae (#2), Polioptilidae (#3), Paridae (#3), Troglodytidae (#3), Trochilidae (#4), Sittidae (#4), Tyrannidae (#4), Picidae (#4), Turdidae (#4), Corvidae (#4), and Cuculidae (#5). This observed order of appearance is shown in Table 1 of Johnston and

Odum.<sup>17</sup> The ruby-throated hummingbird (*Archilochus colubris*), the only representative of Trochilidae in the dataset, is a known specialist that feeds largely on nectar from certain types of flower and sustains one of the highest mass-specific metabolic rates of any vertebrate.<sup>66</sup> Because the unique foraging behavior and outlying physiology of this species relative to others in the dataset has the potential to bias our inferences (e.g., by skewing detection probability in the original data depending on the timing and location of surveys in relation to flowering phenology), we excluded it from the main text analysis as an outlier. However, we obtained qualitatively similar results when including the hummingbird (see Figure S2).

The fifth empirical system of herbivores (presented in Figures 3E and 4E;<sup>18</sup>) was from Gorongosa National Park, in central Mozambique, southeast Africa. The data reflect a 28-year time series of aerial count data on ungulates and elephants from 16 genera (*Tragelaphus* being the only genus with >1 species) from 1994 to 2022.<sup>19</sup> Surveys were conducted both before (6 surveys of 9 genera from 1968–1972) and after (14 surveys of all 16 genera from 1994–2022) the Mozambican Civil War (1977–1992), which caused > 90% declines in all large-herbivore taxa.<sup>19</sup> Densities were determined by dividing the number of animals counted by the maximum area surveyed in each year (1968–2018 data are from Table S3 of Stalmans et al.<sup>19</sup>; 2020 data are from Stalmans and Peel,<sup>67</sup> and 2022 data are from Stalmans et al.<sup>68</sup>). The 14 postwar surveys analyzed here were conducted at varying intervals (1–3 years) using different methodologies (3 fixed-wing, 11 by helicopter) at different scales (106–2195 km<sup>2</sup>). The earlier surveys included the three fixed-wing counts (1994, 1997, 2004) and were conducted at smaller scales (average 472 km<sup>2</sup> from 1994–2007 vs. 1705 km<sup>2</sup> from 2010–2022). To account for sampling artifacts arising from these discrepancies, we calculated moving averages of densities in windows of 3 successive surveys (e.g., the value calculated for 2020 is the average of the 2018, 2020, and 2022 counts). This yielded estimates for 12 postwar intervals. We note that smaller species are harder to detect in aerial surveys, especially in fast-flying fixed-wing aircraft, but this bias is conservative with respect to our prediction (i.e., smaller species are less likely to be detected, particularly in early years). Because low observed densities do not necessarily indicate a viable population, we set three density thresholds (25%, 50%, and 75%), each relative to the maximum density recorded for the genus across all pre- and postwar surveys; a genus is classified as appeared when its density exceeds the threshold for the first time. These thresholds signify different assumptions about when a taxon can be considered to have successfully (re)established inside the park, as opposed to being vagrant individuals or nascent populations vulnerable to stochastic extinction. Taxa that are increasing in density but have not reached a given threshold were assigned the last order of appearance. For example, in Figure S3A, wildebeest (*Connochaetes*), buffalo (*Syncerus*), and zebra (*Equus*) were all present in 2022 but had not yet met the 25% threshold. The implicit assumption that these taxa will arrive is justified by their rapidly increasing densities: in fact, wildebeest exceeded 25% of maximum prewar density in 2022, but this is not captured in our moving-window approach, while buffalo (which are threefold larger with a correspondingly lower rmax) reached 11% and have been increasing by an average of 12% yr<sup>-1</sup> since 2016. The only uncertainty is zebra (>2% of prewar maximum in 2022 and still at risk of failure), but excluding zebra does not alter the result. Notably, the reassembly of large herbivores in Gorongosa is an overwhelmingly natural recovery process (the only genera appreciably supplemented by translocations are *Connochaetes*, *Syncerus*, and *Equus*, which have not yet arrived under any of our thresholds) and has been relatively robust to external perturbations.<sup>68,69</sup> In the main text (Figures 3E and 4E), we report results using a conservative 25% threshold. Figure S3 shows the results under all three thresholds.

The sixth empirical system of herbivores (presented in Figures 3F and 4F;<sup>18</sup>) was also from Gorongosa National Park. We analyzed encounter frequencies with individuals of different large mammalian herbivores (ungulates and elephants, 17–3825 kg) while driving road transects in the south of the park during the progression from the early to late dry season (June–August) of 2016. These data reflect secondary succession in response to seasonal change within a ~250 km<sup>2</sup> focal area and consist of 11 genera (observed order of appearance): *Aepyceros* (#1), *Kobus* (#1), *Phacochoerus* (#2), *Alcelaphus* (#3), *Ourebia* (#4), *Tragelaphus* (#5), *Redunca* (#6), *Hippotragus* (#7), *Loxodonta* (#8), *Connochaetes* (#9), and *Syncerus* (#10). Again, only one genus comprises >1 species (*Tragelaphus* with 3 closely related, ecologically similar species<sup>70</sup>). The observed order of appearance is from the metadata deposited in Dryad Digital Repository accompanying Pansu et al.<sup>18</sup> The seasonal results are consistent with the multi-decadal analysis (Figures 3E, 4E, and S3) and provide support for our theory over extraordinary organismal and spatiotemporal scales.

For the microbial systems, we extracted rrn copy numbers from the rrnDB database,<sup>52</sup> from which we estimated the rmax of infant gut microbes and marine bacteria. In these two systems, the rmax of each taxonomic rank, either a phylum or an order, was calculated as the average rmax estimated by the average rrn copy number across all documented species in the rrnDB database within the respective taxonomic level. For the forest system, we used the diameter at breast height data from the BIEN database<sup>62–65,71</sup> to estimate the rmax of each species. For the animal systems, we used body mass data from the AVONET database for bird species<sup>54</sup> and Pansu et al.<sup>18</sup> for herbivore species to estimate the rmax (Figure 1B). The rmax for higher taxonomic ranks, either a family or a genus, was calculated based on the average body mass across all species in the respective group. We then assessed the correlation between rmax and observed appearance order using Spearman's rank correlation coefficient ( $\rho$ ). Similarly, we used Spearman's  $\rho$  to evaluate the correlation between the observed path and the theoretically predicted path of least resistance and compared this with the distribution of  $\rho$  values obtained from 10<sup>3</sup> randomly generated paths (see results in Figures 3, S2A, and S3A–S3E).

To test our theoretical predictions about the influence of diversity in rmax on developmental order, we drew subsets of taxa from the entire empirical system to generate smaller assemblages with varying degrees of diversity in rmax. We then compared Spearman's rank correlations between rmax and observed order of appearance across these subsets, segmenting diversity (standard deviations) of rmax into different groups (see results in Figures 4, S2B–S2D, and S3B–S3F). For illustration, we delineated these subsets as 60% of the entire system size and the number of diversity groups as three. However, our results are robust to variation in these parameters (see Figures S4–S9).

## METHOD DETAILS

## Feasibility analysis

We illustrate the proposed feasibility premises of development using a hypothetical, competition, three-taxon system  $S$  as an example (Figure 1). We assume that the partition or modification of resources is characterized by the competitive effect  $a_{ij} = a_0(C_j/C_i)$  of taxon  $j$  on taxon  $i$  (Figures 1A–1B), where  $a_0$  is a normalization constant and  $C_i$  is the average consumption rate of taxon  $i$  (Equation 15). To study the maximum effect of growth rates on competition, we assume that all taxa have the same environmental carrying capacity and dispersal ability, and consequently  $a_{ij} \propto a'_0(r_{\max_j}/r_{\max_i})$ , where  $r_{\max_i}$  is the maximum growth rate of taxon  $i$  (Equation 35). This assumption implies that taxa are not constrained to enter at a particular order based on their specific traits (e.g., dispersal ability) other than their maximum growth rates and the current subset of present taxa.<sup>72–74</sup> Next, using metabolic scaling theory,<sup>75</sup> we can approximate maximum growth rates as a function of average body mass ( $M_i$ ) as  $r_{\max_i} \propto M_i^\beta$ , with  $\beta = -1/4$  for plants and metazoans, and  $\beta = 3/4$  for bacteria and archaea.<sup>76,77</sup> For bacteria and archaea, maximum growth rates can also be approximated as a scaling function of ribosomal RNA operons (rrn) copy numbers as  $r_{\max_i} \propto \text{rrn}_i^\beta$  with  $\beta = 5/4$ <sup>78</sup> (Figure 1B). The choice of scaling exponent does not affect the qualitative results of our framework (see below the section “quantification and statistical analysis” in STAR Methods). It is worth noting that our proposed framework can be considered as a basic process upon which one can add more complex contexts. For example, our framework illustrates that carrying capacities can act as an upper limit to consumption rates and can be used instead of growth rates upon the selection of specific trade-offs (Equation 28; Figure S10).

We also assume that any possible subset of taxa (community  $C$ ) from the hypothetical system ( $C \subseteq S$ ) is compatible with a specific *feasibility domain* of environmental space  $D_F(C)$ , and its feasibility is proportional to the size of such a domain and is a function of taxon interactions (Figure 1C). Formally, we determine  $D_F(C)$  following the Lotka-Volterra model (Equation 1):  $dN_i/dt = N_i r_i / \theta_i \left( \theta_i - \sum_{j \in C} a_{ij} N_j \right)$ , where  $N_i$  is the density of taxon  $i$ ,  $r_i$  is the intrinsic growth rates of taxon  $i$ ,  $\theta_i$  is the effective resource availability of taxon  $i$ , and  $a_{ij}$  is the competitive effect described in the preceding paragraph. Then, the feasibility domain corresponds

to the collection of compatible environmental conditions  $\left( D_F(C) = \left\{ \sum_{j \in C} N_j^* \mathbf{a}_j : \mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_{|C|}], N_j^* > 0 \right\} \in \mathbb{R}^{|C|} \right)$  leading to positive equilibrium densities ( $\mathbf{N}^* = \mathbf{A}^{-1} \boldsymbol{\theta} > \mathbf{0}$ ) constrained by the column vectors ( $\mathbf{a}_j$ ) of the competition matrix  $\mathbf{A}$ . Because we are interested in a developmental process, we assume that the system  $S$  can exhibit different community compositions, and therefore, a different resource partition or modification across time.<sup>72–74</sup> For instance, in a hypothetical system under a directional assembly, we may observe that taxon 1 appears first, then taxon 2, and lastly taxon 3. This sequence implies that changes in the feasibility domain are given by  $D_F(\{1\}) \rightarrow D_F(\{1,2\}) \rightarrow D_F(\{1,2,3\})$ . This same logic would apply to any other possible sequence of events, representing possible solutions to time-varying environmental conditions.

Lastly, we assume that a system  $S$  develops through sequenced community changes (the developmental path) that have the highest probability (Figure 1D). We search for this *path of least resistance* following a Markov chain process (the solution with the highest probability). Specifically, we translate each feasibility domain  $D_F(C)$  into a probability  $P(C) = \text{vol}(D_F(C) \cap \Theta^{(C)}) / \text{vol}(\Theta^{(C)})$ , where  $\text{vol}(\Theta^{(C)})$  corresponds to the volume of the parameter space of the environmentally-driven parameters  $\boldsymbol{\theta}$  (a proxy for environmental space). For example, assuming that taxon 1 appears first, the feasibility domain  $D_F(\{1\})$  corresponds to half of  $\text{vol}(\Theta)$  and  $P(\{1\}) = 0.5$ . We assume  $P(\{1\}) = P(\{2\}) = P(\{3\})$ , which implies consistent dispersal probabilities across the three taxa. Because this is a competition system,  $D_F(\{1\}) > D_F(\{1,2\}) > D_F(\{1,2,3\})$ ,<sup>79</sup> which implies  $P(\{1\}) > P(\{1,2\}) > P(\{1,2,3\})$ . However, unless interactions are the same,  $P(\{1,2\}) \neq P(\{1,3\}) \neq P(\{2,3\})$ . Under a Markov chain process, a system can either remain in its current state or change (refer to sections below). In our example, the systems will transition from empty  $\emptyset$  to a one-taxon community with probability  $1 - P(\emptyset \rightarrow \emptyset) = 0.5$ . But, whether it transitions to  $C = \{1\}$  instead of  $C = \{2\}$  or  $C = \{3\}$  depends on the ratio  $P(\{1\}) : P(\{2\}) : P(\{3\})$ . Assuming taxon 1 appears, the system will remain with a probability of  $P(\{1\} \rightarrow \{1\}) = 0.5$ . Alternatively, it will transition to a new state with probability  $P(\{1\} \rightarrow \{1,2\}) + P(\{1\} \rightarrow \{1,3\}) = 1 - P(\{1\} \rightarrow \{1\}) = 0.5$ . But, whether it transitions to  $C = \{1,2\}$  instead of  $C = \{1,3\}$  depends on the ratio  $P(\{1,2\}) : P(\{1,3\})$ . This enables us to quantify the probability of each path by multiplying the transition probabilities along the chain (e.g.,  $P(\emptyset \rightarrow \{1\} \rightarrow \{1,2\} \rightarrow \{1,2,3\}) = P(\emptyset \rightarrow \{1\}) \times P(\{1\} \rightarrow \{1,2\}) \times P(\{1,2\} \rightarrow \{1,2,3\})$ ). Notice that the sequenced probability  $P(\emptyset \rightarrow \{1\} \rightarrow \{1,2\} \rightarrow \{1,2,3\})$  is not necessarily the same as the one single event or simultaneous development  $P(\emptyset \rightarrow \{1,2,3\})$  (Figure 1E). Then, the probabilities among all possible changes can be written as a transition matrix ( $\mathbf{P}$ ). The path of least resistance is mathematically the shortest path of the directed and weighted graph constructed using the adjacency matrix  $-\log \mathbf{P}$  (Equation 44).

To perform our theoretical analysis, we calculated the cumulative probabilities for sequential and simultaneous development in different model-generated systems (Figure 1F). For each system, we assign taxon body masses drawn from a log-normal distribution centered at zero with varying standard deviations. Then, we estimated taxon  $r_{\max}$  using the generated body masses following a scaling relationship with a power of 3/4 (i.e.,  $r_{\max} = (\text{body mass})^{3/4}$ ). Results are qualitatively consistent across different power relationships (Figure S1C). Next, we estimated the elements of each interaction matrix  $\mathbf{A}$  as  $a_{ij} \propto a'_0(r_{\max_j}/r_{\max_i})$  (Equation 35), and calculated the transition matrix  $\mathbf{P}$  for the developmental process (Equation 43). The cumulative probability of development is calculated as the conditional probabilities of transitioning to a particular set of  $n$  taxa (community  $S$ ) in  $x$  trials (attempts) starting from the

empty set  $\emptyset$ . For each initial taxa, the respective conditional probability is the value at the intersection of row  $\emptyset$  and column  $S$  in the  $x$ -step transition matrix  $\mathbf{P}^x$ . Specially, under the simultaneous development, the cumulative probability can be understood as the frequency of observing at least one success in  $x$  independent attempts when having all taxa together  $(1 - (1 - P(\{S\}))^x)$ , where  $P(\{S\})$  is the one-attempt probability (Equation 3). Note that the results are consistent when disassembly is additionally considered (Figures S1D–S1F; STAR Methods).

### Lotka-Volterra model and feasibility

Our theoretical framework is based on the classic Lotka-Volterra model:

$$\frac{dN_i}{dt} = N_i \frac{r_i}{\theta_i} \left( \theta_i - \sum_{j \in S} a_{ij} N_j \right), \quad (\text{Equation 1})$$

where the vector  $\mathbf{N} = (N_1, \dots, N_{|S|})^T$  represents the densities of all taxa  $i \in \{1, \dots, |S|\} \equiv S$  within the system  $S$ . Here,  $|S|$  denotes the number of taxa in the system (i.e., the cardinality of a set). The vector  $\mathbf{r} = (r_1, \dots, r_{|S|})^T \in \mathbb{R}^{|S|}$  represents the intrinsic growth rate of all taxa  $i \in S$ . The vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{|S|})^T \in \mathbb{R}^{|S|}$  represents the effective resource availability (also known as environmental carrying capacity) of all taxa  $i \in S$ . The matrix  $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{|S| \times |S|}$  represents the pairwise interaction matrix. Each element  $a_{ij}$  is unitless, corresponding to the per-capita effect of taxon  $j$  on taxon  $i$  relative to taxon  $i$ 's self-regulation.  $a_{ij} > 0, i \neq j$  represent interspecific competitive interactions. Intraspecific competitive interactions  $a_{ii} = 1$  with the unitless normalization. Note that the classic Lotka-Volterra model Equation 1 is a phenomenological model. The effective resource availability  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{|S|})^T \in \mathbb{R}^{|S|}$  consists of the joint effect of the internal (e.g., intrinsic growth rate) and density-dependence factors acting on the growth rate of each taxon.<sup>74,80,81</sup>

In this paper, we assume that the pairwise interactions serve as a coarse-grained measure of the average interactions under environmental uncertainty, which is in line with a structuralist (also known as internalist) perspective.<sup>72,82,83</sup> Specifically, the structuralist perspective assumes that interactions among taxa (i.e., biological structure) are invariant and determine (or constrain) the environmental conditions compatible with the feasibility of ecological systems. Therefore, while it is expected to observe changes in pairwise interactions, the interaction matrix  $\mathbf{A}$  can be interpreted as an expectation of the invariant biological structure of an ecological system. This conceptual simplification also allows for a rough estimation of interactions using metabolic scaling laws (Figure 1B; STAR Methods). Besides, it is worth noting that our analysis focuses exclusively on competition systems to meet the requirements of the estimation (Figure 1A; STAR Methods).

Additionally, each effective resource availability  $\boldsymbol{\theta}$  captures the phenomenological influence of an environmental challenge on the growth potential of every individual taxon in the system  $S$ . This implies that distinct environmental challenges will lead to varying  $\boldsymbol{\theta}$  of the taxon. The combinations of  $\boldsymbol{\theta}$  compatible with the feasibility of taxa in the system  $S$  (known as the feasibility domain  $D_F$ ) can provide an approximation of the possibility of the system  $S$  in response to environmental challenges.<sup>72,74</sup> Specifically, the feasibility domain  $D_F(S)$  is bounded by the column vectors  $\mathbf{a}$  of interaction matrix  $\mathbf{A}$ :

$$D_F(S) = \left\{ \sum_{j \in S} N_j^* \mathbf{a}_j : \mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_{|S|}], N_j^* \in \mathbb{R}_{>0} \right\} \in \mathbb{R}^{|S|}, \quad (\text{Equation 2})$$

where  $\mathbf{a}_j$  is the  $j^{\text{th}}$  column vector of the interaction matrix  $\mathbf{A}$ . In other words, if  $\boldsymbol{\theta} \in D_F(S)$ , it guarantees that all taxa within the system  $S$  are feasible (i.e.,  $\mathbf{N}^* = \mathbf{A}^{-1} \boldsymbol{\theta} > \mathbf{0}$ ), a necessary condition for coexistence under equilibrium dynamics.<sup>84</sup> For any community  $C$  within the entire system  $S$  (i.e.,  $C \subseteq S$ ), the feasibility domain  $D_F(C)$  is bounded by the column vectors of the  $|C| \times |C|$  sub-matrix within the interaction matrix  $\mathbf{A}$  that corresponds to the community  $C$ .

Without any prior information about the environmental challenges, we assume all directions of  $\boldsymbol{\theta}$  are possible and equally likely, making the parameter space  $\Theta^{|S|}$  of  $\boldsymbol{\theta}$  a closed unit sphere in  $|S|$ -dimensional space. Then, we can normalize combinations of compatible  $\boldsymbol{\theta}$  by the parameter space  $\Theta^{|S|}$  of  $\boldsymbol{\theta}$  to approximate the feasibility (a probability between 0 and 1; Figure 1C):

$$P(S) = \frac{\text{vol}(D_F(S) \cap \Theta^{|S|})}{\text{vol}(\Theta^{|S|})}. \quad (\text{Equation 3})$$

It is worth mentioning that the assumption about an unbiased environment does not affect our qualitative results. This is true because even if we bias the environment (parameter space), this biased space needs to be inside the feasibility domain of all species together (otherwise there is no assembly). This implies that on average the bias will affect all the conditional probabilities in a proportional fashion leading to nearly the same path of least resistance.<sup>74</sup>

Furthermore, we would like to add a brief note that feasibility can be used to compute the transition probability, which represents the likelihood of transitioning from one community to another in the ecological developmental process (refer to the section “[quantification of the transition matrix \( \$\mathbf{P}\$ \) in the Markov chain](#)”; see also Figures 1D and 1E). Moreover, intuitively, this transition probability has the potential to indicate the most likely path of community changes (refer to the section “[identification of the path of least resistance](#)”).



**Estimating competitive interactions ( $a_{ij}$ ) using consumption rates (C)**

The estimation of competitive interactions  $a_{ij}$  is derived from the Consumer-Resource equations, wherein the consumers compete for resources, thus forming a competition system. The mathematical transformation from the Consumer-Resource equations to the Lotka-Volterra (LV) competition equations refers to Vandermeer and Goldberg.<sup>85</sup>

In particular, suppose that there are two self-replicating resources, which leads to the resource equations:

$$\begin{cases} \frac{dR_1}{dt} = r_{rsc_1} R_1 \left[ \left( \frac{k_1 - R_1}{k_1} \right) - c_{11} N_1 - c_{12} N_2 \right], \\ \frac{dR_2}{dt} = r_{rsc_2} R_2 \left[ \left( \frac{k_2 - R_2}{k_2} \right) - c_{21} N_1 - c_{22} N_2 \right], \end{cases} \quad \text{(Equation 4)}$$

where  $R_i$  is the density (or biomass) of the resource  $i$ ,  $r_{rsc_i}$  is the intrinsic rate of increase of resource  $i$ ,  $k_i$  is the carrying capacity of the resource  $i$ ,  $c_{ij}$  is the rate at which the resource  $i$  is consumed by the  $j^{\text{th}}$  consumer, and  $N_j$  is the population density of the  $j^{\text{th}}$  consumer.

Then, considering two consumers, we can write the consumer equations according to the classic LV predator-prey equations<sup>86</sup>:

$$\begin{cases} \frac{dN_1}{dt} = N_1 (b_{11} R_1 + b_{12} R_2 - m_1), \\ \frac{dN_2}{dt} = N_2 (b_{21} R_1 + b_{22} R_2 - m_2), \end{cases} \quad \text{(Equation 5)}$$

where  $b_{ij}$  is the relative conversion of a unit of resource  $j$  by consumer  $i$ , and  $m_i$  is the death rate of the  $i^{\text{th}}$  consumer. Note that  $b_{ij}$  is different from mass/energy transformation efficiency.

Assuming that the dynamics of the resources occur significantly faster than those of the consumers, we can equate the derivatives of Equation 4 to zero and solve for the equilibrium values  $R_1^*$  and  $R_2^*$  of the two resources:

$$\begin{aligned} R_1^* &= k_1 (1 - c_{11} N_1 - c_{12} N_2), \\ R_2^* &= k_2 (1 - c_{21} N_1 - c_{22} N_2). \end{aligned} \quad \text{(Equation 6)}$$

Then, we substitute  $R_1$  and  $R_2$  in Equation 5 by  $R_1^*$  and  $R_2^*$ , respectively. That is:

$$\begin{cases} \frac{dN_1}{dt} = N_1 [b_{11} k_1 (1 - c_{11} N_1 - c_{12} N_2) + b_{12} k_2 (1 - c_{21} N_1 - c_{22} N_2) - m_1], \\ \frac{dN_2}{dt} = N_2 [b_{21} k_1 (1 - c_{11} N_1 - c_{12} N_2) + b_{22} k_2 (1 - c_{21} N_1 - c_{22} N_2) - m_2], \end{cases} \quad \text{(Equation 7)}$$

which can also be written as

$$\begin{cases} \frac{dN_1}{dt} = N_1 [b_{11} k_1 + b_{12} k_2 - m_1 - (b_{11} k_1 c_{11} + b_{12} k_2 c_{21}) N_1 - (b_{11} k_1 c_{12} + b_{12} k_2 c_{22}) N_2], \\ \frac{dN_2}{dt} = N_2 [b_{21} k_1 + b_{22} k_2 - m_2 - (b_{21} k_1 c_{11} + b_{22} k_2 c_{21}) N_1 - (b_{21} k_1 c_{12} + b_{22} k_2 c_{22}) N_2]. \end{cases} \quad \text{(Equation 8)}$$

Multiplying and dividing the two equations in Equation 8 by

$$\begin{aligned} b_{11} k_1 + b_{12} k_2 - m_1, \\ b_{21} k_1 + b_{22} k_2 - m_2, \end{aligned} \quad \text{(Equation 9)}$$

respectively, we obtain

$$\begin{cases} \frac{dN_1}{dt} = (b_{11} k_1 + b_{12} k_2 - m_1) N_1 \left[ \frac{b_{11} k_1 + b_{12} k_2 - m_1 - (b_{11} k_1 c_{11} + b_{12} k_2 c_{21}) N_1 - (b_{11} k_1 c_{12} + b_{12} k_2 c_{22}) N_2}{b_{11} k_1 + b_{12} k_2 - m_1} \right], \\ \frac{dN_2}{dt} = (b_{21} k_1 + b_{22} k_2 - m_2) N_2 \left[ \frac{b_{21} k_1 + b_{22} k_2 - m_2 - (b_{21} k_1 c_{11} + b_{22} k_2 c_{21}) N_1 - (b_{21} k_1 c_{12} + b_{22} k_2 c_{22}) N_2}{b_{21} k_1 + b_{22} k_2 - m_2} \right]. \end{cases} \quad \text{(Equation 10)}$$

Again, dividing both the numerator and denominator of the fractions on the right sides of the two equations in Equation 10 by

$$\begin{aligned} b_{11} k_1 c_{11} + b_{12} k_2 c_{21}, \\ b_{21} k_1 c_{12} + b_{22} k_2 c_{22}, \end{aligned} \quad \text{(Equation 11)}$$

respectively, gives us

$$\begin{cases} \frac{dN_1}{dt} = (b_{11}k_1 + b_{12}k_2 - m_1)N_1 \left[ \frac{b_{11}k_1 + b_{12}k_2 - m_1 - N_1 - \frac{b_{11}k_1c_{12} + b_{12}k_2c_{22}}{b_{11}k_1c_{11} + b_{12}k_2c_{21}}N_2}{b_{11}k_1 + b_{12}k_2 - m_1} \right], \\ \frac{dN_2}{dt} = (b_{21}k_1 + b_{22}k_2 - m_2)N_2 \left[ \frac{b_{21}k_1 + b_{22}k_2 - m_2 - \frac{b_{21}k_1c_{11} + b_{22}k_2c_{21}}{b_{21}k_1c_{12} + b_{22}k_2c_{22}}N_1 - N_2}{b_{21}k_1 + b_{22}k_2 - m_2} \right]. \end{cases} \quad (\text{Equation 12})$$

Essentially, Equation 12 are the classic LV competition equations, which can be easily recognized if we make the following substitutions

$$\begin{aligned} r_1 &\triangleq b_{11}k_1 + b_{12}k_2 - m_1, \\ r_2 &\triangleq b_{21}k_1 + b_{22}k_2 - m_2, \\ K_1 &\triangleq \frac{b_{11}k_1 + b_{12}k_2 - m_1}{b_{11}k_1c_{11} + b_{12}k_2c_{21}}, \\ K_2 &\triangleq \frac{b_{21}k_1 + b_{22}k_2 - m_2}{b_{21}k_1c_{12} + b_{22}k_2c_{22}}, \end{aligned} \quad (\text{Equation 13})$$

and

$$\begin{aligned} a_{12} &\triangleq \frac{b_{11}k_1c_{12} + b_{12}k_2c_{22}}{b_{11}k_1c_{11} + b_{12}k_2c_{21}} = \frac{C_2(b_{11}k_1n_{12} + b_{12}k_2n_{22})}{C_1(b_{11}k_1n_{11} + b_{12}k_2n_{21})} = \frac{C_2}{C_1} \cdot \text{normalization constant}, \\ a_{21} &\triangleq \frac{b_{21}k_1c_{11} + b_{22}k_2c_{21}}{b_{21}k_1c_{12} + b_{22}k_2c_{22}} = \frac{C_1(b_{21}k_1n_{11} + b_{22}k_2n_{21})}{C_2(b_{21}k_1n_{12} + b_{22}k_2n_{22})} = \frac{C_1}{C_2} \cdot \text{normalization constant}, \end{aligned} \quad (\text{Equation 14})$$

where  $r_i$  is the intrinsic rate of natural increase of consumer  $i$ ,  $K_i$  is the carrying capacity of consumer  $i$ , and  $a_{ij}$  is the competition coefficient measuring the competitive effect of consumer  $j$  on consumer  $i$ . Specifically,  $C_j = \frac{c_{ij}}{n_{ij}}$  is the average rate at which consumer  $i$  consumes resources, where  $n_{ij}$  is a normalization constant. Furthermore, it is typically observed that the relative conversion  $b_{ij}$  of a unit of resource  $j$  by consumer  $i$  and the carrying capacity  $k_j$  of resource  $j$  have limited variations for a given consumer  $i$  across different resources  $j$  and can be assumed constant for the sake of model simplification.<sup>75</sup> Therefore, as indicated by Equation 14, it is important to note that the competition coefficient  $a_{ij}$  is proportional to the ratio of consumption rates  $C$  between consumer  $j$  and consumer  $i$ . Formally, we can represent it as

$$a_{ij} = a_0 \left( \frac{C_j}{C_i} \right), \quad (\text{Equation 15})$$

where  $a_0$  is a normalization constant. Following this, we know the intraspecific competition coefficient

$$a_{ii} = a_0 \left( \frac{C_i}{C_i} \right) = a_0. \quad (\text{Equation 16})$$

Without loss of generality, we can set  $a_0 = 1$ . Consequently, Equation 12 become the LV competition equations:

$$\begin{cases} \frac{dN_1}{dt} = r_1 N_1 \left( \frac{K_1 - N_1 - a_{12}N_2}{K_1} \right), \\ \frac{dN_2}{dt} = r_2 N_2 \left( \frac{K_2 - a_{21}N_1 - N_2}{K_2} \right). \end{cases} \quad (\text{Equation 17})$$

Considering the system  $S = \{1, \dots, |S|\}$ , we can write the general form of the LV competition equations as

$$\frac{dN_i}{dt} = r_i N_i \left( \frac{K_i - N_i - \sum_{j \in S \setminus \{i\}} a_{ij} N_j}{K_i} \right), i \in S. \quad (\text{Equation 18})$$

In the previous discussion, our focus has been on the scenario of two consumers and two resources. Next, we will proceed to simplify the scenario by considering only one consumer (also predator, either consumer  $j$  or consumer  $i$ ) and one resource (also prey). Specifically, if the resource is self-replicating, then the resource equation with logistic growth is

$$\frac{dR}{dt} = r_{\text{rsc}} R \left[ \left( \frac{k - R}{k} \right) - CN \right], \quad (\text{Equation 19})$$

where  $R$  is the density of the resource,  $r_{\text{rsc}}$  is its intrinsic rate of increase,  $k$  is the carrying capacity of the resource,  $C$  is the rate at which the resource is consumed by the consumer, and  $N$  is the population density of the consumer.

Following the LV predator-prey equations, we can write the consumer equation as

$$\frac{dN}{dt} = (bR - m)N, \quad (\text{Equation 20})$$

where  $b$  is the predation rate and  $m$  is the death rate of the consumer.

Assuming the resource dynamics are very fast compared to the consumer dynamics, similar to the idea in [Equation 6](#), we can solve the equilibrium value of  $R$  first:

$$R^* = k(1 - CN), \quad (\text{Equation 21})$$

and then substitute  $R^*$  in place of  $R$  in the consumer equation [Equation 20](#). We obtain

$$\frac{dN}{dt} = [bk(1 - CN) - m]N, \quad (\text{Equation 22})$$

which can also be written as

$$\frac{dN}{dt} = N(bk - m - bkCN). \quad (\text{Equation 23})$$

Multiplying and dividing [Equation 23](#) by  $(bk - m)$  gives us

$$\frac{dN}{dt} = (bk - m)N \left( \frac{bk - m - bkCN}{bk - m} \right). \quad (\text{Equation 24})$$

Next, dividing both the numerator and the denominator of the fraction on the right side of the [Equation 24](#) by  $bkC$ , we have

$$\frac{dN}{dt} = (bk - m)N \left( \frac{\frac{bk - m}{bkC} - N}{\frac{bk - m}{bkC}} \right). \quad (\text{Equation 25})$$

If we denote

$$\begin{aligned} r &\triangleq bk - m, \\ K &\triangleq \frac{bk - m}{bkC}, \end{aligned} \quad (\text{Equation 26})$$

then [Equation 25](#) can be written as the classic population model with logistic growth:

$$\frac{dN}{dt} = rN \left( \frac{K - N}{K} \right), \quad (\text{Equation 27})$$

where  $r$  is the intrinsic rate of natural increase of the consumer in isolation and  $K$  is the carrying capacity of the consumer.

Importantly, [Equation 26](#) indicate that the consumption rate  $C$  increases monotonically with the intrinsic growth rate  $r$ . Formally, by substituting the expression for  $bk$  as  $(r + m)$  from the first equation into the second equation, we obtain

$$C = \frac{1}{K \left( 1 + \frac{m}{r} \right)}. \quad (\text{Equation 28})$$

[Figure S10](#) shows the plots of [Equation 28](#) with different death rates  $m$ , indicating that the consumption rate  $C$  exhibits a monotonically increasing behavior with respect to the intrinsic growth rate  $r$ . Also, the presence of an upper limit  $\frac{1}{K}$  on  $C$  aligns with our empirical expectations, as it reflects the fact that the consumption rate cannot infinitely increase. Despite the relationship between  $C$  and  $r$  being nonlinear, it can be effectively approximated using a linear function, especially when  $m$  is large or comparable with the range of  $r$ . Note that the relationship between carrying capacity  $K$  and intrinsic growth rate  $r$  appears to be complex and context-dependent.<sup>87</sup> Therefore, instead of establishing relationships between  $r$  and  $K$ , our primary focus lies on  $r$  due to the relevant datasets available for microbial, plant, and metazoan systems. Nevertheless, if a negative relationship exists between  $r$  and  $K$ , such as  $K = \frac{1}{r}$ , then  $C$  continues to increase monotonically with  $r$ . In this instance, [Equation 28](#) transforms to  $C = \frac{r^2}{r+m}$ , indicating a higher order of scaling in the relationship between  $C$  and  $r$ . This intensified relationship can potentially improve the performance of our theoretical framework. Exploring this additional complexity could be an interesting direction for future research.

### Trade-offs induced by increasing growth rates ( $r$ )

Up to now, we have been discussing how to infer interactions  $a_{ij}$  in the  $K$ -formalism of LV equations by linking them to the Consumer-Resource systems. Following the  $r$ -formalism of LV equations, it is evident that both interspecific and intraspecific interactions ( $\alpha_{ij}, \alpha_{ii}$ , respectively) tend to increase with growth rates  $r$ . Specifically, comparing the LV competition equation [Equation 18](#) in  $K$ -formalism with the generalized LV model below in  $r$ -formalism

$$\frac{dN_i}{dt} = N_i \left( r_i - \sum_{j \in S} \alpha_{ij} N_j \right), \quad (\text{Equation 29})$$

it becomes clear that

$$\alpha_{ii} = \frac{r_i}{K_i} \quad (\text{Equation 30})$$

and

$$\alpha_{ij} = \frac{r_i}{K_i} a_{ij} = \alpha_{ii} a_{ij} \quad (\text{Equation 31})$$

This relationship shows that the intraspecific interaction  $\alpha_{ii}$  is directly proportional to  $r_i$ . In terms of the interspecific interaction  $\alpha_{ij}$ , we first deduce the expression of  $a_{ij}$  explicitly integrating [Equations 15 and 28](#)

$$a_{ij} = \frac{C_j}{C_i} = \frac{\frac{1}{K_j \left(1 + \frac{m_j}{r_j}\right)}}{\frac{1}{K_i \left(1 + \frac{m_i}{r_i}\right)}} = \frac{K_i \left(1 + \frac{m_i}{r_i}\right)}{K_j \left(1 + \frac{m_j}{r_j}\right)}. \quad (\text{Equation 32})$$

Then, following [Equations 30 and 31](#), we have

$$\alpha_{ij} = \alpha_{ii} a_{ij} = \frac{r_i}{K_i} \cdot \frac{K_i \left(1 + \frac{m_i}{r_i}\right)}{K_j \left(1 + \frac{m_j}{r_j}\right)} = \frac{r_i + m_i}{K_j \left(1 + \frac{m_j}{r_j}\right)}. \quad (\text{Equation 33})$$

Therefore, the interspecific interaction  $\alpha_{ij}$  increases with growth rates  $r_i$  and  $r_j$ .

In summary, increases in growth rate translate into a trade-off where they carry increases in interspecific competition ( $\alpha_{ij}$ ), but also they carry increases in density dependence (intraspecific competition  $\alpha_{ii}$ ).

### Inferring consumption rates (C) using maximum growth rates (rmax)

Next, based on [Equation 28](#), the consumption rate  $C$  can be inferred to be approximately proportional to the intrinsic growth rate ( $C \propto r$ ), which can be further approximated by the maximum growth rate ( $r \approx r_{\max}$ ). Formally, this relationship can be expressed as

$$C \propto r_{\max}. \quad (\text{Equation 34})$$

Here, we choose  $r_{\max}$  as a parameter of interest because of its extensive investigation in empirical studies, particularly regarding its metabolic scaling laws. This choice allows for a systematic estimation of consumption rates  $C$  across a wide range of ecological systems, including bacteria, plants, and metazoans, leveraging the wealth of experimental datasets available.

Eventually, by combining [Equation 15](#) with [Equation 34](#), we can derive an estimation of competitive interaction strength  $a_{ij}$  using the ratio of maximum growth rate ( $r_{\max}$ ) between consumer (taxon)  $j$  and consumer  $i$  ([Figure 1B](#)):

$$a_{ij} = a_0 \left( \frac{C_j}{C_i} \right) \propto a'_0 \left( \frac{r_{\max_j}}{r_{\max_i}} \right), \quad (\text{Equation 35})$$

where  $C_i$  and  $C_j$  are the consumption rates of taxon  $i$  and taxon  $j$ , respectively;  $a_0$  and  $a'_0$  are normalization constants. Based on this estimation, we can obtain the pairwise interaction matrix  $\mathbf{A}$  of the Lotka-Volterra model ([Equation 1](#)), which is the only parameter needed to implement the feasibility analysis. Phenomenologically, these competitive effects can represent either resource partition or modification.<sup>29,72–74</sup>

It is important to note that in our theoretical framework, we standardize parameters like carrying capacity (mentioned above) and dispersal ability (next section) to be equal across all taxa, allowing differentiation solely in their  $r_{\max}$ . This level of control offers a coarse-grained explanation of the observed empirical patterns (see [Figures 3 and 4](#)). Indeed, adjustments to these other parameters could further enhance the accuracy of the theoretical path of least resistance, enabling it to better fit the observed order of appearance in empirical systems compared to our current [Figure 3](#). This could be an interesting avenue for future research.

### Quantification of the transition matrix (P) in the Markov chain

In the Markov chain of each developmental process, either sequentially or simultaneously, the transition matrix  $\mathbf{P}$  plays a critical role as it summarizes the likelihood of transitioning from one ecological community to another. Mathematically, the transition matrix of a developmental process is a square matrix, where the rows and columns represent the ecological communities characterized by

different taxon compositions. The matrix size is equal to the number of all possible ecological communities in the developmental process. Each element in the matrix represents the probability of transitioning from the community corresponding to the row to the community corresponding to the column.

Here, we consider bottom-up developmental processes that allow for loops wherein a community may transition back to itself. Specifically, communities comprising  $n$  taxa can either transition to communities comprising  $(n + 1)$  taxa with an additional taxon available in the taxon pool, or they can loop back to themselves and remain unchanged. This coarse-grained model suffices to provide an overall expectation of taxon appearance orders in ecological development. Note that considering cycles (e.g., disassemblies caused by environmental perturbations), which allow a larger community to collapse into a smaller one, does not affect our theoretical conclusions regarding the path of least resistance (refer to the section “[identification of the path of least resistance](#)”).

The likelihood of community transitions within an ecological system is determined by the feasibility of all possible communities within it. Below is a detailed example using the three-taxon system  $S$  illustrated in [Figure 1](#), which comprises taxa tx1, tx2, and tx3 (i.e.,  $S = \{123\}$ ). We will systematically explore the quantification of transition matrices in the Markov chains that describe sequential development from initial communities with one or two taxa, as well as simultaneous development from the community containing all three taxa. Then the three transition matrices are presented. Using these matrices, we can calculate the cumulative probability over  $x$  trials for developing an ecological system.

Initially, all developmental processes start from an empty set  $\emptyset$ . Note that, as mentioned before, we focus on the order in which taxa appear during development, rather than the specific composition of resulting communities. Thus, this empty set provides a standardized baseline, allowing for meaningful comparisons of appearance orders across various types of development, regardless of whether a system starts with no taxa or undergoes a reset due to disturbances. Subsequently, the system may either remain in the empty set or transition into communities containing varying numbers of taxa. On the one hand, the probability of remaining empty is the complement of the average feasibility of all possible initial communities. If the initial system contains one taxon ([Figure 1D](#)), the probability that the system remains empty is  $\mathbf{P}_1[\emptyset, \emptyset] = 1 - \frac{P(\{1\}) + P(\{2\}) + P(\{3\})}{3}$ , where  $P(\{1\})$ ,  $P(\{2\})$ , and  $P(\{3\})$  represent the feasibility of the communities with tx1, tx2, and tx3, respectively (see [Equation 3](#)). Notably, the likelihood of any single taxon being feasible in unknown environments is 50% (i.e.,  $P(\{1\}) = P(\{2\}) = P(\{3\}) = 0.5$ ), similar to flipping a fair coin. This equal feasibility implies that all taxa have the same dispersal ability and are not restricted to appear in a specific order because of this. Instead, their order of appearance is driven by their growth rates and the current subset of present taxa (their interactions). If the initial system contains two taxa, then the probability of remaining empty is  $\mathbf{P}_2[\emptyset, \emptyset] = 1 - \frac{P(\{12\}) + P(\{13\}) + P(\{23\})}{3}$ , where  $P(\{12\})$ ,  $P(\{13\})$ , and  $P(\{23\})$  represent the feasibility of the communities with pairs tx1 and tx2, tx1 and tx3, and tx2 and tx3, respectively. If all three taxa appear initially and simultaneously, the probability of remaining in the empty set is  $\mathbf{P}_3[\emptyset, \emptyset] = 1 - P(\{123\})$ , where  $P(\{123\})$  is the feasibility of all three taxa coexisting together in system  $S$  (see [Figure 1E](#)). On the other hand, the probabilities of transitioning to different initial communities are proportional to their respective feasibility. Importantly, by the definition of a Markov chain, these probabilities collectively sum to  $1 - \mathbf{P}_1[\emptyset, \emptyset]$ . For example, the transition probabilities from the empty set  $\emptyset$  to the three one-taxon initial communities satisfy  $\mathbf{P}_1[\emptyset, \{1\}] : \mathbf{P}_1[\emptyset, \{2\}] : \mathbf{P}_1[\emptyset, \{3\}] = P(\{1\}) : P(\{2\}) : P(\{3\})$  and  $\mathbf{P}_1[\emptyset, \{1\}] + \mathbf{P}_1[\emptyset, \{2\}] + \mathbf{P}_1[\emptyset, \{3\}] = 1 - \mathbf{P}_1[\emptyset, \emptyset]$ . For initial communities containing two taxa, the transition probabilities from empty satisfy  $\mathbf{P}_2[\emptyset, \{12\}] : \mathbf{P}_2[\emptyset, \{13\}] : \mathbf{P}_2[\emptyset, \{23\}] = P(\{12\}) : P(\{13\}) : P(\{23\})$  and  $\mathbf{P}_2[\emptyset, \{12\}] + \mathbf{P}_2[\emptyset, \{13\}] + \mathbf{P}_2[\emptyset, \{23\}] = 1 - \mathbf{P}_2[\emptyset, \emptyset]$ .

For intermediate communities, a community may either remain unchanged or transition into larger communities by incorporating an additional taxon. The probability of remaining in the current community equals the feasibility of that community; for example,  $\mathbf{P}_1[\{1\}, \{1\}] = P(\{1\})$ . Alternatively, the probabilities of transitioning into larger communities are proportional to their respective feasibility. For instance, a community with tx1 may transition to a larger community by including an additional taxon in the pool, either tx2 or tx3. The probabilities for these transitions are proportional to their feasibility, namely,  $\mathbf{P}_1[\{1\}, \{12\}] : \mathbf{P}_1[\{1\}, \{13\}] = P(\{12\}) : P(\{13\})$ . These transition probabilities are located in the row of  $\{1\}$  and the columns of  $\{12\}$  and  $\{13\}$  within the transition matrix  $\mathbf{P}_1$ . Similarly, for the community comprising tx1 and tx2, the probability of maintaining the current composition is equivalent to the feasibility of the community,  $P(\{12\})$ . Alternatively, this community may transition into a larger community that includes tx3, the only remaining candidate in the taxon pool, with a complementary probability of  $(1 - P(\{12\}))$ . These transition probabilities are then cataloged in the row of  $\{12\}$  and the columns of  $\{12\}$  and  $\{123\}$  within the transition matrix  $\mathbf{P}_1$ . In terms of intermediate communities, the transition probabilities for developmental processes that start with two or more taxa (e.g.,  $\mathbf{P}_2$ ) can be quantified in the same way as those described above for processes starting with a single taxon.

Lastly, the transition probability for the community consisting of tx1, tx2, and tx3 (all taxa in the pool) returning to itself is 1. This probability is saved in the row and column  $\{123\}$  of the transition matrices  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ , and  $\mathbf{P}_3$ . This indicates that the full community  $\{123\}$  is an absorbing state. Once all taxa are assembled, the community does not undergo further changes, serving as an endpoint in the sequence of community changes. Note that this absorbing state is the sufficient (but not necessary) condition for development. If there is interest in exploring beyond the endpoint community to understand how additional taxa might integrate, it is also possible to delay reaching an endpoint by expanding the absorbing state. For example, this can be done by increasing the number of taxa in the initial pool and analyzing the dynamics of the intermediate communities that assimilate these additional taxa of interest.

The Markov chain allows both sequential and simultaneous developments to be quantified using the same set of rules, establishing a basis for their comparability. Notably, the simultaneous development, due to its simple structure, can be analytically shown to be equivalent to a binomial process (refer to the following section for more details). For instance, the transition matrix  $\mathbf{P}_1$  for developing a three-taxon system sequentially from one-taxon communities can be formally represented as

$$\begin{matrix} & \emptyset & \{1\} & \{2\} & \{3\} & \{12\} & \{13\} & \{23\} & \{123\} \\ \mathbf{P}_1 = & \begin{bmatrix} 1 - \frac{P(\{1\})+P(\{2\})+P(\{3\})}{3} & \frac{P(\{1\})}{3} & \frac{P(\{2\})}{3} & \frac{P(\{3\})}{3} & 0 & 0 & 0 & 0 \\ 0 & P(\{1\}) & 0 & 0 & \frac{(1-P(\{1\}))P(\{12\})}{P(\{12\})+P(\{13\})} & \frac{(1-P(\{1\}))P(\{13\})}{P(\{12\})+P(\{13\})} & 0 & 0 \\ 0 & 0 & P(\{2\}) & 0 & \frac{(1-P(\{2\}))P(\{12\})}{P(\{12\})+P(\{23\})} & 0 & \frac{(1-P(\{2\}))P(\{23\})}{P(\{12\})+P(\{23\})} & 0 \\ 0 & 0 & 0 & P(\{3\}) & 0 & \frac{(1-P(\{3\}))P(\{13\})}{P(\{13\})+P(\{23\})} & \frac{(1-P(\{3\}))P(\{23\})}{P(\{13\})+P(\{23\})} & 0 \\ 0 & 0 & 0 & 0 & P(\{12\}) & 0 & 0 & 1 - P(\{12\}) \\ 0 & 0 & 0 & 0 & 0 & P(\{13\}) & 0 & 1 - P(\{13\}) \\ 0 & 0 & 0 & 0 & 0 & 0 & P(\{23\}) & 1 - P(\{23\}) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \end{matrix} \quad (\text{Equation 36})$$

Similarly, the transition matrix  $\mathbf{P}_2$  for developing a three-taxon system sequentially from two-taxon communities can be represented as

$$\begin{matrix} & \emptyset & \{12\} & \{13\} & \{23\} & \{123\} \\ \mathbf{P}_2 = & \begin{bmatrix} 1 - \frac{P(\{12\})+P(\{13\})+P(\{23\})}{3} & \frac{P(\{12\})}{3} & \frac{P(\{13\})}{3} & \frac{P(\{23\})}{3} & 0 \\ 0 & P(\{12\}) & 0 & 0 & 1 - P(\{12\}) \\ 0 & 0 & P(\{13\}) & 0 & 1 - P(\{13\}) \\ 0 & 0 & 0 & P(\{23\}) & 1 - P(\{23\}) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \end{matrix} \quad (\text{Equation 37})$$

In simultaneous development, the transition matrix  $\mathbf{P}_3$  for developing a three-taxon system can be represented as

$$\begin{matrix} & \emptyset & \{123\} \\ \mathbf{P}_3 = & \begin{bmatrix} 1 - P(\{123\}) & P(\{123\}) \\ 0 & 1 \end{bmatrix} & \end{matrix} \quad (\text{Equation 38})$$

A key property of the Markov chain is that raising the transition matrix  $\mathbf{P}$  to the power of  $x$  yields the probabilities of transitions between ecological communities with  $x$  steps (or trials). Based on this property, we can calculate the cumulative probability  $\mathbf{P}^x[\emptyset, \{S\}]$  of developing an  $n$ -taxon ecological system  $S$  over  $x$  attempts through sequential processes with varying numbers of initial taxa or the simultaneous process (see Equations 36, 37, and 38 and Figure 1F). Moreover, since  $x$  is an absorbing state, the cumulative probability  $\mathbf{P}^x[\emptyset, \{S\}]$  across all developmental processes would always converge to probability 1 with sufficiently large numbers of trials. However, the rate of convergence can vary depending on the sizes of the initial communities. Therefore, it is possible to identify the process with the highest likelihood of successfully developing the entire system within a limited number of trials (see results in Figures 2A, S1A, and S1E). In short, the larger the initial communities, the slower the convergence. Thus, as detailed in the following section, convergence to probability 1 occurs more slowly in simultaneous development compared to sequential development.

### Understanding simultaneous development as a binomial process

Let  $\mathbf{P}$  be the transition matrix of a Markov chain describing the simultaneous development of an  $n$ -taxon ecological system  $S$ . Analogous to the transition matrix of a three-taxon system in Equation 38, the general form for an  $n$ -taxon system is

$$\begin{matrix} & \emptyset & \{S\} \\ \mathbf{P} = & \begin{bmatrix} 1 - P(S) & P(S) \\ 0 & 1 \end{bmatrix}, & \end{matrix} \quad (\text{Equation 39})$$

where  $\mathbf{P}[\emptyset, \{S\}] = P(S)$  denotes the probability of transitioning from the empty set  $\emptyset$  to the ecological system  $S$  (see Figure 1E). This transition probability is quantified as the feasibility of the  $n$ -taxon system, indicating the likelihood of achieving simultaneous coexistence (see Equation 3).

We propose that the  $x$ -step transition matrix  $\mathbf{P}^x$  is given by the following formula for all integers  $x \geq 1$ :

$$\mathbf{P}^x = \begin{matrix} & \emptyset & \{S\} \\ \emptyset & (1 - P(S))^x & 1 - (1 - P(S))^x \\ \{S\} & 0 & 1 \end{matrix} \quad (\text{Equation 40})$$

This formula will be proven by mathematical induction, starting with the base case when  $x = 1$ . The one-step transition matrix  $\mathbf{P}^1$  is simply

$$\mathbf{P}^1 = \begin{matrix} & \emptyset & \{S\} \\ \emptyset & 1 - P(S) & P(S) \\ \{S\} & 0 & 1 \end{matrix}, \quad (\text{Equation 41})$$

which matches the given matrix in Equation 39. Thus, the base case holds.

Assuming inductively that the formula Equation 40 holds for some arbitrary integer  $x \geq 1$ , we proceed to show that this formula also holds for  $(x + 1)$ . The  $(x + 1)$ -step transition matrix  $\mathbf{P}^{x+1}$  can be computed by multiplying  $\mathbf{P}^x$  by  $\mathbf{P}$ . Thus,

$$\begin{aligned} \mathbf{P}^{x+1} = \mathbf{P}^x \mathbf{P} &= \begin{bmatrix} (1 - P(S))^x & 1 - (1 - P(S))^x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 - P(S) & P(S) \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (1 - P(S))^x \cdot (1 - P(S)) & (1 - P(S))^x \cdot P(S) + (1 - (1 - P(S))^x) \cdot 1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (1 - P(S))^{x+1} & 1 - (1 - P(S))^{x+1} \\ 0 & 1 \end{bmatrix}. \end{aligned} \quad (\text{Equation 42})$$

Clearly, the formula Equation 40 holds for  $(x + 1)$ . By induction, this confirms that for all integers  $x \geq 1$ , the  $x$ -step transition matrix  $\mathbf{P}^x$  is given by Equation 40.

Taking a closer look at Equation 40, it is interesting to notice that the simultaneous process can be characterized not only as a Markov chain (Equation 39) but also as a binomial process, similar to flipping a coin. Specifically, consider the analogy of flipping a biased coin, where

- Heads (success) corresponds to transitioning to the system  $S$ . This occurs with a probability of  $P(S)$ .
- Tails (failure) corresponds to remaining in the empty set  $\emptyset$ . This occurs with a probability of  $1 - P(S)$ .

If the coin is flipped  $x$  times (representing  $x$  steps in the Markov chain), then the probability of flipping at least one head (transitioning to system  $S$  at least once) during these  $x$  attempts is  $1 - (1 - P(S))^x$ . Note that  $(1 - P(S))^x$  is the probability of the complement event, which is getting all tails (never transitioning to system  $S$  in any of the  $x$  attempts). As  $x$  increases, this complement probability approaches zero because the effect of repeatedly multiplying a number less than one diminishes its value towards zero. Thus,  $1 - (1 - P(S))^x$  converges to 1, indicating that with a sufficiently large number of attempts, transitioning to system  $S$  through simultaneous development becomes almost certain. However, since the probability of all taxa coexisting  $P(S)$  is very low (more taxa lead to harder coexistence and thus a lower probability), the convergence of simultaneous development is significantly slower compared to sequential development (Figures 2A, S1A, and S1E).

Therefore, the binomial process offers an intuitive way to understand the inherent stochastic processes that influence the simultaneous development of an ecological system. It quantifies the likelihood that the system successfully reaches simultaneous coexistence at least once within a given number of attempts.

### Identification of the path of least resistance

We define development as the self-organized process of moving from an empty set to a set of  $n$  taxa through a collection of different taxon communities (interactions) whose union forms the set of  $n$  taxa (i.e., the appearance of a group of taxa). In the main text, we focus our analysis on the special case of directional assembly (once a taxon is successfully introduced it cannot go extinct). Effectively, this analysis aims to explain the probability that a group of  $n$  species can be found together at a given point in time. While this analysis is a simplification of the developmental process undergone by natural systems (i.e., no extinctions); under our assumptions, this analysis generates the same results as if additionally considering disassembly processes (see section below). That is, under assembly only, the question is which one path is the system most likely to take. Under assembly and disassembly, the question would be which is the most visited path by the system.

Formally, a directional assembly (one-by-one bottom-up developmental process) of an  $n$ -taxon ecological system  $S$  is a sequence of community changes  $\emptyset, C_1, C_2, \dots, C_{n-1}, S$  consisting of  $0, 1, 2, \dots, n - 1, n$  taxa, respectively (for every  $i$  such that  $1 \leq i \leq n - 1$ ,

$C_i \subset S$ ). In a Markov chain, the probability of observing this sequence can be calculated by multiplying the initial probability with the conditional probabilities of each transition:

$$P(\emptyset, C_1, C_2, \dots, C_{n-1}, S) = P(\emptyset) \times P(C_1|\emptyset) \times P(C_2|C_1) \times \dots \times P(C_{n-1}|C_{n-2}) \times P(S|C_{n-1}), \quad (\text{Equation 43})$$

where  $P(\emptyset) = 1$  is the probability of developing from an empty set. This standardized baseline, with equal probabilities for all paths, allows for meaningful comparisons. In the context of full environmental uncertainty (no a priori information), the probability  $P(C_1|\emptyset)$  of having a single taxon is analogous to flipping a fair coin, with a constant 50% chance of success. Ecologically, this constant probability suggests that all taxa have the same likelihood of appearing. Also, this probability can be adjusted in light of prior environmental information.<sup>74</sup> Moreover, it is important to note that we do not assume any relationship between a taxon's dispersal probability and its maximum growth rate. However, should there be a positive relationship, it would result in an even higher probability  $P(\emptyset, C_1, C_2, \dots, C_{n-1}, S)$  for the path of least resistance, thus enhancing the performance of our theoretical framework and the most likely appearance order from fast- to slow-growing taxa. Incorporating this additional complexity offers a compelling direction for future research. Additionally,  $P(C_m|C_{m-1})$  is the transition probability from community  $C_{m-1}$  to community  $C_m$  ( $m \leq n$ ,  $C_0 = \emptyset$ ,  $C_n = S$ ), which is the element located at the intersection of row  $(m - 1)$  and column  $(m)$  in the transition matrix  $\mathbf{P}$ . This equation holds because the Markov chain has the memoryless property that a future system depends only on the current system, not on any preceding systems in the sequence. The effects of memory in ecological developmental processes could be an interesting avenue for future research.

To identify the path of least resistance, the objective is to find the sequence of ecological systems that maximizes the probability  $P(\emptyset, C_1, C_2, \dots, C_{n-1}, S)$ . This can be conveniently achieved by transforming the multiplication operation in Equation 43 into a summation operation. Specifically, this transformation can be achieved by applying the negative logarithm function ( $-\log$ ) to both sides of the equation. Any base greater than 1 is valid, including common bases such as 10,  $e$ , and 2. That is:

$$\begin{aligned} & -\log[P(\emptyset, C_1, C_2, \dots, C_{n-1}, S)] \\ &= -\log[P(\emptyset) \times P(C_1|\emptyset) \times P(C_2|C_1) \times \dots \times P(C_{n-1}|C_{n-2}) \times P(S|C_{n-1})] \\ &= -\log(1) - \log[P(C_1|\emptyset)] - \log[P(C_2|C_1)] - \dots - \log[P(C_{n-1}|C_{n-2})] - \log[P(S|C_{n-1})], \end{aligned} \quad (\text{Equation 44})$$

where  $P(\emptyset) = 1$  is a constant. The R codes accompanying this paper use the negative natural logarithm function ( $-\ln$ ) with a base of  $e$ .

The objective now shifts to minimizing  $-\log[P(\emptyset, C_1, C_2, \dots, C_{n-1}, S)]$ , which is essentially minimizing  $-\sum_{m=1}^n \log[P(C_m|C_{m-1})]$ , where  $C_0 = \emptyset$  and  $C_n = S$ . In graphical terms, our objective is to find the shortest path in the deterministic tree-like directed graph characterized by the adjacency matrix  $\mathbf{A}$ . In this graph, vertices represent ecological communities of different taxon compositions. Besides, each edge carries a weight  $-\log[P(C_m|C_{m-1})]$  that represents the distance from its connected vertex  $C_{m-1}$  to the other connected vertex  $C_m$ . In the context of ecological development, the shortest path is from a vertex representing the empty set  $\emptyset$  to a vertex representing the most complex community  $S$ . According to graph theory, Dijkstra's algorithm provides an efficient solution to this shortest path problem.<sup>88</sup> The R codes provided use the function "shortest\_paths" from the "igraph" package for these computations. The path with the shortest distance among all paths is defined as the path of least resistance for developing an ecological system (e.g., pink water pipes in Figure 1E and pink arrows in Figure S1D).

All computational tasks were executed using the High-Performance Computing resources available through the MIT SuperCloud and the Lincoln Laboratory Supercomputing Center.<sup>89</sup> The R codes are accessible on GitHub, with the link detailed in the [key resources table](#).

### Theoretical predictions are not influenced by disassembly

As mentioned in the main text, our analysis focuses on the special case of directional assembly, wherein once a taxon is successfully introduced, extinction is precluded. Though this assumption simplifies the natural developmental process, it effectively serves our purpose of determining the expected order of appearance of taxa. It is important to note our theoretical predictions remain consistent even when disassembly processes are considered additionally. With assembly only, the question is which path the system is most likely to follow. Conversely, with both assembly and disassembly, the question is which path is the most frequently visited path by the system.

In this section, we will start by developing ecological intuitions to understand why disassembly does not affect our theoretical predictions. Following this, we will incorporate randomly generated disassembly and repeat the analysis presented in Figure 2. This allows us to systematically and quantitatively assess the influence of disassembly.

In the special case of directional assembly only (see Figure 1D), sequential development from smaller initial communities can diverge into multiple paths. This makes it more efficient to reach the largest community compared to those with a greater number of initial taxa. Even with disassembly present (Figure S1D), these finely bifurcated paths from smaller initial communities can still expedite the progression towards the largest community. Therefore, it can be anticipated that the cumulative probability of sequential development from a single taxon reaches its saturation faster than others with more initial taxa, including simultaneous development (i.e., assembling all taxa in the pool at once). Additionally, a developmental process progresses in a single direction in the case of assembly only, resulting in a saturation probability of 1 (y-axis in Figure 2A). This indicates that once taxa are introduced, the system inevitably advances towards the formation of the largest community without any reversals. However, when disassembly processes are incorporated, a new possibility arises. Disassembly introduces the potential for species loss and regression to less developed



communities, disrupting the progression towards the largest community. Hence, the cumulative probability of reaching the largest community is expected to fall below 1. Essentially, the inclusion of disassembly introduces additional opportunities for deviations from the most likely developmental trajectory, leading to a lower cumulative probability of development.

Interestingly, despite the potential impact of disassembly, the most likely developmental order (the path of least resistance) remains unaffected. Let us take the transition matrix **P** in Equation 36 as an example, which describes the sequential development for a three-taxon system sequentially from one-taxon communities.

$$\begin{matrix}
 & \emptyset & \{1\} & \{2\} & \{3\} & \{12\} & \{13\} & \{23\} & \{123\} \\
 \begin{matrix}
 \emptyset \\
 \{1\} \\
 \{2\} \\
 \{3\} \\
 \{12\} \\
 \{13\} \\
 \{23\} \\
 \{123\}
 \end{matrix} & \mathbf{P} = & \begin{bmatrix}
 1 - \frac{P(\{1\})+P(\{2\})+P(\{3\})}{3} & \frac{P(\{1\})}{3} & \frac{P(\{2\})}{3} & \frac{P(\{3\})}{3} & 0 & 0 & 0 & 0 & 0 \\
 0 & P(\{1\}) & 0 & 0 & \frac{(1-P(\{1\}))P(\{12\})}{P(\{12\})+P(\{13\})} & \frac{(1-P(\{1\}))P(\{13\})}{P(\{12\})+P(\{13\})} & 0 & 0 & 0 \\
 0 & 0 & P(\{2\}) & 0 & \frac{(1-P(\{2\}))P(\{12\})}{P(\{12\})+P(\{23\})} & 0 & \frac{(1-P(\{2\}))P(\{23\})}{P(\{12\})+P(\{23\})} & 0 & 0 \\
 0 & 0 & 0 & P(\{3\}) & 0 & \frac{(1-P(\{3\}))P(\{13\})}{P(\{13\})+P(\{23\})} & \frac{(1-P(\{3\}))P(\{23\})}{P(\{13\})+P(\{23\})} & 0 & 0 \\
 0 & 0 & 0 & 0 & P(\{12\}) & 0 & 0 & 0 & 1 - P(\{12\}) \\
 0 & 0 & 0 & 0 & 0 & P(\{13\}) & 0 & 0 & 1 - P(\{13\}) \\
 0 & 0 & 0 & 0 & 0 & 0 & P(\{23\}) & 0 & 1 - P(\{23\}) \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & P(\{23\}) & 1 - P(\{23\}) \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{bmatrix}
 \end{matrix}$$

(Equation 45)

By definition, each element in the transition matrix is the probability of transitioning from the community in the row to the community in the column. The upper triangular portion represents the assembly process of an ecological system, and the lower triangular portion represents the disassembly process. Recall that the path of least resistance is defined as the path that is the most likely path for the system to take. Let us assume that  $\emptyset \rightarrow \{1\} \rightarrow \{12\} \rightarrow \{123\}$  is the path of least resistance considering the special case of directional assembly only. That is, the probability  $P(\emptyset, C_1, C_2, \{123\})$  of observing the sequence of community changes from  $\emptyset$  to  $\{123\}$ , i.e.,  $\mathbf{P}[\emptyset, C_1] \times \mathbf{P}[C_1, C_2] \times \mathbf{P}[C_2, \{123\}]$ , can be maximized when  $C_1 = \{1\}$  and  $C_2 = \{12\}$  (see Equation 43). Note that  $\mathbf{P}[\emptyset, \{1\}]$ ,  $\mathbf{P}[\{1\}, \{12\}]$ , and  $\mathbf{P}[\{12\}, \{123\}]$  are all located in the upper triangular portion of the transition matrix (see Equation 45) and therefore influenced solely by the assembly process, not the disassembly process. Ecologically speaking, though assembly and disassembly processes are often intertwined in natural systems, the most likely order in which taxa appear for biodiversity development remains independent of disassembly processes driven by uncertain environmental perturbations. This can be intuitively understood by recognizing that the path of least resistance, or the shortest path, is by nature the most straightforward route, which tends to avoid any return cycles to a less developed community as a result of disassembly (see Figure S1D).

With these intuitions in mind, we aim to systematically verify whether our theoretical predictions in Figure 2 remain valid when additionally considering disassembly. To do this, we randomly generated disassembly probabilities for each community. These random probabilities account for the fact that different communities may disassemble with varying probabilities in response to uncertain environmental perturbations. Note that incorporating disassembly mathematically means filling the lower triangular portion of a transition matrix (e.g., Equation 45) with non-zero values. Specifically, these disassembly probabilities were drawn from a uniform distribution ranging from a minimum of 0 to a maximum of 0.1. The results remain robust to other maximum values. To ensure the validity of the studied developmental process, we enforced that the assembly probability for each transition must be greater than the reverse disassembly probability, thus allowing overall biodiversity to increase. In cases where this condition was not met, we iteratively generated another disassembly probability from the same distribution mentioned above.

Using the transition matrix incorporating both assembly and disassembly processes, we conducted the same analysis as presented in Figure 2A. We computed the cumulative probability for development from the empty set  $\emptyset$  through varying initial communities to the full community  $S$  (refer to the section “quantification of the transition matrix(**P**) in the Markov chain” for more details). The results are presented in Figure S1E, indicating that sequential developments starting from smaller initial communities tend to reach saturation more quickly. Notably, in all developments, the cumulative probability of reaching the largest community falls below 1 (grey dashed line), which is in line with our expectations above. Even with disassembly present, sequential development remains more likely to succeed than simultaneous development. These results are consistent with our theoretical prediction in Figure 2A.

Although the most likely developmental paths begin with communities comprising a single taxon, various sequences with different orders of appearance are possible. For example, both sequences  $\emptyset \rightarrow \{1\} \rightarrow \{12\}$  and  $\emptyset \rightarrow \{2\} \rightarrow \{12\}$  start from a single taxon but introduce the two taxa in opposite orders. It is therefore important to identify the single most likely path. To this end, we used Dijkstra’s algorithm to find the most likely path, also referred to as the path of least resistance (see the section “identification of the path of least resistance” for more details). With this path established, we conducted the same analysis as presented in Figure 2B. The results, presented in Figure S1F, reveal that when disassembly is considered, the expected correlations between the most likely appearance

order of a taxon and its  $r_{\max}$  are generally negative (predominantly  $y < 0$ ; Spearman's rank correlation). Also, as the standard deviation of  $r_{\max}$  among taxa (x-axis) increases, these negative correlations tend to be stronger (approaching  $y = -1$ ). These results are consistent with our theoretical prediction in Figure 2B.

In summary, with disassembly processes additionally considered, our theoretical predictions shown in Figure 2 remain consistent. Furthermore, the most likely developmental order along the path of least resistance is not altered by disassembly. Thus, the theoretical most likely order for each empirical system presented in Figures 3 and 4 remains the same.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Proxy for maximum growth rate ( $r_{\max}$ )

To estimate the maximum growth rate ( $r_{\max}$ ) for taxa in each empirical system, we employ metabolic scaling laws, which have been providing valuable insights into the fundamental processes that govern biological structure and function.

Specifically, DeLong et al.<sup>76</sup> presents the scaling relationships between  $r_{\max}$  and body mass across prokaryotes (including bacteria and archaea) and metazoans (including birds and herbivores). Interestingly, the authors discover a shift in the scaling pattern of  $r_{\max}$  from positive in prokaryotes (with a power of  $0.73 \approx 3/4$ ) to negative in metazoans (with a power of  $-0.23 \approx -1/4$ ). Therefore, following the scaling relationship  $r_{\max} \propto (\text{body mass})^{-1/4}$ , we use the body mass information reported in the AVONET database<sup>54</sup> and Pansu et al.<sup>18</sup> and Stalmans et al.<sup>19</sup> to estimate the  $r_{\max}$  for birds and herbivores, respectively (Figures 3D–3F).

Similarly, for plants, Enquist et al.<sup>77</sup> demonstrates an inverse scaling relationship between  $r_{\max}$  and body mass, characterized by a power of  $-1/4$ . Additionally, Marbà et al.<sup>53</sup> provides an allometric scaling relating tree mass ( $M$ ) to diameter at breast height (dbh) via the formula  $M = 78 \times \text{dbh}^{2.47}$ . Consequently, to approximate the  $r_{\max}$  for individual tree species, we use diameter at breast height data from the BIEN database<sup>62–65,71</sup> (Figure 3C).

However, due to the considerable variation in microbial body mass under environmental uncertainties and limited reporting, we use the ribosomal RNA operons ( $rrn$ ) copy number, rather than the body mass, as a proxy for  $r_{\max}$  in microbial systems. Roller et al.<sup>78</sup> presents the positive scaling relationship between the  $rrn$  copy number and the  $r_{\max}$  of bacteria and archaea, i.e.,  $r_{\max} \propto (\text{rrn copy number})^{5/4}$ . Also, the authors supply a comprehensive database of  $rrn$  copy numbers called  $rrnDB$ ,<sup>52</sup> from which we obtain the average  $rrn$  copy numbers for the six phyla of infant gut microbes and the five orders of marine bacteria (Figures 3A–3B).

It is important to note that the choice of scaling exponent does not affect the qualitative results of our framework. Let us explain this with the following example. Considering that the body mass distribution  $M_1, M_2, \dots, M_n$  of  $n$  taxa in an ecological system follows a log-normal distribution with a log mean of  $\mu$  and a log standard deviation of  $\sigma$ , then it is clear that  $\log(M_1), \log(M_2), \dots, \log(M_n)$  are normally distributed with a mean of  $\mu$  and a standard deviation of  $\sigma$ . Assuming that a scaling relationship exists between body mass and  $r_{\max}$ , this can be represented as  $r_{\max_i} = M_i^\alpha$ , where  $\alpha \in (-\infty, \infty)$  is a constant scaling exponent. This scaling relationship can also be expressed logarithmically as  $\log(r_{\max_i}) = \log(M_i^\alpha) = \alpha \cdot \log(M_i)$  for each taxon  $i$  within the range  $1, \dots, n$ . Thus,  $\log(r_{\max_1}), \log(r_{\max_2}), \dots, \log(r_{\max_n})$  are also normally distributed with a mean of  $\alpha\mu$  and a standard deviation of  $|\alpha|\sigma$ , where  $|\cdot|$  denotes absolute value. By definition, this means that  $r_{\max_1}, r_{\max_2}, \dots, r_{\max_n}$  of the  $n$  taxa follow a log-normal distribution with a log mean of  $\alpha\mu$  and a log standard deviation of  $|\alpha|\sigma$ . Therefore, the scaling exponent  $\alpha$ , regardless of its sign (due to  $|\cdot|$ ) or magnitude, does not alter the log-normal nature of the distribution of maximum growth rates ( $r_{\max}$ ).