Scale-invariant topology and bursty branching of evolutionary trees emerge from niche construction

Chi Xue[a,b,c], Zhiru Liu[a,b,c], and Nigel Goldenfeld[a,b,c,1]

[a]Loomis Laboratory of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [b]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and [c]Institute for Universal Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Contributed by Nigel Goldenfeld, February 6, 2020 (sent for review August 30, 2019; reviewed by Marcus W. Feldman, Joachim Krug, and Kim Sneppen)

Phylogenetic trees describe both the evolutionary process and community diversity. Recent work has established that they exhibit scale-invariant topology, which quantifies the fact that their branching lies in between the two extreme cases of balanced binary trees and maximally unbalanced ones. In addition, the backbones of phylogenetic trees exhibit bursts of diversification on all timescales. Here, we present a simple, coarse-grained statistical model of niche construction coupled to speciation. Finite-size scaling analysis of the dynamics shows that the resultant phylogenetic tree topology is scale-invariant due to a singularity arising from large niche construction fluctuations that follow extinction events. The same model recapitulates the bursty pattern of diversification in time. These results show how dynamical scaling laws of phylogenetic trees on long timescales can reflect the indelible imprint of the interplay between ecological and evolutionary processes.

niche construction | evolution | scaling laws | molecular phylogeny

Phylogenetic trees represent the evolutionary history of a group of organisms, usually constructed through an appropriate proxy such as the so-called 16SrRNA gene. This gene codes for a part of the translational machinery of the cell and, as such, is presumed to be an essential part of all cellular life. Canonically, this gene (or the 18S variant in Eukaryotes) is used to define operational taxonomic units (OTUs) that correspond to a generalized notion of species. By evaluating the similarity of DNA sequences, the timeline of speciation and subsequent evolution can be estimated. Nodes on a phylogenetic tree represent OTUs, with the external or leaf nodes being extant organisms that are observed and the internal ones being hypothetical organisms that are inferred based on the similarity and the embedded evolutionary process. When a tree is rooted, the top node, or the root, represents the inferred common ancestor of all nodes in the tree. The lengths of the branches of the tree correspond to nucleotide changes, and the timescale is set by a molecular clock assumption.

It is by no means mandatory that evolutionary history should be tree-like in its topology. Prior to the last universal common ancestor, the translational machinery evolved rapidly, and today's canonical genetic code emerged. This code is not only universal, but also is nearly optimal in minimizing errors, in a sense that has been quantified precisely (1, 2). Simulations of the evolution of the genetic code indicate that it would have been highly unlikely for it to have these properties if the evolution had proceeded in a treelike manner; only with horizontal gene transfer of core translational machinery, and thus a network topology for the phylogeny, can a unique and optimal code evolve rapidly (3). Thus, the evidence suggests that there was a major evolutionary transition that occurred during the emergence of the translational machinery; prior to this transition, the concept of species did not exist in the canonical sense. This transition can be inferred from a topology change in the phylogeny of the 16SrRNA gene. In short, there is much that can be learned about evolution by studying the topology of its timeline.

In this paper, we ask what can be learned from a detailed study of the topology of modern phylogenetic trees, constructed with the benefit of large-scale genomic datasets. In fact, converging lines of evidence strongly suggest the presence of scale invariance in both topological and metric aspects of trees (4–9). These works provide a quantitative and data-rich analysis that is in the same spirit as hypotheses made earlier on the basis of fossil extinction and taxonomic evidence (10, 11) and discussed theoretically with reference to critical models of evolution (12, 13) (see, e.g., refs. 14–16 for detailed reviews of the literature prior to the genomics era). The new element of the recent analyses (4–9) is that they are based on molecular phylogeny, and so contain far more information about the evolutionary process than can be obtained from patterns of extinction.

Nodes of a phylogenetic tree stand for extant organisms (outer leaves) and their hypothesized ancestors (inner leaves). They can be analyzed to quantify the topology of tree branching (4–7). There are several metrics developed in the literature to characterize the topology of phylogenetic trees (4–8). Here, we shall focus on the metrics defined in ref. 5, in which the basic idea is to count the number of subtaxa that diversify from a given node i. We call the first quantity $A(i)$ and define it recursively as the number of nodes of the subtree S_i rooted at node i, including in the counting the node i itself. The second quantity, $C(i)$, is the cumulative sum of $A(i)$ over the subtree, including the

Significance

Phylogenetic trees describe both the evolutionary process and community diversity. Recent work, especially on bacterial sequences, has established that, despite their apparent complexity, they exhibit two unexplained broad structural features which are consistent across evolutionary time. The first is that phylogenetic trees exhibit scale-invariant topology. The second is that the backbones of phylogenetic trees exhibit bursts of diversification on all timescales. Here, we present a coarse-grained model of niche construction coupled to simple models of speciation that recapitulates both the scale-invariant topology and the bursty pattern of diversification in time. These results show, in principle, how dynamical scaling laws of phylogenetic trees on long timescales may emerge from generic aspects of the interplay between ecological and evolutionary processes.

Author contributions: C.X. and N.G. designed research; C.X., Z.L., and N.G. performed research; and C.X., Z.L., and N.G. wrote the paper.

Reviewers: M.W.F., Stanford University; J.K., University of Cologne; and K.S., University of Copenhagen.

Data deposition: All data associated with the manuscript are accessible publicly via GitHub (https://github.com/zhiru-liu/niche-inheritance-trees).

[1]To whom correspondence may be addressed. Email: nigel@illinois.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1915088117/-/DCSupplemental.

First published March 24, 2020.

value at the node i. Both C and A depend on the node i, and, therefore, we can measure how C varies with A over a particular phylogenetic tree. The resulting functional dependence is a strong indicator of the large-scale topology of the tree. For example, if the tree is completely symmetric or balanced, with equal branching into two nodes from every node, as illustrated in Fig. 1*A*, then it can be shown from the definition of C and A that $C = 1 + (A+1)([\ln(A+1)/\ln 2] - 1)$. On the other hand, a maximally unbalanced tree, as shown in Fig. 1*B*, would have each node branching into two: one node that simply persists to the edge of the tree without subsequent branching, and a second node that branches just like the parent, into a nonbranching and a branching node. In this case, $C = A^2/4 + A - 1/4$. These exact results have very distinct asymptotic scaling behavior for large A: $C \sim A \ln A$ (balanced) and $C \sim A^2$ (maximally unbalanced), showing explicitly that topology can be reflected in a scaling law for a phylogenetic tree.

How do real data scale? Remarkably, it is found that, over three orders of magnitude of A, there is a power-law scaling of $C(A) \sim A^{\eta}$, with the exponent $\eta = 1.44 \pm 0.01$ (5–7). The question we are concerned with here is the explanation for this topological scaling law. One possibility is that the result is an artifact influenced by bias due to such factors as uneven speciation rates, choice of taxa, and choice of outgroups for the trees (17). However, these uncertainties do not explain how these effects could lead to power-law behavior of tree topology, especially since there has been more than one independent analysis performed. Moreover, the same metrics have been applied to other trees and networks where these biases are not present, ranging from river-drainage basins to protein networks (18–20), and, once again, a variety of power-law scalings is observed. We take the perspective below that the power-law scaling is indeed real.

There have been many theoretical attempts to model the evolution of phylogenetic trees. See refs. 21–24 for comprehensive reviews. The equal-rates-Markov (ERM) model, first developed by Yule in 1924 (25) and later expanded in the literature (26–28), usually serves as a null hypothesis for the evolutionary process of the tree. The ERM assumes that all extant species on the tree have the same speciation rate. The resultant tree is less unbalanced than the observed ones, and, statistically, it should behave on large scales as a balanced tree with the asymptotic scaling $C \sim A \ln A$. The proportional-to-distinguishable-arrangements (PDA) model assumes that, for a given tree size, all tree topologies are equally likely to appear. The tree, then, is a result of recursively sampling the topology for the subtrees. The original PDA model (29–31) did not give any rules of the growth of the tree, but corresponding evolutionary processes were developed later (32, 33). Again, in the absence of an explicit symmetry-breaking bias, the balanced asymptotic scaling is expected in such models.

Connecting the nodes, edges of a phylogenetic tree represent the evolution time from a parent node to its daughter(s). They can be analyzed by measuring the edge-length abundance distribution (EAD) (9). The EAD is calculated from the number of tips or leaf nodes k that descend from a given internal node i. For a given node i with k tips, $S_i(k)$ is the length of the edge to its immediate ancestor. We define $S(k) \equiv \sum_i S_i(k)$. The result is that, over about three decades of k, $S(k) \sim k^{-\alpha}$ with an exponent in the range $1.3 < \alpha < 1.7$ (9).

As reviewed in ref. 9, although the Yule process (25) and the Kingman coalescent (34) produced power-law EAD, they do not generate exponents that are measured with actual phylogenetic trees. Extended Kingman coalescent with time-varying rate (9) and Λ-coalescent (35, 36) can produce the observed value with a tuning parameter. Yet, the specific biological reason of the parameter choice is not clear.

One way to obtain anomalous power-law scaling for tree topology is to directly include power-law aging behavior into the rules for the generation of trees (4, 37–39). For example, by requiring that branching probabilities are a particular power-law function of the branch age (38), it is possible to obtain both logarithmic and power-law scalings for $C(A)$. However, this approach does not provide a mechanistic interpretation for the scaling laws put into the model, let along those that emerge. Nevertheless, the result does show that the observed power-law scaling can, in principle, arise from a long-term memory introduced into the branching process as an interaction that is nonlocal in time.

The structure of phylogenetic trees can be interpreted as arising from the interplay between evolution and ecosystem dynamics. We view the nodes as revealing information primarily about ecological processes that result in the fixation of beneficial mutations, whereas the edges reveal information primarily about evolutionary processes, because their lengths reflect the numbers of DNA mutations. The presence of a power-law aging or long-term memory implies a breakdown of the separation of scales implicit in the identification of nodes with ecological interactions and edges with evolutionary dynamics. This standard identification would be valid if the ratio R of ecological timescales to evolutionary timescales could be assumed to be zero. However, even though $R \ll 1$, the limit $R \to 0$ of observables, such as the

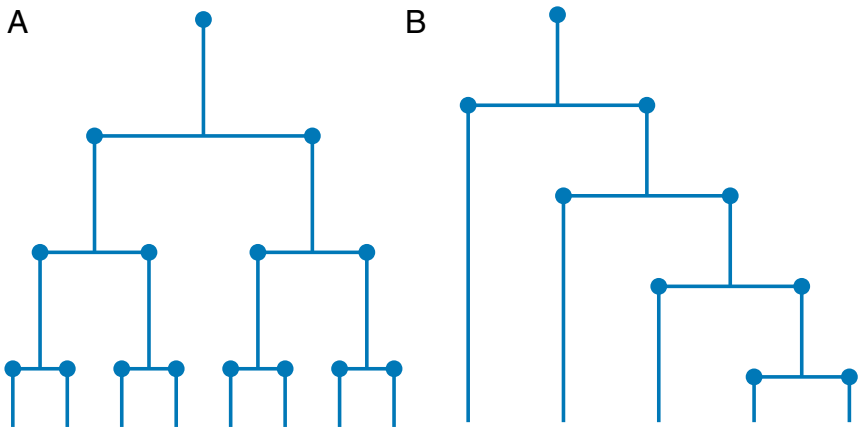

Fig. 1. (*A*) A balanced tree. All nodes have exactly two descendants. The topological relation $C(A) \sim A \ln A$ at large A. (*B*) A maximally unbalanced tree. For any node, only one of the two children continues branching. $C \sim A^2$ at large A. Actual phylogenetic trees have topology and scaling behavior in between the two extreme cases, as studied in ref. 5.

structure of a phylogenetic tree, may not exist due to a singularity; for example, $C(A, R)$ may vary as $R^g F(A)$, where g is a nonzero exponent. Depending on the sign of g, $C(A)$ will either vanish or diverge as the limit is taken, showing that the limit is singular and the timescale separation does not strictly exist. Such problems are common in continuum mechanics, fluid dynamics, and phase transitions (40, 41), and the failure of this limit to exist is responsible for the anomalous scaling laws in phase transitions that naively seem to violate dimensional analysis. In fact, dimensional analysis is not violated, of course, due to the surprising way in which the lattice scale (on the order of angstroms) influences the correlations, even on the scale of the correlation length itself, which may be many orders of magnitude greater than the lattice scale (40). This so-called "scale-interference" in space apparently has an analogue in the present problem, but it is a scale-interference in time, something well-documented in other dynamical system problems (40, 42). The breakdown of the timescale separation suggested here implies that there is feedback between the ecological and evolutionary processes. This feedback is sometimes known in broad terms as "niche construction," and we will see below that the role of niche construction in evolution is observable, even at large timescales. In short, the goal of this paper is to test whether or not such scale-interference does occur, using as a model the way in which it is detected in phase-transition phenomena.

Niche construction (43–51) is a term that describes the fact that organisms modify the environment and, thus, create new ecological niches; in turn, these niches affect the evolutionary trajectory of other organisms that share the environment (44). The resulting dynamics is a coevolution of the coupled dynamical variables for the organisms (52–55) or their genomes (45), as well as the environment itself. This coupled dynamics contains two-way feedbacks between the organisms and the environment, which are local in time. However, phylogenetic trees follow only the dynamics of the organisms themselves. The effective theory for the organismal degrees of freedom can be obtained conceptually by integrating out the environmental variables (e.g., using functional integration), and the resulting description would then contain interactions that are nonlocal in time, leading to an effective long-term memory in the branching process.

We will see that niche construction indeed introduces long-lived memory into the evolutionary process, and even a very simple caricature of niche construction over evolutionary time can capture the power-law scaling of $C(A)$, with an exponent that is close to the one observed empirically. Moreover, the same model also leads to a power-law EAD, with an exponent within range of the empirical estimates. The analyses we perform in this article mirror the physics of anomalous scaling exponents in critical phenomena, and we establish our main point through the techniques of cross-over scaling (40).

The niche of a species generally refers to its role or function in an ecosystem and can be thought of as the "variables by which species in a given community are adaptively related" and which control the species population response to each other and their environment (56). The habitat occupied by a given species is distinct from the niche, in this definition, and the two together comprise the ecotope (56). The ecotope involves the environmental factors that the species relies on, including the geographic configuration, the climate, etc., and the interactions with other species in the same ecosystem, represented by the species' position in the food web and their dynamical history. The phenomenon that organisms modify the environment and thus create new niches is termed "niche construction" (44). In contrast to natural-selection theory, which treats the environment as a static stage on which population dynamics and genetics occur, niche construction theory emphasizes the modification of the environment by the organisms as an explicit process and a key factor in evolution (43–48), although there remain controversies (49–51). Specifically, organisms can shape the environment they live in, change the selection pressure, and, as a result, reroute their own evolutionary path. A similar feedback of organisms on their environment is often referred to in ecology as ecosystem engineering (57–60), which is on a shorter timescale than that of niche construction. Even though it can be argued that niche construction really means ecotype construction, here, we will use the term niche, as is frequently done.

Ecological models, such as the MacArthur–Levins model (61), typically treat a niche as an abstract one-dimensional space which the organisms inhabit (although multidimensional niche models also exist; see, for example, refs. 56 and 62). We will regard a niche as the total available growth space or evolutionary degrees of freedom of the organism. We will call n the available niche value in the analysis below. When we say an organism has a large niche value, we mean that it has a large number of possible ways to adapt to the environment and to eventually survive or to reach genome-type fixation. On the other hand, an organism with a small niche value is very discriminating with regard to its environment, and this will affect its ability to be resilient within a wide range of environmental fluctuations. Our usage of niche means that, following a speciation event, the daughter species have a similar niche value, but with some fluctuation. The niche value can change either by organismal influence on the environment or through the evolution of mutations that enable key innovations to arise.

Our work is, of course, not the first to attempt to model niche construction, but the ingredient here is the invoking of critical scaling theory to explain the observed topological scaling laws of the large-scale evolutionary dynamics. Of earlier work in this area, we specifically draw attention to applied population-dynamics models (52–55) and population-genetics models (45) that study the effect of niche construction or ecosystem engineering on organism populations and evolution.

This article is organized as follows. First, we present a minimal model of the large-scale effects of niche construction, which we call the Niche Inheritance Model. In this model, the descendant species inherit the parent's niche with fluctuations due to niche construction and evolution. The model is a caricature of the most significant ecological interactions that influence a phylogenetic tree in our assessment, but, based on a huge body of work on scaling laws, we anticipate that such a minimal model will yield nontrivial predictions that can be in agreement with experimental data (40). Next, we show that an apparent power-law regime develops when strong niche construction (destruction) leads nodes to be deactivated due to a lack of niche. The scaling laws are revealed through data-collapse scaling. Finally, we end with some discussion about the use of minimal models in evolutionary ecology at large timescales.

Results

Niche Inheritance Model. We assign each species node three attributes: the amount of available niche n, the speciation rate r, and the extinction probability e. The tree-generation algorithm is as follows. Let the parent node be represented by its parameters (n_0, r_0, e_0). We first calculate the time interval until the first speciation event, assuming a Poisson process with a rate r_0. Then, we forward time to the speciation event. We let the parent diversify into two children (n_1, r_1, e_1) and (n_2, r_2, e_2). We treat the branching to be binary, because a multifurcation can be viewed as a coarse-grained bifurcation. The niche sizes n_1 and n_2 are inherited from the parent with fluctuations due to construction/destruction, as expressed below:

$$n_1 = n_0 + \Delta n_1, \quad \textbf{[1a]}$$

$$n_2 = n_0 + \Delta n_2. \quad \textbf{[1b]}$$

The fluctuations Δn_i, $i = 1, 2$, are assumed to be generated by the following distribution:

$$\frac{\Delta n_i}{n_0} \sim \mathcal{N}(\mu_n, \sigma_n^2), \quad [2]$$

where $\mathcal{N}(\mu_n, \sigma_n^2)$ stands for a normal distribution with mean $\mu_n \equiv 0$ and variance σ_n^2. For each child node, we calculate r_i and e_i according to certain mathematical rules, such as Eqs. **3** and **4**, which will be discussed in the next paragraphs. Then, we test whether the node goes extinct or remains to bifurcate later. The test is done by drawing a uniformly distributed random number in $[0, 1]$ and comparing it with the extinction probability e_i. The child goes extinct and is removed from the tree if the random number is smaller than e_i. All inferred nodes and branches dependent on the extinct child are also pruned away. The pruning is necessary to make the simulation result directly comparable to actual phylogenetic trees. In real trees, only nodes that are on the same lineage as leaf nodes can be hypothetically inferred, and those who have gone extinct with no descendants are not visible. Hence, in the simulation, when a node fails to pass the extinction test, we remove it, together with its ancestors.

The speciation rate r is treated as an increasing function of niche to reflect the fact that the more available niche space there is, the more likely it is for a speciation event to be successful. Specifically, with the interpretation of available niche n as the total available growth space or evolutionary degrees of freedom, when an organism has a large niche value, it has a large number of possible ways to adapt to the environment and to eventually survive. In this way, the speciation rate r naturally positively correlates with the niche n. To capture the first-order feature and to develop a minimal description, we set the relation between r and n to be linear, as shown below.

$$r(n) = \begin{cases} n, & n \geq 0, \\ r_\epsilon, & n < 0. \end{cases} \quad [3]$$

Here, r_ϵ is the background speciation rate when the available niche n is negative. With this relation, we also enforce $r_0 = n_0$ for the root node.

The extinction probability e implicitly incorporates all ecological interactions that can lead to extinction after a species emerges. Despite the fact that there are multiple factors driving species extinction, there is no well-accepted way to quantify the strength of each factor or to braid them together. In this model, we assume that the extinction probability e increases with the speciation rate r, based on the reasoning that a large speciation rate results in a big group of competitors with similar niche and, thus, reduces the survival probability of individual species. In our effort to build a minimal model, we assume a simple functional form of $e(r)$ to capture the positive correlation between e and r, as shown below. We also use the following function of $e(r)$ to effectively limit the bifurcation rate of the tree:

$$e(r) = \frac{r}{r + R_0}. \quad [4]$$

The rate r and probability e represent all ecological interactions among species in this very simplified minimal model. Note that the main purpose of extinction in this model is to limit the growth of the tree. Without it, the successful speciation rate would quickly diverge, and the number of nodes would diverge exponentially in time. Our form of the extinction probability is effectively a cap on the speciation rate and may be thought of as representing competition between coexisting organisms. This form of the extinction probability is broadly consistent with the finding that longevity is unrelated to order size and is, in some sense, heritable and, thus, dependent on environmental factors (63).

In a numerical simulation, we start with a root node and evolve the tree based on the above rules until it reaches a certain size. Then, we compute A and C using the following definitions. For an arbitrary node i on the tree, let S_i be the subtree rooted at node i. Define $A(i)$ as the size, or number of nodes, of S_i, and $C(i)$ as the cumulative size of S_i, $C(i) \equiv \sum_{j \in S_i} A(j)$. We calculate the (A, C) pair for every node and, thus, obtain the relation $C(A)$.

Existence of the Absorbing Boundary. In the above framework, there exists a boundary case of $r_\epsilon = 0$, which means that nodes with negative niche values will never bifurcate. In actual evolution, we observe species that seem not to be changing phenotypically while their relatives actively diversify—for example, the "living fossil" species coelacanth. Therefore, this boundary case is biologically meaningful. We use it as the starting point of our analysis.

Imagine the left node starts by chance with a larger n and, thus, a higher r than the right one. If $r_\epsilon \neq 0$ and all succeeding nodes are able to branch, then the right node, by fluctuation, will eventually gain a descendant with high r, and the left node will gain a descendant with low r. The two subtrees, in general, undergo the same random process and are symmetric. Therefore, on a long timescale, the entire tree is balanced due to this "catch-up" effect. However, if $r_\epsilon = 0$, once a node gets a negative niche, it is deactivated and will not be able to contribute any descendants in the evolutionary process. This eliminates the possibility of catching up the growth between the left and right subtrees. Therefore, $r_\epsilon = 0$ drives the asymmetry of the tree and leads it to be unbalanced. We will refer to the case of $r_\epsilon = 0$ as the absorbing boundary, since it effectively removes bifurcating species from the tree.

Effect of Niche Construction Strength. While the absorbing boundary induces imbalance in the tree, the frequency of nodes hitting the boundary also matters. Based on Eq. **2**, we see that σ_n tunes the probability for the child node to have a negative niche value. Therefore, it determines how often nodes reach the absorbing boundary. When $\sigma_n = 0$, the niche does not fluctuate, and all nodes have the same value of niche as well as the same speciation rate. The Niche Inheritance Model is thus equivalent to the Yule process (25). We expect the resultant tree to be balanced. When σ_n is large, however, the access to the absorbing boundary is frequent. There will be many nodes turning inactive and many branches being terminated during the evolution. The effect of imbalance exerted by the absorbing boundary now becomes visible.

We demonstrate in Fig. 2 $C(A)$ relations for different niche construction strengths. We next point out the important issue of undersampling. If two nodes have subtrees of the same size A but different topologies, then they will very likely have different values of C (except if one tree can be transformed to the other by mirroring the left and right branches). For a given size, there can be many subtrees of distinct topologies. Therefore, we usually have multiple C values associated with the same A, especially when A is not too small. However, if A is large and comparable to the total size of the tree, then there may only be a few subtrees to sample from, and, thus, there will be only a few values of C over which to sample. In the extreme case, when A is equal to the size of the phylogenetic tree, there is only one topology present, that of the tree itself, and C is single valued. The (C, A) pair now is associated with the root. This means that when we look for scaling laws of phylogenetic trees, we must generate trees that are much larger than the range of A where we measure scaling, in order that there are enough subtrees to generate $C(A)$ with good statistical accuracy. Correspondingly, at large values

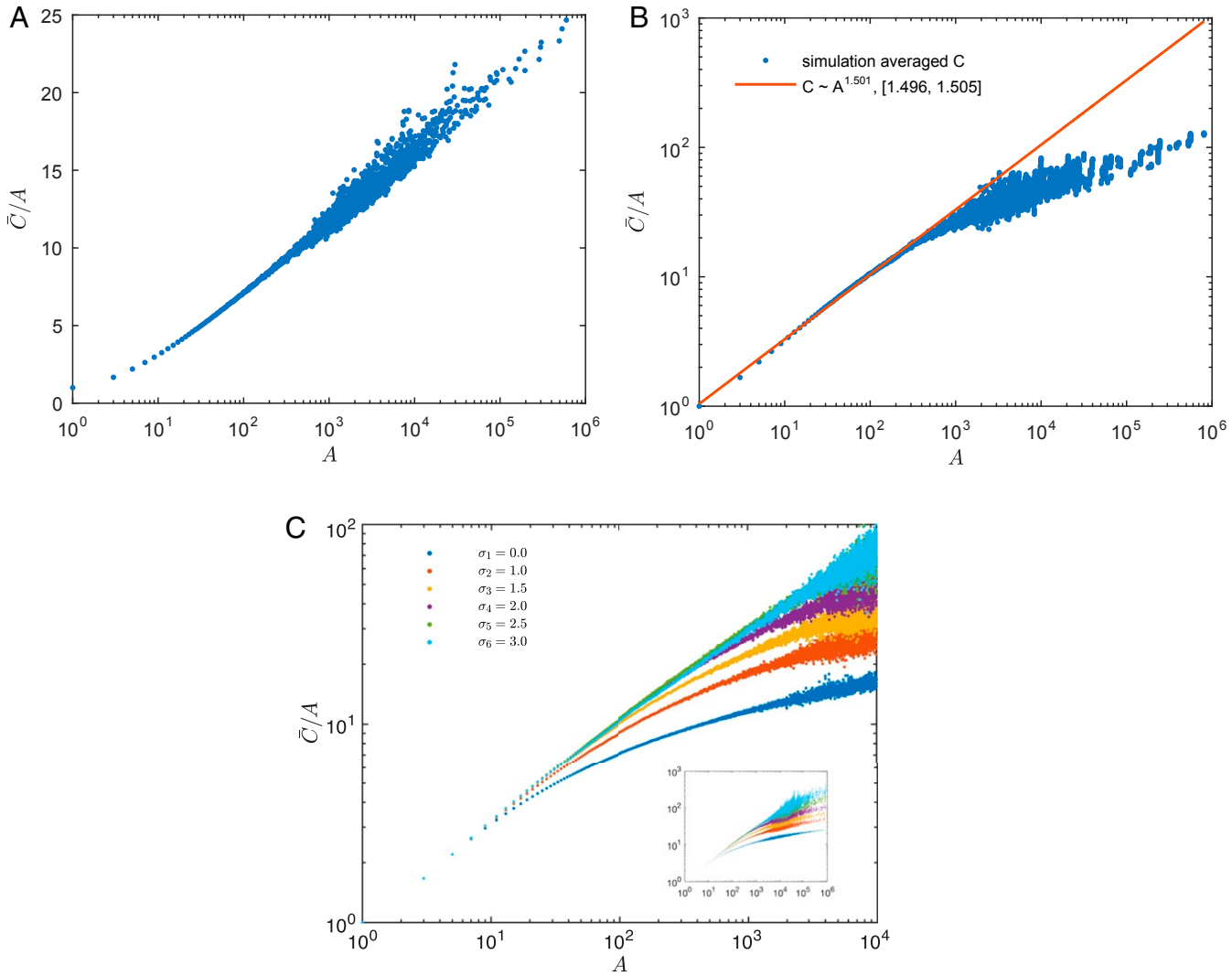

Fig. 2. (*A*) Averaged $C(A)$ calculated for a typical tree generated by the Niche Inheritance Model, with $\sigma_n = 0$. The dots scatter along a straight line in the linear-logarithmic scale, indicating $\bar{C}/A \sim \ln A$. (*B*) Averaged typical $C(A)$ calculated with $\sigma_n = 2$. The scale is double logarithmic. Fitting the well-averaged region, $A < 200$, to a power function $\bar{C} \sim A^{\eta}$ gives an exponent of $\eta = 1.501$, with the 95% CI being [1.496, 1.505]. The red line stands for the fitted function. (*C*) Dependence of averaged $C(A)$ on σ_n, with $r_\epsilon = 0$. As σ_n increases, the apparent power-law region of $\bar{C}(A)$ also stretches. Since subtrees with $A > 10^4$ are undersampled, those data points are not shown in the main plot. The full data are shown in *C*, *Inset*. All of the following $C(A)$ plots are handled in a similar fashion. Other parameters for all sets of simulations are $r_\epsilon = 0$, $\mu_n = 0$, $R_0 = 10$, and $n_0 = 1$ for the root node.

of A, the $C(A)$ data will be undersampled and will show large statistical fluctuations that are artifacts of undersampling.

In Fig. 2, we average the C values corresponding to the same A and plot the resulting quantity $\bar{C}$ vs. A. At small A, there are many samples of subtrees, and $\bar{C}$ reasonably represents the expected value. This is indicated by the thin and smooth region in the $\bar{C}$–A graph. However, at large A, there are few subtrees present, and the tree topology is thus heavily undersampled for the reasons explained above. $\bar{C}$ then does not reflect the correct value corresponding to an asymptotically large tree. This is illustrated as the broad scattered region in the $\bar{C}$-A graph.

For zero niche construction, $\sigma_n = 0$, the model reduces to a Yule process, and we expect the tree to be balanced with a $C(A) \sim A \ln A$ asymptotic behavior. This is verified in Fig. 2*A*. Notice that the scale is linear-logarithmic, and the $A \ln A$ behavior is illustrated by the dots scattering along a straight line. When there is a strong niche construction effect or a large σ_n, we expect the tree to be unbalanced, with $C(A)$ deviating from the balanced scaling. This is demonstrated in Fig. 2*B*. Instead of $A \ln A$, in the range comparable with the observed data in ref. 5, $A < 200$, $C(A)$ falls roughly along a straight line in the double-logarithmic scale, indicating a power-law behavior, $C(A) \sim A^{\eta}$. A fitting to the power-law function gives the exponent of $\eta \approx 1.50$.

Fig. 2*C* shows the averaged $\bar{C}(A)$ curves at different values of σ_n. Judging from the smooth regions at small A of $\bar{C}$-A curves, as σ_n increases from 0 to 3, $\bar{C}(A)$ transitions from $A \ln A$ to an apparent power-law behavior. The power-law regime grows in range as σ_n increases. We do not have a strong conclusion at the large A end due to the undersampling issue.

So far, we have shown that niche construction together with the absorbing boundary induces a power-law scaling regime in the $C(A)$ relation. In the next sections, we will explore the origin of the scaling.

Singularity Induced by the Absorbing Boundary. We attempted to describe the scaling behavior using a simplified mean field

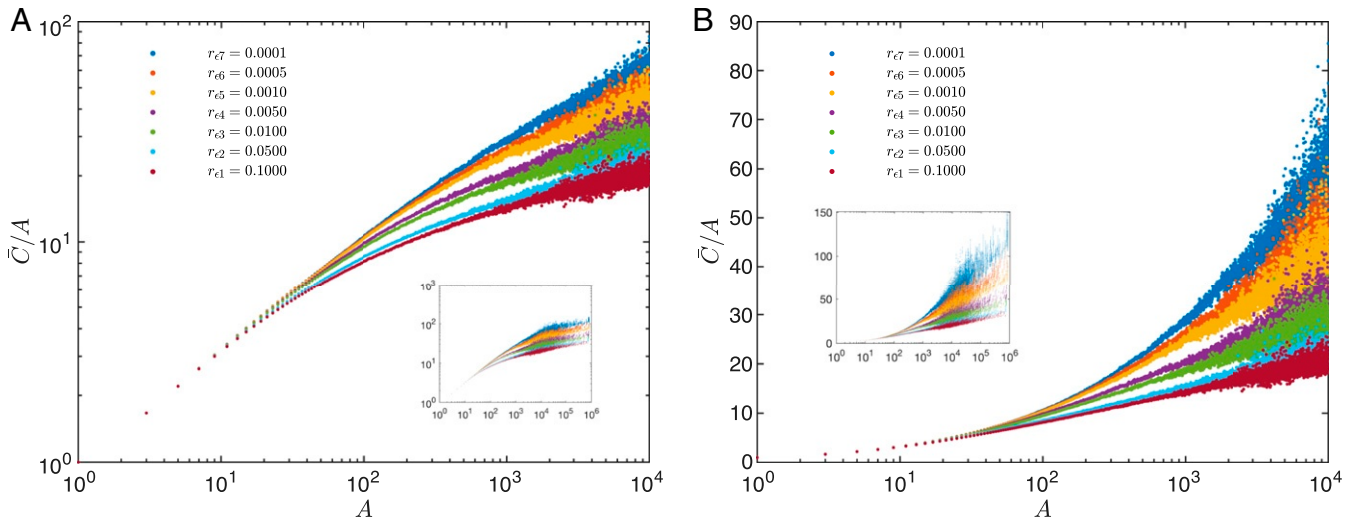

Fig. 3. (*A*) Dependence of averaged $C(A)$ on r_ϵ, with $\sigma_n = 2.5$. As r_ϵ approaches zero, $\bar{C}(A)$ expands the power-law range. Other parameters are $\mu_n = 0$, $R_0 = 10$, and $n_0 = 1$ for the root node. (*B*) The same data as in *A* plotted in the linear-logarithmic scale. The scaling turns $\bar{C}(A) \sim A \ln A$, as r_ϵ becomes large. *Insets* in *A* and *B* show the full data range, respectively.

that ignores the fluctuations of the tree topology due to the random-species birth and death processes (see the nonextinctive mean-field model in *SI Appendix*). Although the mean-field calculation explains the exact scalings for the two extrema of balanced and maximally unbalanced trees, it fails to recapitulate the scaling laws that we observe in the numerical simulations. However, it is well known from the theory of critical phenomena that nontrivial power-law scaling arises from singularities in limit processes (40) that cannot be captured by mean-field theory. Thus, we now focus on the singularity induced by the absorbing boundary.

The imbalance induced by a large niche construction effect is crucially related to the condition $r(n) = r_\epsilon = 0$ for $n < 0$, which means that nodes stop branching when $n < 0$. It can be relaxed to a positive r_ϵ, if the tree has a finite growing time T. When r_ϵ is nonzero but still small enough such that $1/r_\epsilon \gg T$, then very few nodes with negative niches will be able to complete the bifurcation before the termination of the tree growth. So, effectively, a small r_ϵ acts as an absorbing boundary as well. As r_ϵ increases, the deactivation effect due to a finite T only acts on nodes near the tips. The symmetry between the left and right branches is gradually restored. Therefore, at large r_ϵ, the tree becomes balanced.

In the above argument, we have implied that nodes are able to reach the $n < 0$ region, so that r_ϵ can play a role in the evolution. This condition applies in the presence of a strong niche construction effect, and, in the calculations reported below, we have taken $\sigma_n = 2.5$, so that the influence of the absorbing boundary is clearly visible in the range of A that we can easily simulate with good statistics.

In Fig. 3, we show the dependence of $\bar{C}(A)$ on r_ϵ for trees terminated at a finite size $A \approx 10^6$. When $r_\epsilon = 0$, we observe the same apparent power-law regime of $\bar{C}(A)$ in the well-averaged range of A as in Fig. 2*B*. This is demonstrated as the segment of straight line under the double-logarithmic scale in Fig. 3*A*. As r_ϵ increases, the apparent power-law region of $\bar{C}(A)$ reduces in range. Eventually, a behavior of $A \ln A$ becomes significant in the entire range of A at a large r_ϵ, as illustrated by the straight line with $r_\epsilon = 0.1$ under the linear-logarithmic scale in Fig. 3*B*.

Critical Scaling at the Absorbing Boundary. Although the behavior of $\bar{C}(A)$ at large A in Fig. 3 is not well represented due to undersampling, we conjecture that, for a nonzero r_ϵ, $C(A)$ consists of two distinct asymptotic limits: at small A, $C(A) \sim A^\eta$; and at large A, $C(A) \sim A \ln A$. The cross-over happens around the transition point $A = A_T$. In fact, in Fig. 3, the curve corresponding to $r_\epsilon = 0.01$ has both a significant power-law region and a smooth $A \ln A$ region, before the issue of undersampling smears the data.

With this conjecture, we observe that A_T divides $C(A)$ into two regimes and that the transition point A_T increases as r_ϵ approaches 0. We propose that the dependence of A_T on r_ϵ is critical. Then, in the terminology of critical phenomena (40), there exists a so-called cross-over scaling function

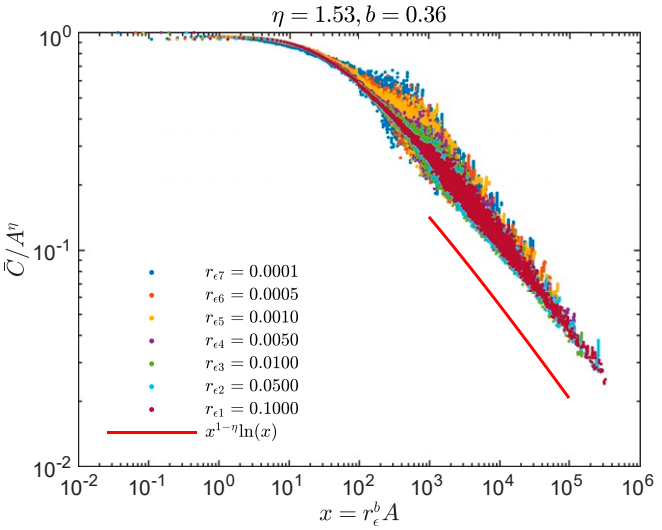

Fig. 4. Critical scaling of $\bar{C}(A)$ as r_ϵ decreases, indicated by the data collapse. By tuning η and b, we reach a data collapse of the eight $\bar{C}(A)$ datasets corresponding to eight values of r_ϵ ranging from 0 to 0.1 across four orders of magnitude. $b = 0.36$ and $\eta = 1.53$ gives the best result. The tail matches the expected $x^{1-\eta} \ln x$ behavior, which is indicated by the straight reference line in the double-logarithmic scale. Other parameters for all datasets are $\sigma_n = 2.5$, $\mu_n = 0$, $R_0 = 10$, and $n_0 = 1$ for the root node.

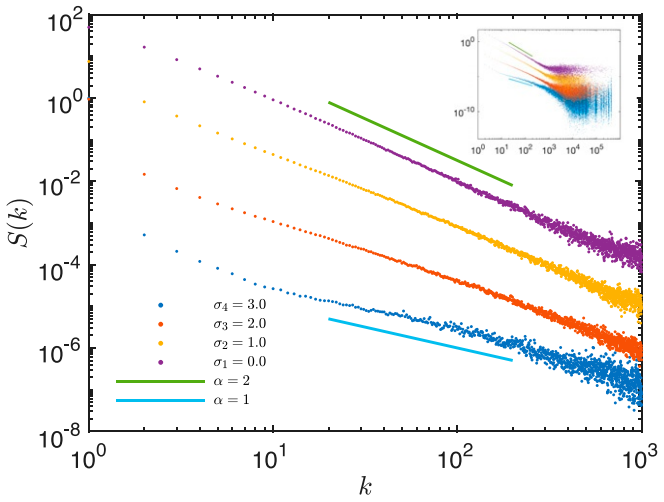

Fig. 5. Dependence of EAD on σ_n, with $r_\epsilon = 0$. Data are shifted for clarity. Increasing σ_n changes the power-law exponent approximately from $\alpha = 2$ to $\alpha = 1$. Since large subtrees are not well sampled, data beyond $k > 10^3$ become noisy and, thus, are not informative. *Inset* shows the EADs plotted with full data. Other parameters used are $\mu_n = 0$, $R_0 = 10$, and $n_0 = 1$.

$F(x)$, defined as

$$x = r_\epsilon^b A, \qquad [5]$$

such that

$$C(A, r_\epsilon) = A^\eta F(x). \qquad [6]$$

The functional form of $F(x)$ should accommodate the fact that

$$C(A) \sim \begin{cases} A^\eta, & \text{small } A, \\ A \ln A, & \text{large } A. \end{cases} \qquad [7]$$

Here, we have included the $\ln A$ correction at large A. Following the standard procedure for finite-size scaling at the upper critical dimension, where there are logarithmic corrections to scaling (64), we require that $F(x)$ has the following asymptotic behaviors:

$$F(x) \to \begin{cases} \text{const}, & \text{small } A \text{ and } x \to 0, \\ x^{1-\eta} \ln x, & \text{large } A \text{ and } x \to +\infty. \end{cases} \qquad [8]$$

If there is critical scaling, and the function $F(x)$ exists, then different datasets corresponding to different r_ϵ values should collapse onto the same curve, when plotted as C/A^η vs. $x = r_\epsilon^b A$. Indeed, in Fig. 4, we show the data collapse obtained by tuning exponents b and η. For the presented eight datasets, $b = 0.36$ and $\eta = 1.53$ gives the best data collapse.

The data collapse indicates a critical behavior of $C(A)$ as r_ϵ approaches 0. Notice that the value of η found from the data collapse is slightly different from the one obtained via fitting in Fig. 2*B*. From the function $F(x)$, obtained in the data collapse, we can read off the value x_T at which $F(x)$ crosses over from a constant to $x^{1-\eta}$. In Fig. 4, this value is $x_T \approx 3.85$. Then, for an arbitrary r_ϵ, we can calculate A_T as follows:

$$A_T = x_T r_\epsilon^{-b}. \qquad [9]$$

As $r_\epsilon \to 0$, $A_T \to +\infty$ and the power-law scaling of $C(A) \sim A^\eta$ expands to the entire range of A.

In the above two sections, we have considered a soft absorbing boundary of finite and small r_ϵ. By studying the scaling behavior as r_ϵ goes from finite to 0, we show that the boundary effect induces critical dynamics associated with a phase transition and predict a measurable transition point.

Power Law in the EAD. In this section, we show that the niche construction model also reproduces another characteristic of phylogenetic trees: the power law in the EAD, first discovered by O'Dwyer et al. (9). For convenience, we briefly revisit the definition of this distribution. The distribution is denoted by a function $S(k)$, where k is the clade size, the number of leaf nodes a subtree has. The edge length of a node is the time interval between its birth and speciation. $S(k)$ is then the sum of all edge lengths of nodes whose descendant trees have clade size k. It was found that the EAD of phylogenetic trees follows a power-law behavior, $S(k) \sim k^{-\alpha}$, where α is estimated to be between 1.3 and 1.7 (9).

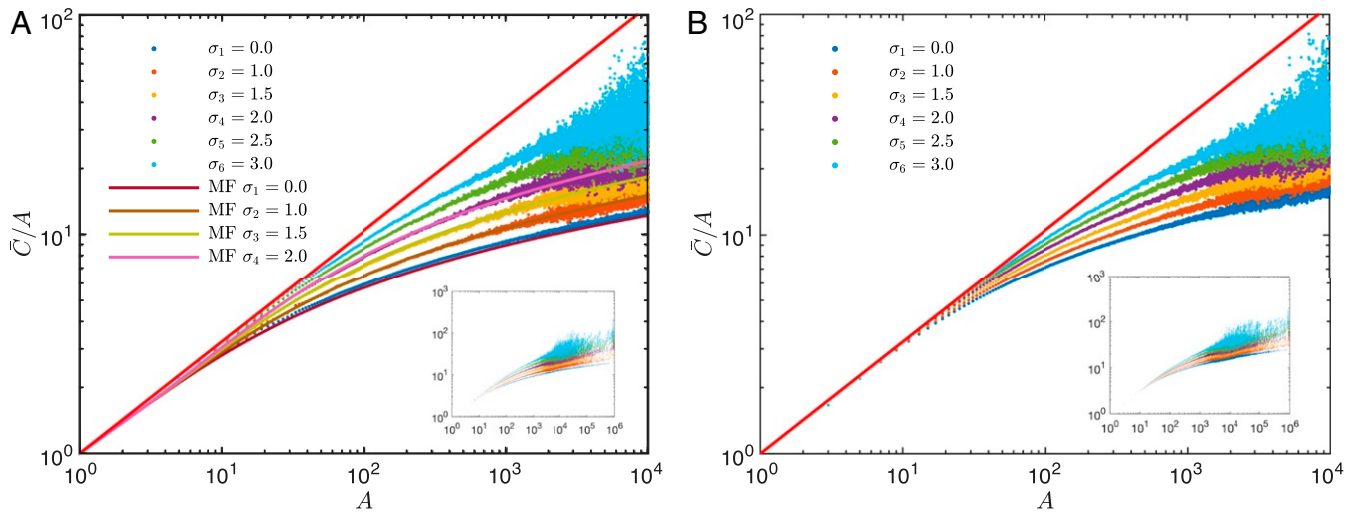

Fig. 6. (*A*) Averaged $C(A)$ for the modified model with a constant waiting time. Theoretical predictions from mean field analysis in *SI Appendix* for small niche construction strengths are also plotted. None shows power-law behavior. (*B*) Averaged $C(A)$ for the modified model with a constant bifurcation rate. Mean-field analysis no longer applies to this case, and power-law behavior is still not recovered. For both plots, other parameters used are $r_\epsilon = 0$, $\mu_n = 0$, $R_0 = 100$, and $n_0 = 1$. The reference line is the power-law function $C = A^{1.51}$. The simulated tree has 10^6 nodes, and the data are averaged over 10 repetitions. The *Insets* in *A* and *B* show the full data range, respectively.

In Fig. 5, EADs of the niche construction model are presented. Each EAD has been scaled differently to reduce overlap in the plot. In the double-logarithmic scale, $S(k)$ falls roughly along a straight line for $k < 1{,}000$, indicating a power-law behavior. The scaling range is comparable with the observation of real phylogenetic trees (9). The data points on the right side are more scattered because large clades are sampled insufficiently, as in the case of $C(A)$. As shown by the two reference lines, increasing the niche construction strength changes the exponent of the power law. When $\sigma_n = 0$, the model reduces to the Yule process, whose EAD follows

$$S_{Yule} \sim 1/k(k-1), \qquad [10]$$

which corresponds to the $\alpha = 2$ power law for large k. These results show that the niche construction model exhibits scaling in the EAD, but the scaling exponent depends on the niche construction strength.

The Necessity of Eco-Evolutionary Feedback in a Minimal Model. In this section, we will connect the scaling behavior of $C(A)$ with a central element of our model: the coupling between the speciation rate and the niche. Mathematically formulated in Eq. **3**, this coupling represents perhaps the simplest nontrivial feedback between the ecological variable, niche values, and the evolutionary variable, edge lengths. To explore what are the essential ingredients of a minimal model for phylogenetic tree structure, we will demonstrate the effects of modifying our model. We will show that, without this feedback, it is not possible to recapitulate scale invariance and the anomalous scaling laws, and so this is an essential part of a minimal model for phylogenetic tree structure.

To begin with, we recall that the edge length of a particular node is determined by its speciation rate r. More specifically, we required the edge length, or time till speciation, to follow an exponential distribution with parameter r. We could modify this element of the model by requiring all edge lengths to be equal to a constant number instead. We will set r_ϵ to be zero, so that nodes with negative niche values will still be turned inactive. With constant edge lengths, only inactive nodes will contribute to the imbalance of a tree. Therefore, this modification has isolated the node deactivation as the only mechanism affecting the topological structure. As shown in Fig. 6*A*, none of the curves have a noticeable range of power-law scaling. It's worth noting that the finite and constant edge lengths effectively satisfy the infinite time assumption in the mean-field calculation (*SI Appendix*). As a result, when extinction events are not frequent, the analytical result in *SI Appendix* (*SI Appendix*, Eq. **S11**) should apply to this modified model. Indeed, for small fluctuation strengths, the analytical curves agree well with the simulation in Fig. 6*A*. This plot demonstrates that node deactivation alone is not sufficient to produce the power-law behavior in $C(A)$.

Eliminating variability of edge lengths might be too drastic a change. Therefore, in the next modification, we allow the edge length to follow an exponential distribution with a constant rate. In contrast to the original model, this modified version does not retain the coupling between the niche value and the speciation rate (Eq. **3**). Fig. 6*B* shows that, even with variable edge lengths, $C(A)$ still loses the power-law behavior. As a direct comparison, in Fig. 2*C*, data for $\sigma_n > 2$ clearly display a power-law behavior over nearly three decades.

The above two modifications illustrate that edge lengths need to be not only varying, but also coupled to niche values in order to generate realistic topological structures. We could understand the effect of edge lengths by considering a node with a very small niche value and, hence, a very small speciation rate. Such a node most likely has a very long waiting time before speciation, but because the simulation time in real life is finite, the node cannot speciate. On the other hand, nodes with large niche values will speciate more often and have more child nodes, thus causing the imbalance in the tree. This distinction between large and small niche values is present only because the model couples niche values with speciation rates. Without this coupling, an identical distribution of edge lengths is insufficient to induce enough imbalance in the tree, and $C(A)$ does not exhibit power-law behavior, as we saw in Fig. 6*B*. In conclusion, our specific mechanism to assign edge lengths not only reproduces realistic statistics of edge lengths, but also is essential for a realistic topological structure.

All data associated with the manuscript are accessible publicly, via https://github.com/zhiru-liu/niche-inheritance-trees.

Discussion

We have presented a model to explain the observed universal scaling of phylogenetic trees. We incorporate niche construction as an explicit evolutionary process in the tree growth. By analyzing the Niche Inheritance Model, we make two significant conclusions. First, a large niche construction effect, together with the absorbing boundary, leads to an apparent power-law regime in the tree topology. This is in the same range of A as observed in actual phylogenetic trees (5). The existence of the power-law $C(A)$ relation is a critical phenomenon, arising from the scale interference in time due to the singular dependence in the small speciation rate and small niche size limit. We demonstrate this by analyzing the cross-over of $C(A)$ from A^η at small A to $A \ln A$ at large A, with a r_ϵ-dependent threshold, reflecting a singular behavior in the Niche Inheritance Model as $r_\epsilon \to 0$. The second conclusion is that the Niche Inheritance Model is also able to recapitulate the scaling of the EAD. The EAD is not quite as sensitive a test of scaling as the topological scaling law for $C(A)$, since the Kingman coalescent and Yule processes both exhibit power-law scale invariance in the EAD. However, quantitatively, the power-law exponents are different from what one sees in nature. In our model, the niche construction effect generates a power-law scaling, and the exponent depends on the construction strength, which reflects the long memory of niche construction through the growth of the phylogenetic tree.

Our model has simple rules for the evolution of the tree. The significance is that there is a local in-time interplay between the speciation rate and niche availability and that this can generate a critical behavior in $C(A)$ because of the singularity induced by the cutoff of $r_\epsilon = 0$ at negative n. Our model shows that one must search for singular effects if a power-law $C(A)$ is to be recovered, just as is the case in the modern theory of critical phenomena. It might come as a surprise that such a simple model is able to recapitulate the otherwise inexplicable finding of a topological scaling law in phylogenetic trees. However, in matters of scaling, it is well established that minimal models suffice to capture the phenomena precisely, because extra layers of realism do not introduce new singularities that can change the scaling predictions (40).

There are several issues that require further investigation. First, we have predicted a scaling form for the cross-over point A_T as a function of r_ϵ, separating the power law and the $A \ln A$ regions. Actual phylogenetic trees, however, have small sizes, and the cross-over is undetectable. Therefore, we cannot be sure whether or not actual phylogenetic trees follow the critical scaling. Second, the exponent of the power-law behavior in our model is close to, but not exactly equal to, the reported values. We do not yet know if the scaling laws and scaling functions are universal, and, if not, what are the relevant or marginal operators in the branching process that control the scaling laws. The fact that the exponent for the EAD is sensitive to σ_n suggests that the absorbing boundary may actually be a marginal variable and not a relevant one. In order to understand this point, a technical

renormalization group analysis is required and would be the next step. Third, our model does not capture the decreasing cladogenesis rate that has been reported in actual phylogenetic trees (24, 65, 66). It remains to be examined whether incorporating mechanisms to model the empirical cladogenesis rate reduction would change the scaling and how.

Our results show that niche construction is more than a feedback between evolutionary and ecological processes arising when their timescales are not widely separated. Niche construction not only leads to a perturbation in the evolutionary trajectories of all components of an ecosystem, but also creates an indelible footprint on the evolutionary process that cannot be eliminated, even for very long times. These memory effects manifest themselves through the anomalous scaling laws that characterize observed phylogenetic trees.

ACKNOWLEDGMENTS. We thank James O'Dwyer and Kevin Laland for their critical reading of the manuscript and helpful suggestions. A portion of this paper was adapted from the dissertation of C.X. This work was supported by the NASA Astrobiology Institute under Cooperative Agreement NNA13AA91A issued through the Science Mission Directorate.

1. D. Haig, L. D. Hurst, A quantitative measure of error minimization in the genetic code. *J. Mol. Evol.* **33**, 412–417 (1991).
2. E. V. Koonin, A. S. Novozhilov, Origin and evolution of the universal genetic code. *Annu. Rev. Genet.* **51**, 45–62 (2017).
3. K. Vetsigian, C. Woese, N. Goldenfeld, Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10696–10701 (2006).
4. E. Hernandez-Garcia, M. Tuğrul, E. Alejandro Herrada, V. M. Eguiluz, K. Klemm, Simple models for scaling in phylogenetic trees. *Int. J. Bifurcation Chaos* **20**, 805–811 (2010).
5. E. A. Herrada *et al.*, Universal scaling in the branching of the tree of life. *PloS One* **3**, e2757 (2008).
6. P. J. Maldonado, "Computational approaches to stochastic systems in physics and ecology," PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL (2012).
7. N. Goldenfeld, Looking in the right direction: Carl Woese and evolutionary biology. *RNA Biol.* **11**, 248–253 (2014).
8. C. Colijn, G. Plazzotta, A metric on phylogenetic tree shapes. *Syst. Biol.* **67**, 113–126 (2018).
9. J. P. O'Dwyer, S. W. Kembel, T. J. Sharpton, Backbones of evolutionary history test biodiversity theory for microbes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8356–8361 (2015).
10. B. Burlando, The fractal dimension of taxonomic systems. *J. Theor. Biol.* **146**, 99–114 (1990).
11. B. Burlando, The fractal geometry of evolution. *J. Theor. Biol.* **163**, 161–172 (1993).
12. P. Bak, K. Sneppen, Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.* **71**, 4083–4086 (1993).
13. J. Chu, C. Adami, A simple explanation for taxon abundance patterns. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15017–15019 (1999).
14. R. V. Solé, J. Bascompte, Are critical phenomena relevant to large-scale evolution? *Proc. Roy. Soc. Lond. B* **263**, 161–168 (1996).
15. M. Newman, Self-organized criticality, evolution and the fossil extinction record. *Proc. Roy. Soc. Lond. B* **263**, 1605–1610 (1996).
16. R. V. Solé, S. C. Manrubia, M. Benton, S. Kauffman, P. Bak, Criticality and scaling in evolutionary ecology. *Trends Ecol. Evol.* **14**, 156–160 (1999).
17. C. R. Altaba, Universal artifacts affect the branching of phylogenetic trees, not universal scaling laws. *PloS One* **4**, e4611 (2009).
18. J. R. Banavar, A. Maritan, A. Rinaldo, Size and form in efficient transportation networks. *Nature* **399**, 130–132 (1999).
19. A. Masucci, Formal versus self-organised knowledge systems: A network approach. *Phys. Stat. Mech. Appl.* **390**, 4652–4659 (2011).
20. A. Herrada, V. M. Eguíluz, E. Hernández-García, C. M. Duarte, Scaling properties of protein family phylogenies. *BMC Evol. Biol.* **11**, 155 (2011).
21. A. O. Mooers, S. B. Heard, Inferring evolutionary process from phylogenetic tree shape. *QRB Q. Rev. Biol.* **72**, 31–54 (1997).
22. D. J. Aldous, Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci.* **16**, 23–34 (2001).
23. M. Blum, O. François, Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Syst. Biol.* **55**, 685–691 (2006).
24. H. Morlon, Phylogenetic approaches for studying diversification. *Ecol. Lett.* **17**, 508–525 (2014).
25. G. U. Yule, A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.*Philos. Trans. R. Soc. London. Ser. B* **213**, 21–87 (1924).
26. D. G. Kendall, On the generalized "birth-and-death" process. *Ann. Math. Stat.* **19**, 1–15 (1948).
27. E. Harding, The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Probab.* **3**, 44–77 (1971).
28. L. L. Cavalli-Sforza, A. W. Edwards, Phylogenetic analysis: Models and estimation procedures. *Evolution* **21**, 550–570 (1967).
29. D. E. Rosen, Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* **27**, 159–188 (1978).
30. J. S. Rogers, Central moments and probability distribution of Colless's coefficient of tree imbalance. *Evolution* **48**, 2026–2036 (1994).
31. D. Aldous, "Probability distributions on cladograms" *in Random Discrete Structures*, D. Aldous, R. Pemantle, Eds. (Springer, Berlin, Germany, 1996), pp. 1–18.
32. M. Steel, A. McKenzie, Properties of phylogenetic trees generated by Yule-type speciation models. *Math. Biosci.* **170**, 91–112 (2001).
33. I. Pinelis, Evolutionary models of phylogenetic trees. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 1425–1431 (2003).
34. J. F. C. Kingman, The coalescent. *Stoch. Process. Appl.* **13**, 235–248 (1982).
35. J. Pitman, Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870–1902 (1999).
36. N. Berestycki, Recent progress in coalescent theory. *Ensaios Matemáticos* **16**, 1–193 (2009).
37. M. Stich, S. Manrubia, Topological properties of phylogenetic trees in evolutionary models. *Eur. Phys. J. B* **70**, 583–592 (2009).
38. S. Keller-Schmidt, M. Tuğrul, V. M. Eguíluz, E. Hernández-García, K. Klemm, Anomalous scaling in an age-dependent branching model. *Phys. Rev.* **91**, 022803 (2015).
39. S. Keller-Schmidt, K. Klemm, A model of macroevolution as a branching process based on innovations. *Adv. Complex Syst.* **15**, 1250043 (2012).
40. N. D. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Boston, MA, 1992).
41. G. I. Barenblatt, *Scaling, Self-Similarity, and Intermediate Asymptotics* (Cambridge University Press, Cambridge, UK, 1996).
42. L. Y. Chen, N. Goldenfeld, Y. Oono, Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory. *Phys. Rev.* **54**, 376–394 (1996).
43. R. C. Lewontin, "Gene, organism and environment" in *Evolution from Molecules to Men*, D. S. Bendall, Ed. (Cambridge University Press, Cambridge, UK, 1983), pp. 273–285.
44. F. J. Odling-Smee, "Niche-constructing phenotypes" in *The Role of Behavior in Evolution*, H. C. Plotkin, Ed. (MIT Press, Cambridge, MA, 1988), pp. 73–132.
45. K. N. Laland, F. J. Odling-Smee, M. W. Feldman, Evolutionary consequences of niche construction and their implications for ecology. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10242–10247 (1999).
46. F. J. Odling-Smee, K. N. Laland, M. W. Feldman, *Niche Construction: The Neglected Process in Evolution* (Princeton University Press, Princeton, NJ, 2003), p. 37.
47. K. Laland, B. Matthews, M. W. Feldman, An introduction to niche construction theory. *Evol. Ecol.* **30**, 191–202 (2016).
48. K. N. Laland, J. Odling-Smee, M. W. Feldman, Causing a commotion. *Nature* **429**, 609–609 (2004).
49. K. N. Laland, K. Sterelny, Perspective: Seven reasons (not) to neglect niche construction. *Evolution* **60**, 1751–1762 (2006).
50. M. Gupta, N. Prasad, S. Dey, A. Joshi, T. Vidya, Niche construction in evolutionary theory: The construction of an academic niche? *J. Genet.* **96**, 491–504 (2017).
51. M. W. Feldman, J. Odling-Smee, K. N. Laland, Why Gupta et al.'s critique of niche construction theory is off target. *J. Genet.* **96**, 505–508 (2017).
52. E. Gilad, J. von Hardenberg, A. Provenzale, M. Shachak, E. Meron, A mathematical model of plants as ecosystem engineers. *J. Theor. Biol.* **244**, 680–691 (2007).
53. K. Cuddington, W. G. Wilson, A. Hastings, Ecosystem engineers: Feedback and population dynamics. *Am. Nat.* **173**, 488–498 (2009).
54. W. Gurney, J. Lawton, The population dynamics of ecosystem engineers. *Oikos* **76**, 273–283 (1996).
55. D. C. Krakauer, K. M. Page, D. H. Erwin, Diversity, dilemmas, and monopolies of niche construction. *Am. Nat.* **173**, 26–40 (2009).
56. R. H. Whittaker, S. A. Levin, R. B. Root, Niche, habitat, and ecotope. *Am. Nat.* **107**, 321–338 (1973).
57. C. G. Jones, J. H. Lawton, M. Shachak, Organisms as ecosystem engineers. *Oikos* **69**, 373–386 (1994).
58. A. Hastings *et al.*, Ecosystem engineering in space and time. *Ecol. Lett.* **10**, 153–164 (2007).
59. G. Barker, J. Odling-Smee, "Integrating ecology and evolution: Niche construction and ecological engineering" in *Entangled Life*, G. Barker, E. Desjardins, T. Pearce, Eds. (Springer, Dordrecht, Netherlands, 2014), pp. 187–211.
60. J. Odling-Smee, D. H. Erwin, E. P. Palkovacs, M. W. Feldman, K. N. Laland, Niche construction theory: A practical guide for ecologists. *Q. Rev. Biol.* **88**, 3–28 (2013).
61. R. MacArthur, R. Levins, The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.* **101**, 377–385 (1967).
62. T. Biancalani, L. DeVille, N. Goldenfeld, Framework for analyzing ecological trait-based models in multidimensional niche spaces. *Phys. Rev.* **91**, 052107 (2015).
63. S. Bornholdt, K. Sneppen, H. Westphal, Longevity of orders is related to the longevity of their constituent genera rather than genus richness. *Theor. Biosci.* **128**, 75–83 (2009).
64. N. Aktekin, The finite-size scaling functions of the four-dimensional Ising model. *J. Stat. Phys.* **104**, 1397–1406 (2001).
65. S. Nee, A. O. Mooers, P. H. Harvey, Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8322–8326 (1992).
66. D. L. Rabosky, I. J. Lovette, Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? *Evolution* **62**, 1866–1875 (2008).

PNAS
www.pnas.org

Supplementary Information for

Scale-invariant topology and bursty branching of evolutionary trees emerge from niche construction

Chi Xue, Zhiru Liu, Nigel Goldenfeld

Corresponding Author: Nigel Goldenfeld
E-mail: nigel@uiuc.edu

This PDF file includes:

Supplementary text
Figs. S1 to S4
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1915088117

Supporting Information Text

1. Robustness Upon Changes in Parameters

Niche Construction Strength:- As already discussed in the main text, the stronger the niche construction strength σ_n is, the more unbalanced the tree will be. Increasing σ_n drives $C(A)$ to the power-law behavior. However, if σ_n gets too large, it becomes possible that all the species cease to bifurcate and the tree can no longer grow to the desired size. Therefore, only a finite range of σ_n is available to investigate. We can characterize this range by plotting σ_n versus the probability of generating a sufficiently large tree. The probability is estimated by repeating the simulation many times and record the number of runs that generate a tree of size $N = 10000$. The result is shown in Fig. S1. We observe that beyond $\sigma_n = 3$, the probability of generating a large tree is near zero. Those trees with significant niche construction strength, though possible to generate at the cost of thousands of failures, should not be expected in nature and thus are not considered in our analysis.

As reported in the main text, the power-law exponent in $C(A)$ is robust throughout the physical range of the niche construction strength. On the other hand, σ_n does affect the power-law in EADs.

Extinction Strength:- Another adjustable parameter in the model is the extinction strength R_0. The motivation of defining the extinction rate in the given form is to use R_0 to limit the bifurcation rate. Indeed, if r of a node becomes comparable to R_0, $e(r) \approx 0.5$ and the node will be most likely removed from the tree. Therefore, we expect the niches and effectively bifurcation rates to stop growing after reaching close to R_0. However, if R_0 is very large, then it might take longer for the niche to reach R_0, and the stage before stagnation of niche growth could possibly alter the behavior of $C(A)$ and $S(k)$. We need to study how niches of nodes change during the simulation in order to understand whether the model is robust upon variation in R_0.

We assign an index to each node according to its birth time, and then collect the niche values of all nodes. In Fig. S2, we plot the absolute value of niche versus index for three vastly different extinction strengths. We observe two distinct stages in the simulation. First, for all R_0, the system undergoes a fast growing stage before saturating to the limit posed by R_0. No matter how large the limit is, the number of nodes in this stage is always around the order of a hundred, which is far fewer than the total number of nodes. Second, after the fast growing stage ends, the system enters a stage of fluctuation in niches values. In this stage, at any moment, active nodes possess a range of niche values, which increases with time as well as σ_n. In logarithmic scale, the center of the fluctuation is around R_0, but the range is independent of R_0.

With the distribution of niches in mind, we proceed to present the effect of changing R_0. Since the majority of nodes are in the second stage, we assert that it is the fluctuation stage that leads to the power-law behavior in both $C(A)$ and EAD. In Fig. S2d, we show that even for significantly different R_0, $C(A)$ is essentially identical to the one in the main text. In Fig. S3a, we plot the EAD for $\sigma_n = 2.5$ and $R_0 = 10^9$. The power-law is preserved, but small and large clades have branch lengths much larger than the other ones. If we keep only the fluctuation stage by removing the first 100 nodes from the tree, a distribution identical to $R_0 = 10$ is recovered in Fig. S3b. In the fast growing stage for large R_0, nodes have niches, thus bifurcation rates, orders of magnitudes smaller than nodes in the fluctuation stage. As a result, the edge lengths of the growing stage are orders of magnitudes longer than the fluctuation stage. The distribution is broken at both ends, because a node in the first stage can either be turned inactive, hence a small clade, or bifurcate into the rest of the tree, hence a large clade. Therefore, removing these nodes eliminates the peculiarity in the EAD, as shown in Fig. S3b.

To sum up, extinction strength brings nothing new to the behavior of the system. As long as we focus only on the fluctuation stage, both $C(A)$ and EAD are robust for all R_0 we tested.

Other Parameters:- The other parameters in the model are n_0, r_ϵ and μ_n. Since the inheritance of niche is multiplicative, n_0 only indirectly sets the length scale of the tree. As discussed in the main text, a nonzero r_ϵ reduces the boundary effect and gradually restores the symmetry. Lastly, a nonzero μ_n will adjust the frequency of generating nodes with negative niche, serving a similar role of σ_n. Based on the above argument, we conclude that none of these parameters will change the qualitative behavior of our model.

2. Mean Field Analysis of the Non-extinctive Niche Inheritance Model at the Infinite Time Limit

The extinction probability in the previous section depends positively on the speciation rate and thus induces a bias toward small rates in the evolutionary process. Here, we are interested in a mathematically simpler version without such a bias. This can be done by setting the extinction probability to be constant for all nodes. Furthermore, since any nonzero constant e can be mapped to $e = 0$ by effectively offsetting the speciation rates to $r(1 - e)$, we only need to look at the simplest situation with

$$e = 0. \qquad [S1]$$

The resulting model has no extinction and so we term it the Non-extinctive Niche Inheritance Model.

It should be pointed out that removing the bound on the values of niche and speciation rate will result in the speciation rate growing exponentially large in a short time, since the niche of a child changes proportionally to its parent's niche as in Eq. (2) in the main text. This is not biologically meaningful. Still, this simplified model can be handled mathematically from a mean-field point of view, and provides insights to the actual behavior of the non-extinctive model, as will be discussed below.

In the rest of the section, we conduct a mean-field theory calculation for the non-extinctive Niche Inheritance Model to derive the dependence of $C(A)$ on σ_n. We again work in the presence of the absorbing boundary $r_\epsilon = 0$, so that a certain number of nodes will turn inactive and not branch during the evolutionary process. We will discuss the regime of validity of the mean-field assumption at the end of this section.

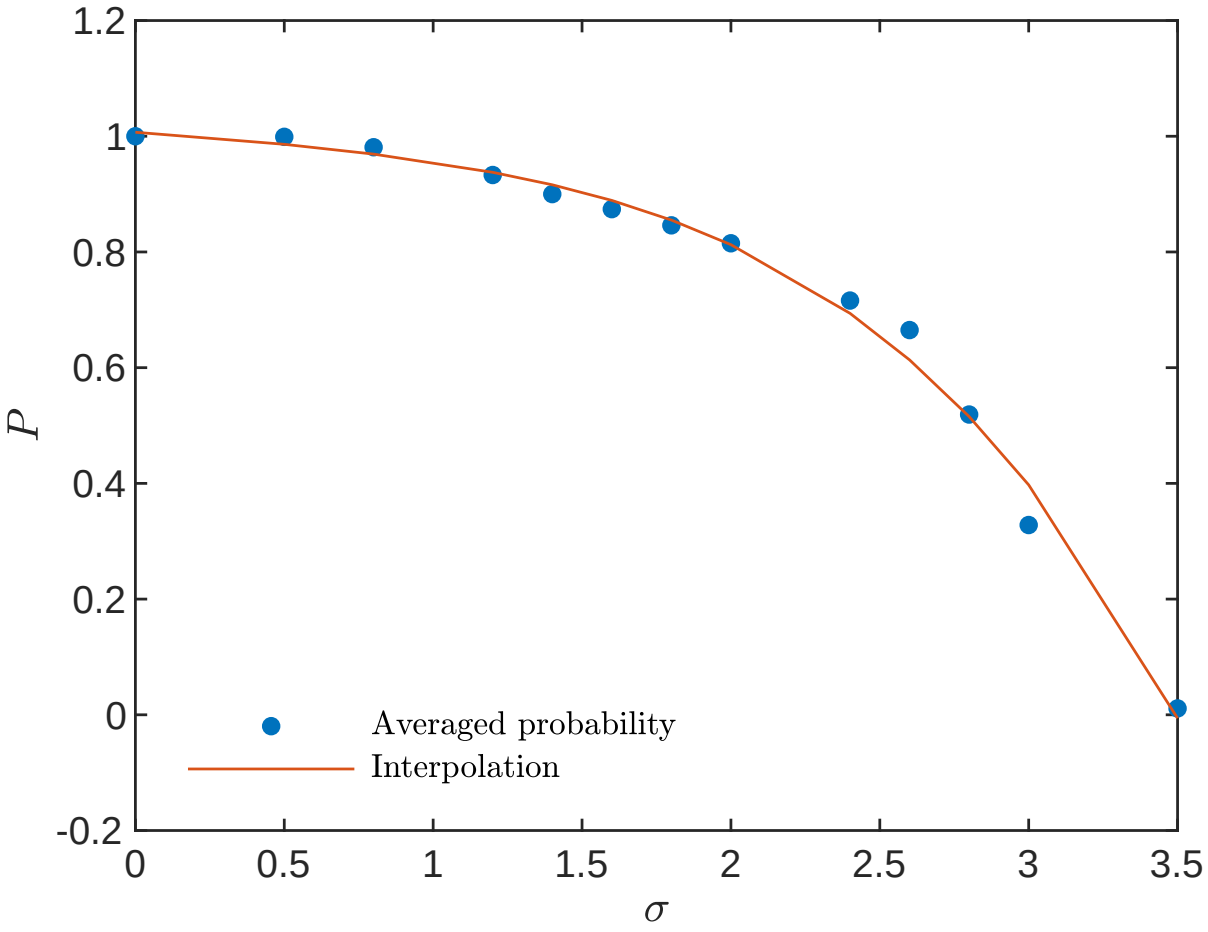

Fig. S1. The estimated probability of generating an infinite tree. Each data point is calculated by simulating 1000 times and recording the number of trees larger than $N = 10000$. The probability decreases to near zero beyond $\sigma_n = 3$. Other parameters used are $r_\epsilon = 0$, $\mu_n = 0$, $R_0 = 10$ and $n_0 = 1$.

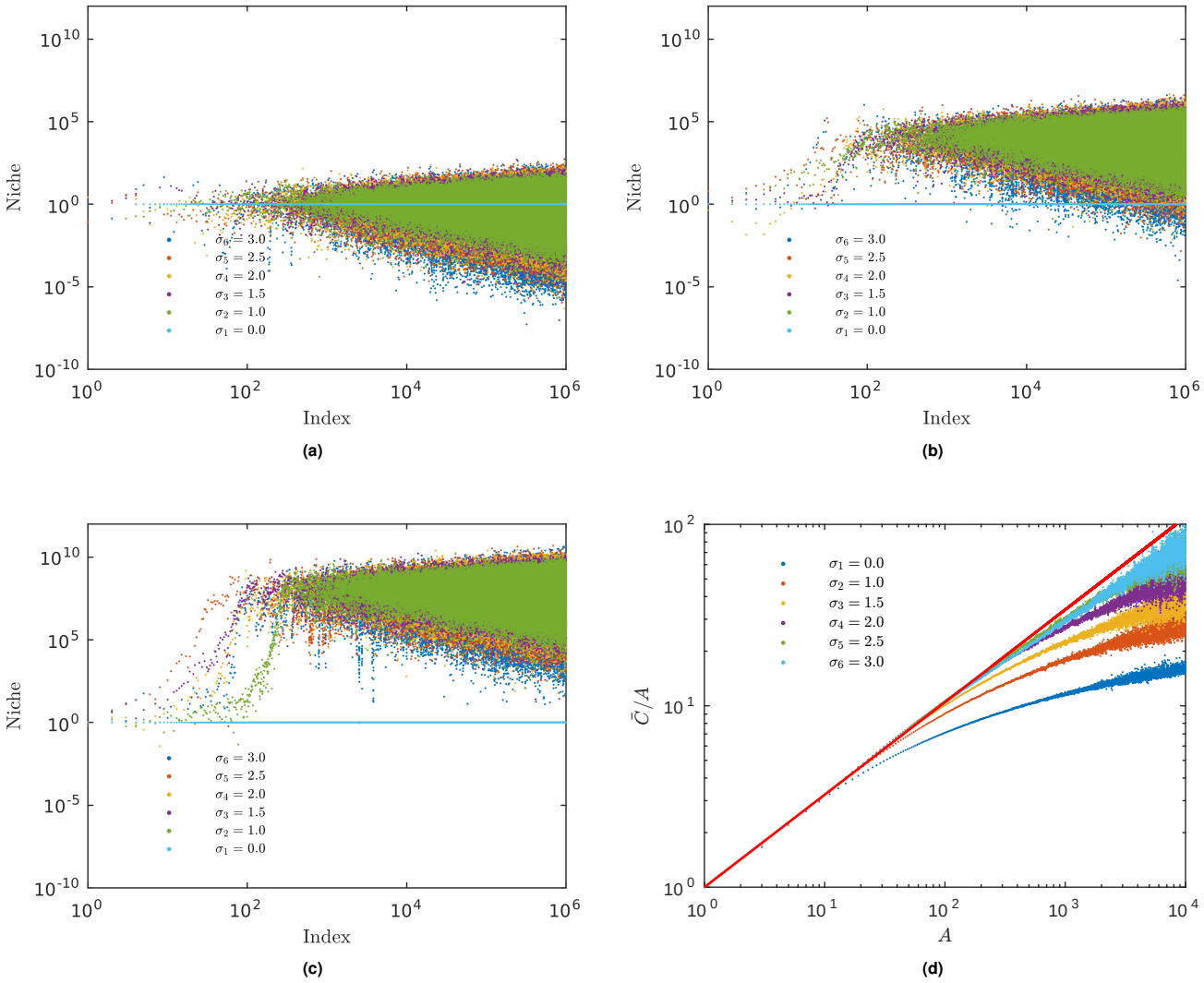

Fig. S2. (a) Absolute value of node niche. Node index is labeled according to its birth time. After around a hundred nodes, the simulation enters a fluctuation stage, where niche values differ greatly among nodes close in birth time. The fluctuation range is seen to be growing steadily as the simulation goes on. (b)(c) Niche-index graphs generated in the same way as (a), except with $R_0 = 10^5$ and $R_0 = 10^9$, respectively. The stage before fluctuation features a fast growth in niche values, and this stage ends when the niche values saturate to the limit posed by R_0. Notice that in logarithmic scale, the fluctuation range is the same regardless of R_0. (d) Averaged $C(A)$ with $R_0 = 10^9$. Both the power-law exponent and the scaling range are identical to $R_0 = 10$ in the main text, demonstrating the robustness of $C(A)$ under a drastic change in R_0.

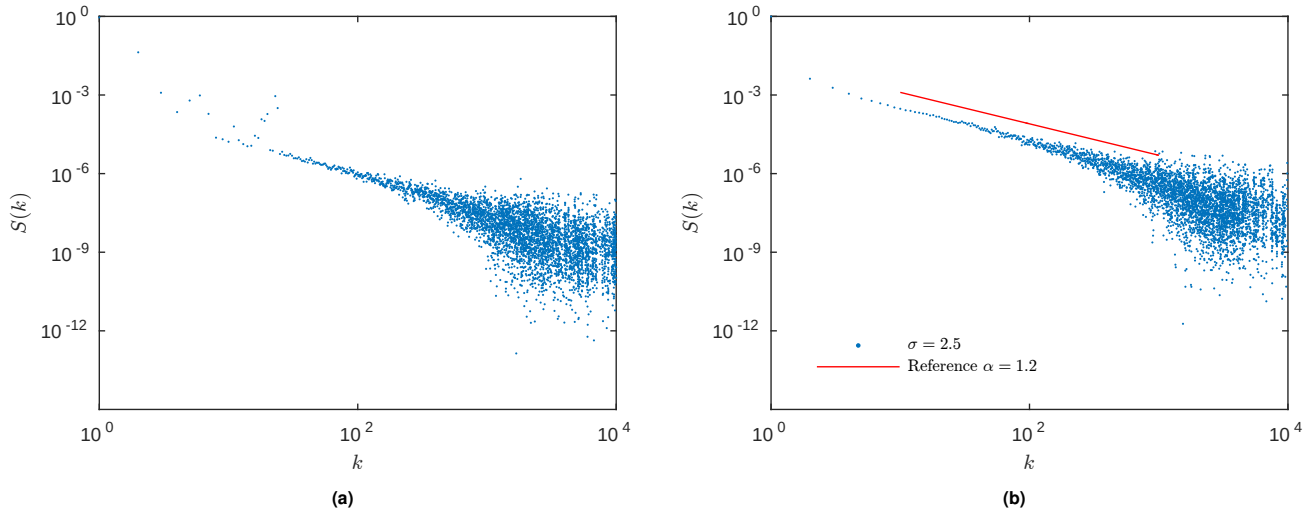

Fig. S3. (a) EAD calculated with all nodes for $\sigma_n = 2.5$, $R_0 = 10^9$. The region between clade size $k = 10$ to $k = 10^4$ is seen to behave similarly with $R_0 = 10$ in the EAD plot in the main text. (b) EAD calculated with nodes in the fluctuation stage only. The number of nodes removed is 100, determined by Fig. S2c. The behavior of EAD is now identical to $R_0 = 10$. Therefore, the power-law behavior originates from the fluctuation stage, which is robust under changes in R_0.

A. Deactivation Probability of Nodes. Comparing trees with different topologies, we observe the following facts. In a completely balanced binary tree, the leaf nodes are all active and can branch. In a completely unbalanced binary tree, only one child can branch and the other is inactive. A phylogenetic tree should lie somewhere in between the two extreme cases. If we define the deactivation probability of a leaf node as q, then

$$\begin{cases} q = 0, & \text{completely balanced binary tree,} \\ q = 0.5, & \text{completely unbalanced binary tree.} \end{cases} \quad [S2]$$

B. Dependence of C(A) on the Deactivation Probability. Suppose there are n_d leaf nodes, when the tree evolves to depth d. Then on average, $n_d q$ of the leaves will turn inactive, and each of the remaining $n_d(1-q)$ nodes will branch into two leaves at depth $d+1$. Therefore, we have a recursive relation for n_d,

$$n_{d+1} = 2n_d(1-q). \quad [S3]$$

The general expression for n_d is then calculated to be

$$\begin{cases} n_0 = 1, \\ n_1 = 2, \\ n_d = n_1 a^{d-1}, \end{cases} \quad [S4]$$

with $a = 2(1-q)$ as the average number of active children of one parent node. The full parameter range is $0 \leq q \leq 1$ and correspondingly $0 \leq a \leq 2$. However, for $1/2 < q \leq 1$ and $0 \leq a < 1$, the tree can not grow to a significant size, We thus exclude this situation from the consideration.

The subtree size A of a node at depth D is given by

$$A = \sum_{d=0}^{D} n_d. \quad [S5]$$

The average depth of nodes in the subtree is

$$\langle d \rangle = \frac{\sum_{d=0}^{D} d n_d}{A}. \quad [S6]$$

From the definition, C of node i can be written as $C(i) = \sum_{S_i} A(i)$, where S_i is the subtree rooted at node i. Let d_{ij} be the depth of node j in subtree i, or equivalently the number of edges between node i and j. It's easy to check that the definition can be rewritten using d_{ij} as

$$C(i) = \sum_{j \in S_i} (d_{ij} + 1) = \sum_{j \in S_i} d_{ij} + A(i) \quad [S7]$$

Dividing both sides by A leads to

$$C = A(\langle d \rangle + 1). \quad [S8]$$

which allows us to calculate C.

With Eq. (S4), we obtain the following explicit expressions of A and C in terms of D, for $0 \leq q < 1/2$ and $1 < a \leq 2$.

$$A = 1 + 2\frac{a^D - 1}{a - 1}, \quad [S9]$$

$$C = \frac{2}{a-1}[(D+1)a^D - 1] - \frac{2a(a^D - 1)}{(a-1)^2} + A. \quad [S10]$$

The $C(A)$ relation is further given as follows, by eliminating D from the above two equations.

$$C(A) = (A-1)\log_a\left[\frac{(A-1)(a-1)}{2} + 1\right] + \frac{2}{a-1}\log_a\left[\frac{(A-1)(a-1)}{2} + 1\right] + (A-1) + \frac{a-A}{a-1}. \quad [S11]$$

For the completely balanced tree with $a = 2$, $C(A)$ is reduced to

$$C(A) = (A+1)\log_2(A+1) - A, \quad [S12]$$

which has $C \sim A\ln A$ asymptotic behavior for large A.

As $q \to 1/2$ and $a \to 1$, we can derive the limit form of $C(A)$, given below, using L'Hôpital's rule.

$$C(A) \to \frac{A^2 + 1}{2}. \quad [S13]$$

Despite the different functional form, it has the same asymptotic scaling $C(A) \sim A^2$ at large A. A similar mean-field calculation has been performed in Ref. (S1). The authors reported the same qualitative result as what we have derived above, but with a slightly different model.

C. Dependence of C(A) on the Niche Construction Strength. Now we analyze the relationship between the parameter σ_n in the Niche Inheritance Model and the deactivation probability q.

Based on Eq. (1) and Eq. (2) in the main text, the child node turns inactive if $n_0 + n_0 x < 0$, where x is the random number drawn from the distribution $\mathcal{N}(\mu_n, \sigma_n^2)$ to characterize the niche construction effect. Therefore, we have the following equation to link q and σ_n.

$$q = \text{Prob}(r = 0) = \text{Prob}(x < -1) = \frac{1}{2}\text{erfc}\left(\frac{1}{\sqrt{2}\sigma_n}\right). \quad [S14]$$

So, for a given σ_n, we can compute q using the above equation, and then, with $a = 2(1-q)$, calculate $C(A)$ following Eq. (S11). This calculation allows us to use σ_n as a proxy for the niche construction strength, and by varying it, we can estimate its influence on the scaling laws within mean field theory. Figure S4 shows the $C(A)$ relations for different values of σ_n. When σ_n is finite, $C(A)$ always approaches $A \ln A$ when A is large. This can also be derived from Eq. (S11).

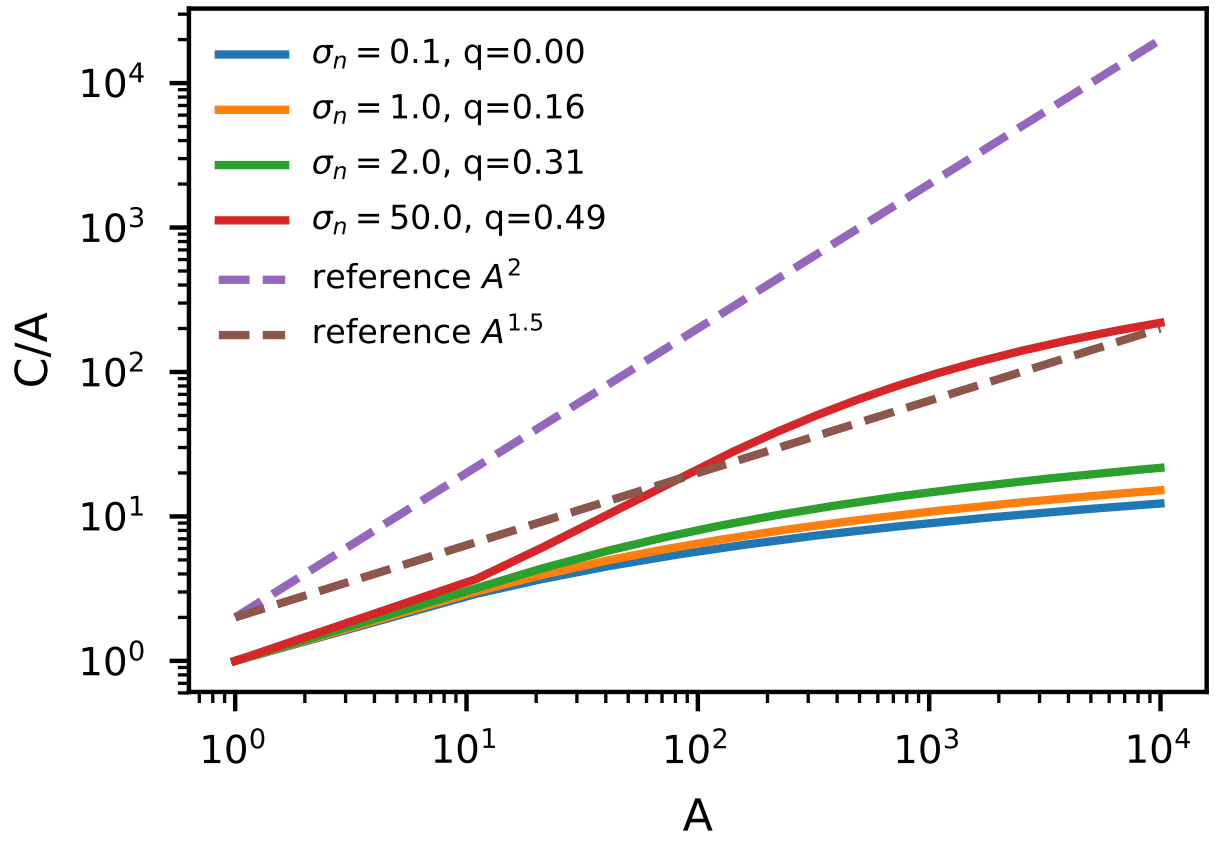

Fig. S4. Mean-field analytical $C(A)$ at different values of σ_n. For finite σ_n, $C(A)$ always approaches $A \ln A$. As $\sigma_n \to +\infty$, $q \to 1/2$ and the asymptotic behavior approaches $C(A) \sim A^2$.

In the presence of the absorbing boundary $r_\epsilon = 0$, the strength of the niche construction effect strongly impacts the topology of the phylogenetic tree. In the absence of niche construction, species are equivalent taxonomically and have the same bifurcation rate. The resultant tree is balanced, and $C(A) \sim A \ln A$.

D. Comments on the Mean-field Calculation. The above mean-field calculation succeeds in describing the qualitative behavior of $C(A)$ in extreme cases of $\sigma_n = 0$ and $\sigma_n \to +\infty$. However, it does not explain the power-law scaling at intermediate σ_n , as an effective scaling for an intermediate range of A.

There are two main caveats in the above calculation. First, the calculation does not account for any stochasticity in the process. It is applicable to an averaged situation, since n_d is the expected number of nodes at depth d. Second, the calculation is only correct with an infinite growth time of the tree. In order to use the recursion relation Eq. (S3), all active nodes at depth d have to be able to complete the branching process. This premise can always be achieved if the growth is terminated at

$T = +\infty$. However, if T is finite, then there will always be some active nodes that will not branch before the termination. This effectively leads to a larger deactivation probability than the constant q. This effect is significant for nodes with small speciation rates, which can occur at any depth. Therefore, the effective deactivation probability $\tilde{q}_d$ should be dependent on the distribution of speciation rate at depth d.

References

[S1] Stich M, Manrubia S (2009) Topological properties of phylogenetic trees in evolutionary models. *The European Physical Journal B* 70(4):583–592.