

# A consumer-resource model of phylogenetic diversification

Jessica Metzger

(Dated: January 8, 2025)

All taxonomic scales are important in the tree of life. But are some scales more important than others, and if so, why? Despite the rapid progress in recent decades towards understanding the multiscale structure of the phylogenetic tree, questions of scale-selection remain inconclusive. This is due in part to the over-reliance on phenomenological models in which scales must be imposed by hand. I develop a simple bottom-up model that (following recent literature) reproduces the anomalous power-law scaling observed in natural phylogenetic data, but also exhibits what appears to be scale-selection through a broken power-law crossover.

## I. INTRODUCTION & LITERATURE REVIEW

Are the familiar discrete taxonomic classifications (species, genus, etc.) motivated, and if so, what are the most important ones? Numerous phenomenological models for phylogenetic diversification were developed in the last century [1], predicting (with the support of fossil data) the existence of structure at all taxonomic scales. Moreover, a recent surge in “bottom-up” models invoking explicit eco-evolutionary coupling spurred by advanced molecular sequencing data has enabled interrogation of the more detailed structure of the phylogenetic tree across scales, e.g. [2]. Nevertheless, these models have remained focused on coarser multiscale trends like average power-law scaling, and not on peculiarities of any particular scale. This is at least partially because many existing models, even “bottom-up” ones, are too simplified to examine more fine-grained structure.

We refer to the over-importance of a particular scale as “scale-selection.” It can be thought of as a symmetry-breaking similar to the choice of wavelength in microphase separation, and as a discontinuity in the renormalization-group flow induced by coarse-graining genotypic (or phenotypic) similarities. It should manifest itself as a transition of various metrics, such as power-law exponents and fractal dimensions, occurring when the scale of observation is shifted.

We suspect that scale-selection ought to occur, for instance, at the species scale for sexually-reproducing organisms, sharply delineated by reproductive viability. One can invent mechanisms for scale-selection at almost any scale (e.g. multicellularity as a fundamental transition from single-cellularity; membrane-bound nucleus; vertebrate vs. invertebrate; etc.), thus introducing arbitrarily many possible places for scale-selection. Two questions arise: (1) what does quantitative data say about scale-selection, and (2) what simplified models reproduce characteristics of the observed scale-selection statistics?

Models of *scale-invariant* phylogenetic topology, e.g. [3, 4], were compared with pre-sequencing phylogenetic data (fossils, etc.) in Refs. [5–10] to moderate success. After the advent of molecular sequencing, the field was revolutionized, and more robust comparisons were made in works like [11–13]. Simulta-

neously, new macroscopic, top-down/phenomenological models for scale-invariant phylogenetic branching continue to abound, e.g. [2, 14–16]. Most models don’t attempt to predict scale-selection, and comparisons to data have either focused on multiscale averages of various metrics (neglecting the possibility for them to change with respect to scale) or examined different scale ranges separately. Even when power-law exponents are found to be conserved across different scales, e.g. [13], analyzing the scales separately may miss bumps or discontinuities that a full-scale analysis would find.

At a more mechanistic level, theoretical ecology has been revolutionized by simplified frameworks like consumer-resource models [17] and variants [18–22], Lotka-Volterra models and variants [23–27], and evolutionary game theory [27–30]; along with substantial progress in experiments. This has facilitated the development of more “microscopic,” bottom-up models of diversification, e.g. [18, 21, 28, 30–44]. For comprehensive reviews of diversification models (both top-down and bottom-up), see [1, 10, 45, 46].

My goal in this paper is to make some progress in helping unify these top-down and bottom-up approaches. Because bottom-up models don’t force multiscale scale homogeneity like many top-down models do, this type of approach can be useful for studying scale-selection, as well.

There are a few models and scenarios where scale-selection occurs trivially. It happens straightforwardly in sexually-reproducing organisms at the scale of genetic immiscibility, meriting the “species” taxonomic scale [31, 47–50]. Some models of asexual reproduction explicitly model the competitive interactions between species through a Gaussian kernel in continuous trait space which causes speciation on the kernel’s length scale [29, 30]. However, I am interested in the possibility of spontaneous selection of a scale that isn’t imposed by some existing length-scale; thus, I will not examine these types of scenarios.

I seek to unify bottom-up and top-down approaches through a physically-motivated “microscopic” model that can make “macroscopic” phylogenetic predictions. Thus, I work with a simplified consumer resource (CR) model. I will follow the evolution of the system over long periods of time, and study the statistics of the resulting phylogenetic tree. I will find that the tree obeys the clas-

sic power-law scaling, but exhibits an additional feature at an intermediate scale that could represent some kind of scale selection. I also develop a simplified model for the diversification process that can allow one to make analytical predictions about the branching topology. This simplified model could be helpful to guide us in searching for new, more robust metrics of scale-invariance and scale-selection in real systems.

## II. RESULTS

### A. Consumer-resource model

Suppose there are  $N$  individuals, who consume  $R$  self-replenishing resources. Each individual consumes  $R_c < R$  different resources. Denote the consumption preferences of individual  $i$  for resource  $\alpha$  as  $t_{i\alpha} \in \{0, 1\}$ . Denote the abundance of resource  $\alpha$  as  $r_\alpha$ .

Note that, unlike typical CR models, consumption preferences are discrete. This choice is made to force diversification: when consumption preferences are allowed to continuously vary, the population remains full of generalists [51]. It can be shown that specializing strategies, while able to fixate in the absence of mutations [18], relax to the generalist strategy under evolution due to its evolutionary stability [29]. This may be avoided by imposing nonlinear trade-off constraints on the resource consumption preferences [35]. For simplicity, I simply force them to be all-or-nothing, a reasonable choice for realistic systems.

The individuals undergo Poisson death and reproduction, with rates

$$r_{\text{death}} = \delta, \quad r_{\text{reproduce}, i} = \sum_{\alpha=1}^R t_{i\alpha} r_\alpha. \quad (1)$$

Upon reproduction of individual  $i$ , a new individual is created with strategy  $\mathbf{t}'$ , which is a mutated version of the parent strategy  $\mathbf{t}_i$  such that with probability  $\mu$  one consumption preference is switched to a randomly-chosen new one (distinct from individual  $i$ 's existing preferences).

The resources are continually replenished at a rate  $\gamma$  until they reach a carrying capacity  $K$ . They follow the overall dynamics

$$\dot{r}_\alpha(t) = \gamma \left(1 - \frac{r_\alpha}{K}\right) - \sum_{i=1}^{N(t)} t_{i\alpha} r_\alpha. \quad (2)$$

I have set various prefactors to unity since the units of time and of each  $r_\alpha$  are arbitrary. I will further restrict myself to  $\gamma = K = 1$ , and  $\delta = 0.1$ .

I will focus on the case of  $R_c = 2$ , where each individual may consume only 2 resources.

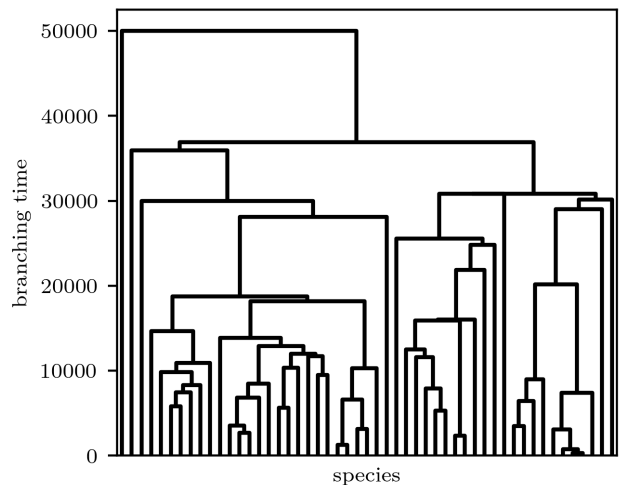


FIG. 1. Phylogenetic tree for  $R = 100$ ,  $\mu = 10^{-4}$ , and  $t_f = 5 \times 10^4$ .

### B. Branch size and cumulative branch size scaling

Simulations generate nontrivial phylogenetic trees (Fig. 1). I wish to compare them quantitatively with the literature via some metric. Following [2, 11] and others, I consider the metric given by the relative scaling of each node's branch size  $A$  and cumulated branch size  $C$ . For node  $i$ , define  $A(i)$  as the number of nodes in the sub-tree rooted at  $i$ , including  $i$  itself. If we let  $S_i$  be the sub-tree rooted at  $i$ , then  $A(i) = |S_i|$ . Define the cumulative branch size  $C(i)$  as  $\sum_{j \in S(i)} A(j)$ .

In trees with perfectly balanced topologies, i.e. where each node branches, one can show that  $C \sim A \ln A$  in the limit of large  $A$ . On the other hand, for a perfectly unbalanced topology where only one of each node's two daughter nodes branches, the scaling is  $C \sim A^2$  [14]. In general, the exponent of the power-law scaling is called  $\eta$ , i.e.  $C \sim A^\eta$ . An  $\eta$  between 1 and 2 is sometimes called “anomalous” because it doesn't match either of these “null” models [2].

The model not only possesses anomalous  $C$  vs.  $A$  scaling, but has a structure near  $A = A_c \approx 100$  (Fig. 2). This feature appears to be either a bump indicating systematic preference towards size- $A_c$  subtrees, or a transition between two different scaling regimes. This suggests the existence of scale-selection, though the mechanism is yet unclear. To gain a deeper understanding of the system, I turn to a more analytically tractable simplification.

### C. Resource utilization network

The model with  $R_c = 2$  has a convenient intuitive reformulation in terms of a network of interacting resources and its dual graph of interacting species. Construct a graph of  $R$  nodes, where each node represents a resource.

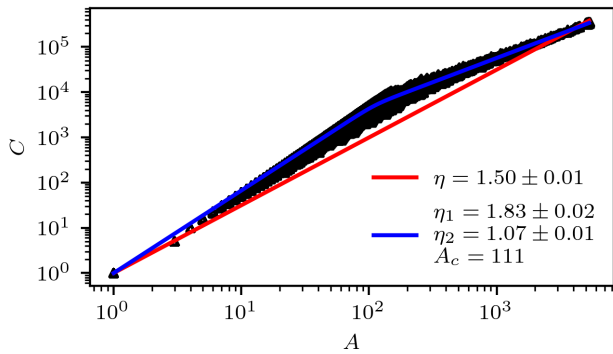


FIG. 2.  $C$  vs.  $A$  for  $R = 100$ ,  $\mu = 10^{-4}$ , and  $t_f = 10^6$ . Points are gathered from 26 independent seeds.

Since each species can consume only two resources, a species can be represented by an edge in the graph. Speciation corresponds to connecting a new resource node to an existing connected component (since mutations occur one at a time).

### 1. Coarsening

I will first consider the early stages of network formation, where the number of connected nodes (resources being consumed) is much smaller than  $R$ . Thus, I will assume that all new edges are formed with new nodes, making the resource network tree-like.

Given a tree topology, the population dynamics (1)-(2) determine the abundance of each species (edge) and resource (node). I take a deterministic approach, allowing species abundances to vary continuously, and assume the steady state is reached on timescales faster than mutations. One can in fact show that the only stable topology is an “aster,” a hub with an arbitrary number of spokes. Any time a path of length 3 is formed, the middle link goes extinct (this proof is relatively straightforward but not included here). Thus the resource utilization network can be represented as a set of asters connected by “ghost” edges (Fig. 3, above).

The steady state of (1)-(2) dictates that the total population of a  $k$ -aster (hub with  $k$  spokes) is  $(k + 1)\gamma/\delta$ , evenly distributed across all edges. Assuming speciation occurs evenly across all species at rates commensurate with their population sizes, larger asters thus gain connections faster. However, for  $k > 1$ , speciation at an outer node of a  $k$ -aster results in its fragmentation into a  $(k - 1)$ -aster and a 1-aster (isolated species), connected by a ghost edge. Once a member of an aster has been chosen for speciation, there is a 50% chance this leads to fragmentation of the aster and 50% chance this increases the order of the aster, depending on which resource preference is mutated (its hub or its spoke). Taking this all into account, the steady-state distribution over aster size

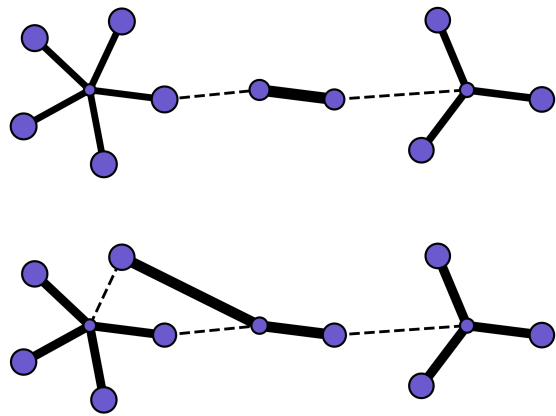


FIG. 3. Above: “linked aster” topology of the resource network, with dashed “ghost edges” representing extinct species. Edge width and node size are proportional to species and resource abundance, respectively. Below: as the population begins to saturate the resource space, cycles begin to form in the resource network, disrupting the existing structure.

$k$  turns out to be

$$p(k) \sim k^{-1}. \quad (3)$$

As a consequence, the degree distribution of the network without ghost edges also follows a power-law distribution.

The connection of this network with the phylogenetic tree is obfuscated by pruning the ghost edges. However, if we keep track of them, we find that the overall network has a similar topology to the phylogenetic tree. Namely, the phylogenetic tree is found by taking the dual of the full resource utilization network and resolving multifurcations appropriately. The last step, however, is not trivial, and even random outgroup assignment may affect the resulting network statistics in a biased way [12]. Nevertheless, the resource utilization network remains a good proxy for the phylogenetic tree which is both analytically manageable and realistic in the regime where the number of resources currently utilized is much less than  $R$ . In future work, I intend to use it to generate predictions for network scaling that can be compared to known scaling, e.g. [5, 6, 13].

### 2. Saturation

Above, I have assumed the number of available resources to be effectively infinite, so mutations will not generate cycles in the network. However, as the size of the cluster becomes nontrivial relative to  $R$ , this assumption will no longer hold and cycles will begin to appear, corresponding to “convergent evolution” in consumption preference space. (Species may also mutate back to their parent genotype, but this is significantly less likely, with a probability  $1/R$ .) Introducing a cycle will not merge the clusters; an edge will die to maintain the linked-aster

topology (Fig. 3, below). Once cycles begin to appear, the topology of the network can no longer be related to the topology of the phylogenetic tree.

In the simulation data depicted in Fig. 2, the number of species is as high as  $\sim 5 \times 10^3$ , while the number of resources is only  $R = 100$ . This is well into the saturated regime, indicating that the full resource network is highly cyclical.

We may conjecture that the crossover into the saturated regime corresponds to the power-law break in the branch size scaling, Fig. 2.

### III. DISCUSSION

While the qualitative features are conserved across a range of parameters and realizations, the exponents and crossover scales are by no means robust. Moreover, I have not removed extinct species from the phylogenetic trees before analysis. This represents a substantial departure from many real datasets, where only extant genotypes appear. For these and other reasons, this work is far from conclusive, and many loose ends remain to be tied

up.

I remarked earlier that the genotypic lengthscale-selection induced in sexually-reproducing individuals by reproductive inviability doesn't constitute a true "spontaneous" scale-selection in the purest sense. Ironically, the scale-selection observed in this model appears to itself likely be a direct consequence of the size of the resource pool. While this is a less direct mechanism, it isn't exactly "spontaneous."

Nevertheless, the model is undoubtedly able to reproduce some qualitative features of phylogenetic tree scaling [11], and has the advantage of being both physically-motivated and analytically tractable. There are many possible future directions. I have restricted my analysis to only one feature of tree topology. Future analysis may, for instance, compare the taxonomic abundance distributions across different taxonomic scales, something that has been studied extensively in real systems but primarily at the species level [52]. Non-topological statistics (e.g. branching times) also provide useful avenues to compare with real phylogenetic trees, which are known to display temporally inhomogeneous bursts of diversification [1, 2]. These represent many promising directions in which models of this type could be useful.

- 
- [1] H. Morlon, enPhylogenetic approaches for studying diversification, *Ecology Letters* **17**, 508 (2014).
  - [2] C. Xue, Z. Liu, and N. Goldenfeld, engScale-invariant topology and bursty branching of evolutionary trees emerge from niche construction, *Proceedings of the National Academy of Sciences of the United States of America* **117**, 7879 (2020).
  - [3] P. Bak and K. Sneppen, Punctuated equilibrium and criticality in a simple model of evolution, *Physical Review Letters* **71**, 4083 (1993), publisher: American Physical Society.
  - [4] J. Chu and C. Adami, A Simple Explanation for Taxon Abundance Patterns, *Proceedings of the National Academy of Sciences* **96**, 15017 (1999), arXiv:physics/0002001.
  - [5] B. Burlando, The fractal dimension of taxonomic systems, *Journal of Theoretical Biology* **146**, 99 (1990).
  - [6] B. Burlando, engThe fractal geometry of evolution, *Journal of Theoretical Biology* **163**, 161 (1993).
  - [7] M. E. J. Newman, Self-organized criticality, evolution and the fossil extinction record, *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**, 1605 (1997), publisher: Royal Society.
  - [8] R. V. Solé and J. Bascompte, engAre critical phenomena relevant to large-scale evolution?, *Proceedings. Biological Sciences* **263**, 161 (1996).
  - [9] R. V. Solé, S. C. Manrubia, M. Benton, S. Kauffman, and P. Bak, engCriticality and scaling in evolutionary ecology, *Trends in Ecology & Evolution* **14**, 156 (1999).
  - [10] M. G. B. Blum and O. François, Which Random Processes Describe the Tree of Life? A Large-Scale Study of Phylogenetic Tree Imbalance, *Systematic Biology* **55**, 685 (2006).
  - [11] A. Herrada, V. M. Eguíluz, E. Hernández-García, and C. M. Duarte, Scaling properties of protein family phylogenies, *BMC Evolutionary Biology* **11**, 155 (2011).
  - [12] C. R. Altaba, enUniversal Artifacts Affect the Branching of Phylogenetic Trees, Not Universal Scaling Laws, *PLOS ONE* **4**, e4611 (2009), publisher: Public Library of Science.
  - [13] E. A. Herrada, C. J. Tessone, K. Klemm, V. M. Eguíluz, E. Hernández-García, and C. M. Duarte, enUniversal Scaling in the Branching of the Tree of Life, *PLOS ONE* **3**, e2757 (2008), publisher: Public Library of Science.
  - [14] M. Stich and S. C. Manrubia, enTopological properties of phylogenetic trees in evolutionary models, *The European Physical Journal B* **70**, 583 (2009).
  - [15] E. Hernandez-Garcia, M. Tugrul, E. A. Herrada, V. M. Eguiluz, and K. Klemm, Simple models for scaling in phylogenetic trees, *International Journal of Bifurcation and Chaos* **20**, 805 (2010), arXiv:0810.3877 [q-bio].
  - [16] T. J. Davies, A. P. Allen, L. Borda-de Águia, J. Regetz, and C. J. Melián, enNeutral Biodiversity Theory Can Explain the Imbalance of Phylogenetic Trees but Not the Tempo of Their Diversification, *Evolution* **65**, 1841 (2011), eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.2011.01265.x>.
  - [17] R. M. Arthur, Species packing, and what competition minimizes\*, *Proceedings of the National Academy of Sciences* **64**, 1369 (1969), publisher: Proceedings of the National Academy of Sciences.
  - [18] A. Posfai, T. Taillefumier, and N. S. Wingreen, enMetabolic Trade-Offs Promote Diversity in a Model Ecosystem, *Physical Review Letters* **118**, 028103 (2017).
  - [19] R. Marsland, W. Cui, and P. Mehta, enA minimal model



- for microbial biodiversity can reproduce experimentally observed ecological patterns, *Scientific Reports* **10**, 3308 (2020), publisher: Nature Publishing Group.
- [20] R. Marsland, W. Cui, and P. Mehta, The Minimum Environmental Perturbation Principle: A New Perspective on Niche Theory, *The American Naturalist* **196**, 291 (2020), publisher: The University of Chicago Press.
- [21] M. Sireci, M. A. Muñoz, and J. Grilli, enEnvironmental fluctuations explain the universal decay of species-abundance correlations with phylogenetic distance, *Proceedings of the National Academy of Sciences* **120**, e2217144120 (2023).
- [22] W. Cui, R. M. III, and P. Mehta, enLes Houches Lectures on Community Ecology: From Niche Theory to Statistical Mechanics (2024), arXiv:2403.05497 [q-bio].
- [23] A. N. Kolmogorov, Sulla teoria di Volterra della lotta per l'esistenza, *Giornale Istituto Ital. Attuari*, 74 (1936).
- [24] A. Rescigno and I. W. Richardson, enThe struggle for life: I. Two species, *The bulletin of mathematical biophysics* **29**, 377 (1967).
- [25] enStability and Complexity in Model Ecosystems | Princeton University Press (2001), ISBN: 9780691088617.
- [26] S. Smale, enOn the differential equations of species in competition, *Journal of Mathematical Biology* **3**, 5 (1976).
- [27] M. Doebeli, enA model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms, *Population Ecology* **44**, 59 (2002), eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1007/s101440200005>.
- [28] S. A. Kauffman and S. Johnsen, enCoevolution to the edge of chaos: Coupled fitness landscapes, poised states, and coevolutionary avalanches, *Journal of Theoretical Biology* **149**, 467 (1991).
- [29] S. Geritz, Kisdi, G. Meszéna, and J. Metz, enEvolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree, *Evolutionary Ecology* **12**, 35 (1998).
- [30] U. Dieckmann and M. Doebeli, enOn the origin of species by sympatric speciation, *Nature* **400**, 354 (1999), publisher: Nature Publishing Group.
- [31] H. A. Orr, enThe population genetics of speciation: the evolution of hybrid incompatibilities., *Genetics* **139**, 1805 (1995).
- [32] M. Ackermann and M. Doebeli, Evolution of Niche Width and Adaptive Diversification, *Evolution* **58**, 2599 (2004), publisher: [Society for the Study of Evolution, Wiley].
- [33] B. F. D. Blasio and F. V. D. Blasio, *Dynamics of competing species in a model of adaptive radiations and macroevolution* (2005), arXiv:q-bio/0501025.
- [34] M. McPeck, enThe Ecological Dynamics of Clade Diversification and Community Assembly, *The American Naturalist* **172**, E270 (2008).
- [35] B. Østman, R. Lin, and C. Adami, enTrade-offs drive resource specialization and the gradual establishment of ecotypes (2014), arXiv:1310.8634 [q-bio].
- [36] I. Ispolatov, V. Madhok, and M. Doebeli, Individual-Based models for adaptive diversification in high-dimensional phenotype spaces, *Journal of Theoretical Biology* **390**, 97 (2016), arXiv:1507.04416 [q-bio].
- [37] L. F. Lafuerza and A. J. McKane, enThe role of demographic stochasticity in a speciation model with sexual reproduction, *Physical Review E* **93**, 032121 (2016), arXiv:1509.04568 [q-bio].
- [38] B. H. Good, S. Martis, and O. Hallatschek, enAdaptation limits ecological diversification and promotes ecological tinkering during the competition for substitutable resources, *Proceedings of the National Academy of Sciences* **115**, 10.1073/pnas.1807530115 (2018).
- [39] N. Anceschi, J. Hidalgo, T. Bellini, A. Maritan, and S. Suweis, enHow neutral and niche forces contribute to speciation processes? (2018), arXiv:1811.07623 [q-bio].
- [40] M. Amicone and I. Gordo, enMolecular signatures of resource competition: Clonal interference favors ecological diversification and can lead to incipient speciation, *Evolution* **75**, 2641 (2021), eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.14315>.
- [41] C. A. Maya-Lastra and D. A. R. Eaton, enGenetic incompatibilities do not snowball in a demographic model of speciation (2021).
- [42] S. V. Koniakhin, enEffects of resource competition on evolution and adaptive radiation (2023), arXiv:1907.12082 [q-bio].
- [43] E. Bellavere, C. H. S. Hamster, and J. A. Dijkman, enSpeciation in a MacArthur model predicts growth, stability, and adaptation in ecosystem dynamics, *Theoretical Ecology* **16**, 209 (2023).
- [44] V. M. Marquioni and M. A. M. d. Aguiar, enThe transition to speciation in the finite genome Derrida-Higgs model: a heuristic solution (2024), arXiv:2412.06565 [q-bio].
- [45] A. O. Mooers and S. B. Heard, Inferring Evolutionary Process from Phylogenetic Tree Shape, *The Quarterly Review of Biology* **72**, 31 (1997), publisher: The University of Chicago Press.
- [46] D. J. Aldous, Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today, *Statistical Science* **16**, 23 (2001), publisher: Institute of Mathematical Statistics.
- [47] W. Bateson, Heredity and Variation in Modern Lights, in *Darwin and Modern Science: Essays in Commemoration of the Centenary of the Birth of Charles Darwin and of the Fiftieth Anniversary of the Publication of The Origin of Species*, Cambridge Library Collection - Darwin, Evolution and Genetics, edited by A. C. Seward (Cambridge University Press, Cambridge, 2009) pp. 85–101.
- [48] T. Dobzhansky, enStudies on hybrid sterility, (1934).
- [49] T. Dobzhansky, enStudies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids, *Genetics* **21**, 113 (1936).
- [50] H. J. Muller, enReversibility in Evolution Considered from the Standpoint of Genetics, *Biological Reviews* **14**, 261 (1939), eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-185X.1939.tb00934.x>.
- [51] I have demonstrated this in simulations, which are not included in this paper.
- [52] B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White, enSpecies abundance distributions: moving beyond single prediction theories to integration within an ecological framework, *Ecology Letters* **10**, 995 (2007).