

RESEARCH ARTICLE

Generalism drives abundance: A computational causal discovery approach

Chuliang Song^{1,2*}, Benno I. Simmons³, Marie-Josée Fortin², Andrew Gonzalez¹

1 Department of Biology, Quebec Centre for Biodiversity Science, McGill University, Montreal, Canada, **2** Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, **3** Department of Ecology and Conservation, University of Exeter, Cornwall Campus, Penryn, United Kingdom

* chuliang.song@utoronto.ca



OPEN ACCESS

Citation: Song C, Simmons BI, Fortin M-J, Gonzalez A (2022) Generalism drives abundance: A computational causal discovery approach. *PLoS Comput Biol* 18(9): e1010302. <https://doi.org/10.1371/journal.pcbi.1010302>

Editor: Mercedes Pascual, University of Chicago, UNITED STATES

Received: February 1, 2022

Accepted: June 14, 2022

Published: September 29, 2022

Copyright: © 2022 Song et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the datasets analyzed in this study are publicly available. The dataset of plant-hummingbird communities is available from Dryad Digital Repository [dx.doi.org/10.5061/dryad.c270ft8](https://doi.org/10.5061/dryad.c270ft8). The dataset of coral reef fishes is available from the Reef Life Survey website [dx.doi.org/10.1038/s41559-020-01342-7](https://doi.org/10.1038/s41559-020-01342-7). The source code to produce the results is available on GitHub at <https://github.com/clsong/ReproduceChickenEgg>.

Funding: B.I.S. was supported by a Royal Commission for the Exhibition of 1851 Research

Abstract

A ubiquitous pattern in ecological systems is that more abundant species tend to be more generalist; that is, they interact with more species or can occur in wider range of habitats. However, there is no consensus on whether generalism drives abundance (a selection process) or abundance drives generalism (a drift process). As it is difficult to conduct direct experiments to solve this chicken-and-egg dilemma, previous studies have used a causal discovery method based on formal logic and have found that abundance drives generalism. Here, we refine this method by correcting its bias regarding skewed distributions, and employ two other independent causal discovery methods based on nonparametric regression and on information theory, respectively. Contrary to previous work, all three independent methods strongly indicate that generalism drives abundance when applied to datasets on plant-hummingbird communities and reef fishes. Furthermore, we find that selection processes are more important than drift processes in structuring multispecies systems when the environment is variable. Our results showcase the power of the computational causal discovery approach to aid ecological research.

Author summary

Ever since Aristotle, the chicken-or-egg causality dilemma has baffled researchers. Such causality dilemmas are abundant in ecological research, where causal directions are often assumed but not tested. An archetypal example is whether being a generalist causes a species to be more abundant, or whether being more abundant causes a species to be generalists. Without doubt, the gold standard to establish causal directions is controlled experiments. However, controlled experiments that can disentangle the direction of causality in this case are challenging because it involves controlling biotic or abiotic niche breadth. These challenges create an opportunity for computational tools to detect the most likely causal direction. Here, by adapting a set of recently developed computational methods, we provide strong evidence that generalism drives abundance, overturning the previously established direction. We hope our work raises awareness of the potential for

Fellowship. M.-J.F. acknowledges the funding of the CRC in Spatial Ecology. A.G. is supported by the Liber Ero Chair in Biodiversity Conservation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

computational discovery methods to address long-standing questions in ecology, especially increasingly large datasets become available.

Introduction

Identifying the causes of species abundance is a central question in ecology with direct implications for conservation management [1–3]. A ubiquitous ecological pattern in ecological communities is the skewed distribution of abundance with a few abundant species accompanied by many less abundant and rare species [4]. What causes this species abundance distribution is one of the most studied questions in ecological research. Theoretical studies have provided a diverse array of explanations for the emergence of uneven species abundance distributions, such as neutral theory [5, 6], niche partitioning [7–10], emergent neutrality [11, 12], and even statistical artifacts [13–15]. However, while these well-established theoretical explanations operate under different mechanisms, they generate similar and often empirically indistinguishable patterns (e.g., log-normal distributions). Thus, by considering only species abundance distributions, it is difficult to discern the main ecological drivers of this ecological pattern [16–18].

Here, we examine the role of species generalism as a predictor of abundance. Generalism is the (biotic or abiotic) niche breadth of a species [19, 20]; this is an archetypal feature of a species that strongly correlates with its abundance [21]. Specifically, we focus on biotic niche breadth and we will refer to the number of interacting partners in an interaction network as a measure of breadth. Despite the strong correlation, identifying the causal direction between abundance and generalism is not a trivial problem. Indeed, we have a chicken-and-egg dilemma: both causal directions make intuitive sense, and it is difficult to discern *a priori* which direction is correct. Specifically, whether being a generalist causes a species to be more abundant, or whether being more abundant causes a species to be more generalized [3, 21–25]. These two causal directions present fundamentally different views on multispecies dynamics in a local community (Fig 1). In one causal direction—via selection processes such as resource limitation (niche-based)—generalist species are competitively advantaged by having access to a wider range of resources, which causes higher abundance. In the other causal direction—via drift processes (neutrality-based)—abundant species are more likely to occupy more biotic niche space simply by coming into contact with more interaction partners than rare species, resulting in greater generalism. To further add to the complexity, the causal direction between abundance and generalism may not be unidirectional because it is unlikely that only selection or drift processes are occurring [26]. However, we do not know whether and when selection or drift processes *predominantly* structure multispecies dynamics in a local community [3, 26]. Thus, understanding the relative causal direction between abundance and generalism would increase our understanding of the roles of selection and drift in the structuring of ecological communities.

We employ a computational approach—without assuming a mechanistic model of how species abundances are generated—to directly identify the *causal* explanations for species abundance and other ecological patterns, such as species generalism. The computational problem of identifying the causal direction is known as *causal discovery* or structural identification in the field of statistics [27]. Note that causal discovery is different from *causal inference*, where causal discovery aims to find the causal direction while causal inference aims to find the causal strength given the preassigned causal directions. While causal inference has become a popular tool in quantitative ecology [28–31], causal discovery remains rarely used in ecology. This is

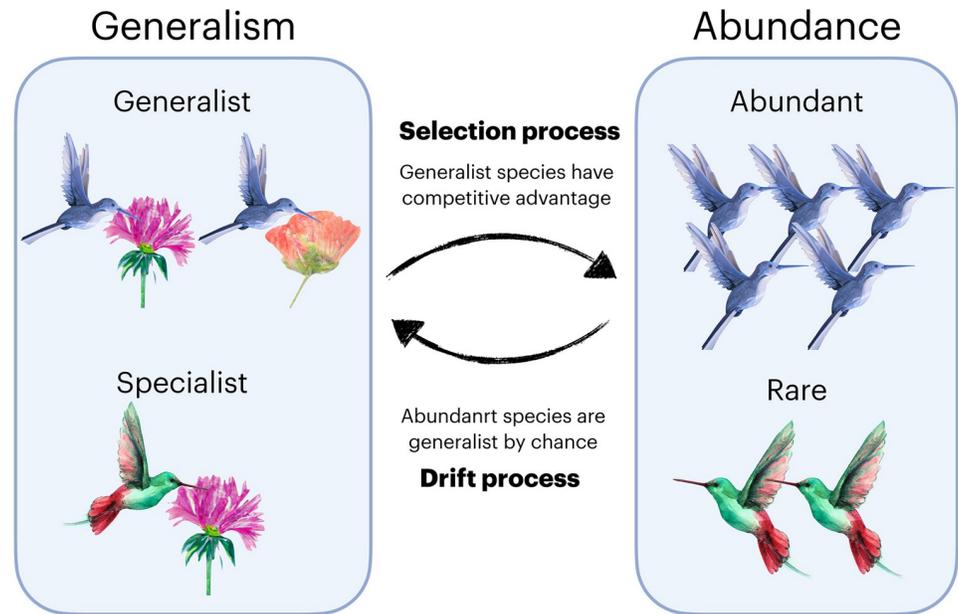


Fig 1. A chicken-and-egg dilemma of generalism and abundance. Empirical evidence shows that abundant species are also generalists. However, the causal direction is debated. If the community is mainly structured by selection processes, then species are more abundant because generalists have a competitive advantage. In contrast, if the community is mainly structured by drift processes, then species are more generalized because abundant populations have a higher chance of encountering more partners. The clip-arts of flowers and hummingbirds are made with DALL-E.

<https://doi.org/10.1371/journal.pcbi.1010302.g001>

partly because causal discovery is a notoriously difficult problem and has only taken off in the past decade [32–34]. Without any assumption, it is mathematically impossible to correctly distinguish causes and effects [35, 36].

To address this fundamental constraint, researchers in the field of causal discovery have recently developed a set of computational methods that can operate under minimal assumptions of the causal forms (Chapter 4 of [27]). In particular, two methods have a firm theoretical foundation and are widely applicable. One method is the nonlinear additive noise model based on nonparametric regression [27, 37], and the other method is the geometric-information inference based on information theory [38, 39]. Both methods take advantage of “asymmetry” resulting from the one-way causal direction: the nonlinear additive noise model focuses on the asymmetry of noise, while the geometric-information inference focuses on the asymmetry of information. In parallel, in the field of community ecology, a new method with a different theoretical foundation has been proposed to identify the causal direction between generalism and abundance [24]. A key assumption of Fort et al.’s method is the need to classify continuous data into binary categories (e.g. classifying species abundance data into either abundant or rare). However, the method can be sensitive to how the data are binarized, especially given the log-normal nature of species abundance distributions.

We take a computational causal discovery approach to the chicken-and-egg dilemma between generalism and abundance, and we apply three methods with independent theoretical foundations: (i) our refinement on Fort et al.’s method based on formal logic [24], (ii) the nonlinear additive noise model based on nonparametric regression [27, 37], and (iii) the geometric-information inference based on information theory [38, 39]. Our computational approach not only allows us to detect the causal directions, but also to use the relative strength of the causal directions as a proxy of the relative roles of either selection or drift processes. We

evaluate the sensitivity of these three methods to plant–hummingbird data across the Americas [25, 40] and reef fishes data from the Reef Life Survey program [41, 42]. All three methods consistently found strong evidence that generalism drives abundance in these plant-hummingbird communities and reef fish datasets. In addition, we found strong evidence that selection processes act more strongly than drift processes when local temperatures are more variable.

Methods

Data

We have used two cross-sectional datasets. One dataset of reef fishes contains measures of species abundance and generalism [42]. Species generalism in this dataset empirically estimates the habitat niche breadth of each fish species. The details of how niche breadth is inferred can be found in [42]. The other dataset has 25 plant-hummingbird systems across the Americas, which contains high-quality network structures and independent measures of plant and hummingbird abundance [43]. To empirically test the causal direction, we need independent estimates of species generalism and abundance (e.g., they cannot be both estimated from visitation data) [44, 45]. The details of this plant-hummingbird dataset can be found in [25]. Species generalism can be measured in different ways from network topology [3]. We have adopted two network metrics of generalism: the degree (number of interacting partners) and the normalized degree (the percentage of interacting partners over all locally available partners). These two metrics exhibit strong correlations with species abundance (S1 Data). Note that the generalism of plants is likely to be underestimated because we do not have the full pollinator-plant networks (e.g., insect pollinators are not sampled).

To understand the mechanisms of the causal directions, we further used the structural and environmental factors in the dataset of plant-hummingbird systems. For the structural factor, we computed four network properties: nestedness, modularity, connectance, and motif frequency. Nestedness is a key community property that has been shown to support biodiversity in mutualistic systems [46, 47]. Specifically, we measure nestedness with the combined nestedness metric (NODF_c), because it can provide an unbiased assessment of the level of nestedness across networks [47–50]. Modularity measures the extent to which a network can be divided into subnetworks that are strongly interacting within but are weakly connected across. Specifically, we measure modularity as Newman’s Q value [51]. Connectance measures the number of realized interactions over the number of all possible interactions. Motif frequency measures the relative occurrences of a motif in the whole network [52]. Specifically, we use all motifs with 2 to 4 species.

For the environmental factors, we gathered the temperature data for networks with location information from the public repository Terraclimate [53]. Fig P in S1 Data shows a strong association between the nestedness factor and the temperature factor, confirming previous findings [47].

Causal discovery method: Refining Fort et al. (2016)’s algorithm

We briefly describe the method proposed by [24]. Suppose species’ abundance and generalism are both binarized. Specifically, the abundance of a species is categorized as either abundant or rare, while the generalism of a species is categorized as either generalist or specialist. This combines to give us four qualitative categories. Given the strong association between abundance and generalism, we would frequently observe two combinatory categories—species being abundant and generalist concurrently or being rare and specialist concurrently—regardless of the causal direction between abundance and generalism. The crux of this method is that the causal direction would determine the frequency of the other two combinatory categories. If

abundance causes generalism, then we are unlikely to observe species being abundant and specialist concurrently. In contrast, if generalism causes abundance, then we are unlikely to observe species being rare and generalist concurrently.

The justification of Fort et al.'s method [24] relies on the formal logic of binarized categories. However, abundance and generalism are continuous gradients. To address this issue, [24] has proposed two methods to binarize continuous data of abundance and generalism, which was further refined in [25]. We follow the refinement proposed by [25]: specifically, a species is classified as abundant with probability of the rescaled abundance $\frac{\text{abundance} - \min(\text{abundance})}{\max(\text{abundance}) - \min(\text{abundance})}$, and classified as rare with one minus the rescaled abundance. Similarly, we can classify species into generalist and specialist: a species is classified as generalist with probability of the rescaled generalism $\frac{\text{generalism} - \min(\text{generalism})}{\max(\text{generalism}) - \min(\text{generalism})}$, and classified as specialist with one minus the rescaled generalism. However, the argument of the formal logic on the binary variables does not immediately provide a justification of this new methodology on probabilistic variables. Thus, we argue that it needs to be tested with a proper null model [54].

We generate, therefore, a simple null model to show that the original Fort et al.'s method [24] is sensitive to nonlinear causal relationships. We consider two cases where the variable X causes variable Y . In the first case, X is generated from a uniform distribution, while Y is generated as X plus some noise, where the magnitude of the noise decreases as X increases. The decreasing magnitude of the noise fills the assumption in the method (i.e., X being large and Y being small are unlikely to co-occur), as follows

$$\begin{aligned} x &\sim \text{Uniform}(0, 1), \\ y &\sim x + \text{Normal}(0, g(x)), \end{aligned} \tag{1}$$

where $g(x)$ is a monotonically decreasing function of x . In the other case, everything remains the same except that Y is generated from X with a non-linear function. Formally, we have

$$\begin{aligned} x &\sim \text{Uniform}(0, 1), \\ y &\sim x^2 + \text{Normal}(0, g(x)). \end{aligned} \tag{2}$$

The key difference between these two cases lies in the marginal distributions of the effects Y . In the first case, the noise induces more “large” values of Y than “small” values (Panel A in Fig B in [S1 Data](#)). In the second case, the presence of nonlinear causal relationships reverses this pattern, inducing more “small” values of Y than “large” values (Fig Panel C in Fig B in [S1 Data](#)). As a result, we would expect to see more species with small x and large y concurrently in the first case (indication of X causing Y), while more species with large x and small y in the second case (indication of Y causing X). However, the ground truth is that X causes Y in both cases. Despite the appeal of the method when applied to binary data, the original Fort et al.'s method [24] essentially compares which variable has a more skewed marginal distribution when applied to continuous data. In other words, whether species abundance, without reference to its generalism, is more skewed, or vice versa. This is not ideal as this causal discovery method should be based on the joint distribution (i.e., the association between abundance and generalism) instead of the marginal distribution (i.e., the isolated property of either abundance or generalism).

The potential bias of the method can be remedied by rescaling the marginal distributions. We first consider the marginal distribution of species abundance. The lognormal distribution of species abundance is one of the few universal patterns in community ecology [55]. The dataset we use unsurprisingly exhibits lognormal species abundance distribution (Fig E in

[S1 Data](#)). We then consider the marginal distribution of generalism. Unlike species abundance, the distribution of generalism is not log-normal (Fig F in [S1 Data](#)). If we use the degree (number of interacting partners) as a measure of generalism, the distribution of degree commonly follows a truncated power law $P(k) \sim k^{-\gamma} e^{k/k_c}$, where k is the degree, γ is the critical exponent, and k_c is the cutoff degree [56–58]. However, the distribution of normalized degree (i.e., the percentage of interacting partners over all locally available partners) does not tend to follow a truncated power law. Thus, we face a problem of model selection requiring us to transform different distributions with different metrics.

To solve this issue, we use a semiparametric approach to transform the skewed marginal distributions of both species' abundance and generalism [59]. The marginal distribution is transformed into a normal distribution with a mean 0.5. The choice of the mean is to balance the proportions of “small” and “large” values. We then follow the same procedure adopted in the previous method by [25] to normalize the values into probability. Fig 2B illustrates this method. Importantly, the transformation of the marginal distributions preserves the strong association between abundance and generalism in the joint distribution (Fig A in [S1 Data](#)). Fig C in [S1 Data](#) shows, with simulated data, that our refined method works for some ecologically relevant nonlinear causal relationships.

Testing other methods of casual discovery

To provide stronger confidence in the detected causal directions, we used two widely adopted methods in causal discovery that are distinct from Fort et al.'s method [24]: nonlinear additive noise model and information-geometric inference [27, 37]. Here, we provide a succinct summary of these methods.

First, we focus on the method based on nonlinear additive noise model. Suppose the variable Y is caused by variable X . This method assumes that Y is generated by a (potentially nonlinear) function f of variable X with some noise ϵ that is independent of X . Formally, we can write it as

$$Y = f(X) + \epsilon, \quad \epsilon \perp\!\!\!\perp X. \quad (3)$$

The assumption of additive noise is common in ecological models. For example, additive noise may represent measurement error [60–62] or stochastic dynamics [63, 64].

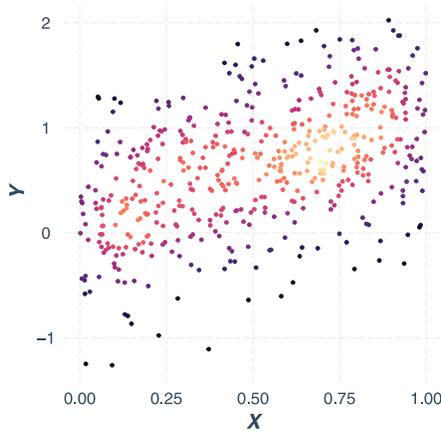
This method takes advantage of the asymmetry of noise. Specifically, the noise $\epsilon(Y)$ is only independent of X but not independent of Y if Y is caused by X . Vice versa, the noise $\epsilon(X)$ would only be independent of Y but not of X if X is caused by Y . Thus, by testing whether the residuals are independent of X or Y , we can know which variable is the cause. To implement this method, we first fit the nonlinear generating function using the non-parametric generalized additive model [65, 66]. We then calculate the residual and use the kernel-based tests for independence [67, 68]. Details can be found in [S1 Data](#). Fig 2C illustrates this method.

Then, we focus on the method based on information-geometric inference. Suppose again that the variable Y is caused by variable X . This method assumes a deterministic mapping between X and Y . More formally, we have $Y = f(X, \epsilon)$, where f is any diffeomorphism (i.e., bijective and its inverse is differentiable) function and the noise ϵ is constant (note that noise does not have to be additive as in Eq 3).

This method takes advantage of the asymmetry of information [38, 39]. Intuitively, if X causes Y in a nonlinear manner with noise, then some information in X would be lost in Y . More formally, the marginal distribution of Y tends to peak when f has a small slope (i.e., there is negative correlation between $P(Y)$ and f), while the marginal distribution of X does not correlate with the slope of f . Researchers have proved that this result can be equivalently expressed

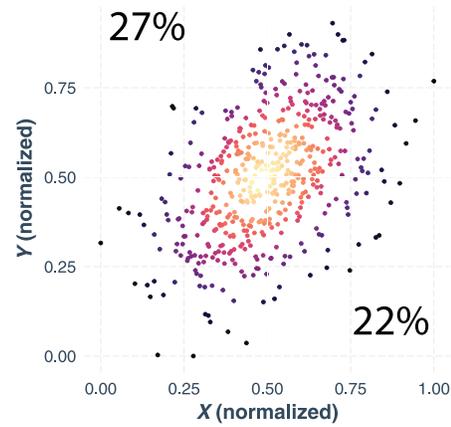
A. Illustrated example

$$X \rightarrow Y$$



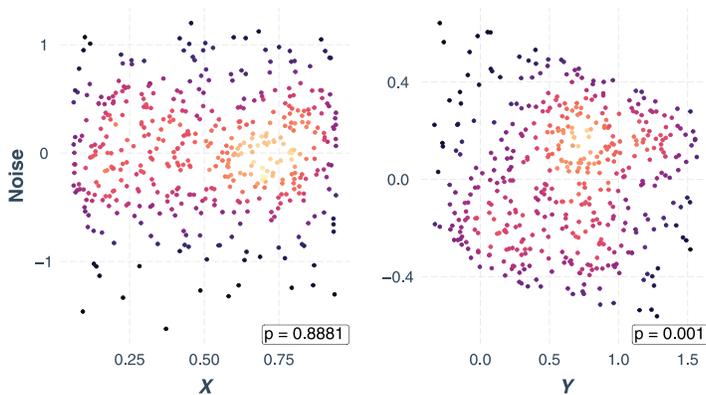
B. Formal logic

$$P(\text{small } X \ \& \ \text{large } Y) > P(\text{large } X \ \& \ \text{small } Y)$$



C. Nonlinear additive noise model

$$Y = f(X) + \epsilon(Y) \Rightarrow \epsilon \perp\!\!\!\perp X \ \& \ \epsilon \not\perp\!\!\!\perp Y$$



D. Geometric-information inference

$$H(X) \geq H(Y)$$

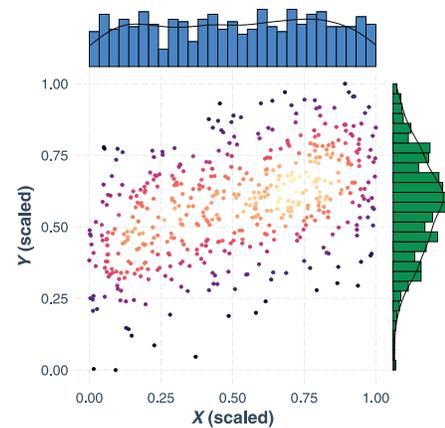


Fig 2. Illustration of the causal discovery methods. Panel (A) shows a hypothetical dataset where X is the cause and Y is the effect. We generate this dataset as an ensemble $y \sim x^d$ with white noise, where d ranges from 0.5 to 1. The color of the points represents the density of data: the darker it is, less points it represents. Panel (B)-(D) illustrate three causal discovery methods we adopted. Panel (B) illustrates the method based on formal logic. The criteria of X being the cause is that the proportion of points with small x and large y values is greater than that of points with large x and small y values. The marginal distributions are both normalized to keep roughly equal proportions of small and large samples. Normalization can help avoid the bias in the method regarding skewed marginal distributions. Panel (C) illustrates the method based on nonlinear additive noise model. This method assumes that the effect (i.e., Y) is some potentially nonlinear transformation of the cause (i.e., $f(X)$) plus some noise that is independent of the cause (i.e., $\epsilon(Y)$). The criteria of X being the cause is that the residuals from the regressions should be independent of X but not of Y . The form of the nonlinear transformation $f(X)$ and consequently the residuals can be fitted through nonparametric regressions. Panel (D) illustrates the method based on information theory. The criteria of X being the cause is that X has a higher entropy than Y . The marginal distributions are both scaled via a linear transformation between 0 and 1 for a fair comparison of entropy.

<https://doi.org/10.1371/journal.pcbi.1010302.g002>

as an entropy-based criterion. Formally, if X causes Y , we have

$$H(X) \geq H(Y), \tag{4}$$

where H denotes the differential Shannon entropy. Vice versa, if Y causes X . To implement

this method, we estimate the empirical entropy of X via [69]:

$$\hat{H}(X) = \psi(m) - \psi(1) + \frac{1}{m-1} \sum_{i=1}^{m-1} \log(x_{i+1} - x_i), \quad (5)$$

where $\psi = \frac{d \ln \Gamma(x)}{dx}$ is the digamma function, x are ordered ascendingly ($x_i \leq x_{i+1}$) and rescaled to be bounded between 0 and 1, and $\log(0)$ is defined as 0. The empirical entropy $\hat{H}(Y)$ can be similarly calculated. Details can be found in [S1 Data](#). [Fig 2D](#) illustrates this method.

Results

Using our refinement of Fort et al.'s method [24], we found that generalism drives abundance for species in both datasets. For the sake of brevity, we focus on the results of the plant-hummingbird system here and refer the reader to [S1 Data](#) for the details of the results obtained with the reef fishes dataset. For the dataset of plant-hummingbird systems, [Fig 3A and 3B](#) show that the proportion of species that are abundant and specialist concurrently (indication of generalism as the cause) is generally higher than the species that are rare and generalist concurrently (indication of abundance as the cause). Specifically, 19 out of 25 communities exhibit evidence that generalism is the cause in these hummingbird species, while 23 out of 25 communities exhibit such evidence for plant species. This result is not obtained when analyzed with the original Fort et al.'s method [24] ([Fig 3C and 3D](#)). This contrast is expected given that the original method is biased to assign the variable with a more skewed marginal distribution as the cause (see [Fig G](#) in [S1 Data](#) with the transformed unskewed marginal distributions). Formal analysis of statistical significance can be found in [S1 Data](#).

In addition, these exceptions, where abundances appear to be the cause, only occur when the communities are structured with poorly nested networks and under relatively low environmental variation. [Fig 4\(A\) and 4\(C\)](#) shows a strong correlation between the level of nestedness and the relative strength of the causal direction [48, 50]. Specifically, the causal evidence for generalism as a driver of observed patterns, especially for plants, is stronger when the community has a more nested structure. Among four structural factors we studied (nestedness, modularity, connectance, and motif analysis), nestedness shows the most strong association with the causal strength (see [S1 Data](#) for details). [Fig 4\(B\) and 4\(D\)](#) shows a strong correlation between the mean temperature and the relative strength of the causal direction. Specifically, the causal evidence for generalism as a driver of observed patterns, especially for hummingbirds, is stronger when the community is under greater environmental variations. Further statistical analysis with dimension reduction and nonparametric statistics can be found in [S1 Data](#).

The casual discovery approaches based on additive noise model and information-geometric inference support the result that generalism drives abundance. With the additive noise model, we found that generalism is likely to be the cause (p value = 0.73 for animals and = 0.64 for plants) while abundance is highly unlikely to be the cause (p value $< 10^{-3}$ for both animal and plants). More details on the model fit (including uncertainty estimation) can be found in [S1 Data](#). With the information-geometric inference, we found that the marginal distributions of generalism are much higher than those of abundance, which indicates that generalism is more likely to be the cause ([S1 Data](#)). The results for the reef fishes dataset are qualitatively the same for all methods ([S1 Data](#)).

Discussion

We have studied whether abundance drives generalism or the other way around via a computational causal discovery approach. We have used three independent methods of causal

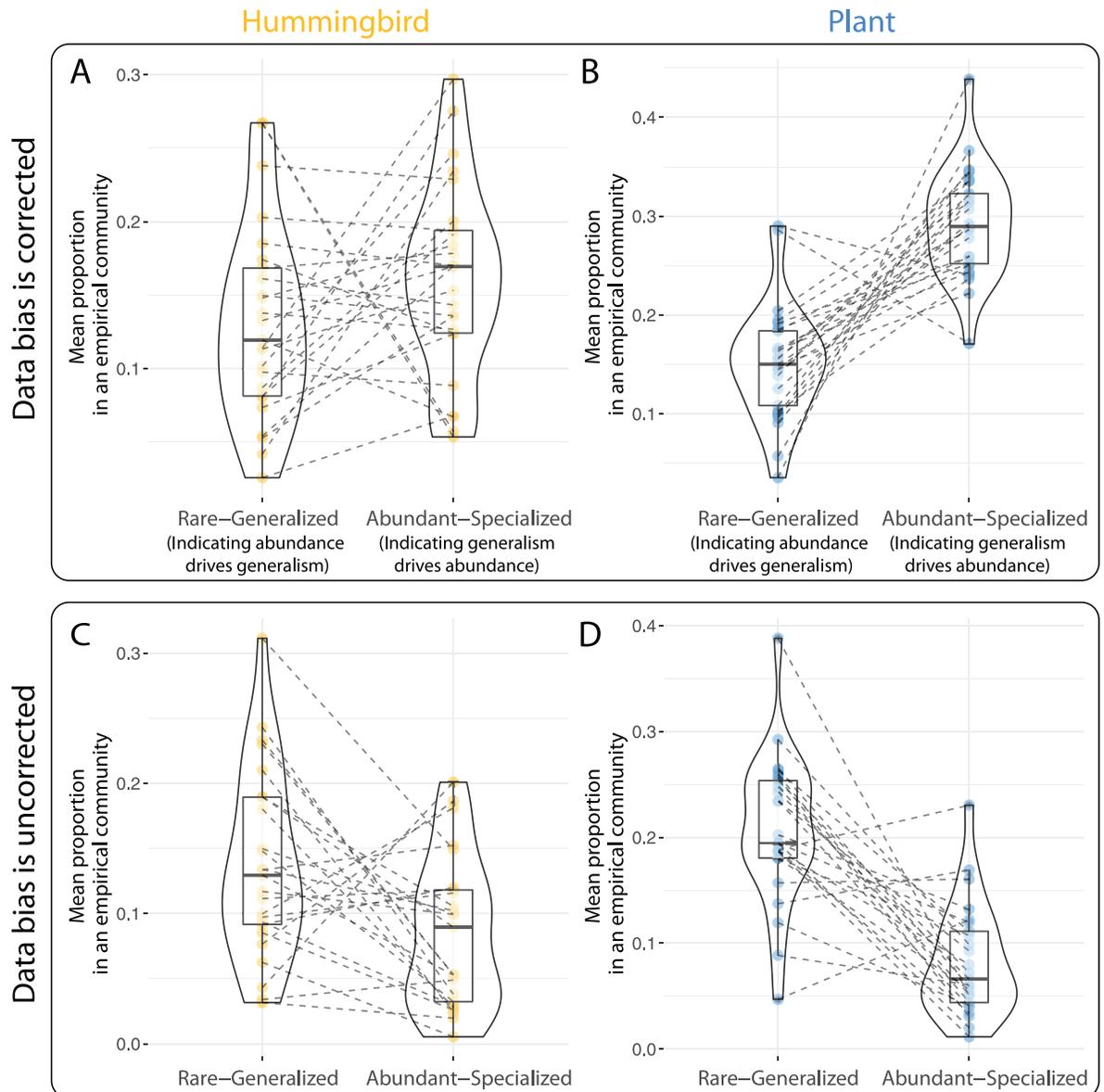


Fig 3. Generalism drives abundance. We apply the causal discovery method based on formal logic to detect the causal direction in an empirical dataset of 25 hummingbird-plant communities. In each panel, the *x* axis shows the two categories (rare and generalized species versus abundant and species species), and the *y* axis shows the mean proportion of species in that category. Each point denotes a different empirical community. Each line connects two points in the same empirical community. If the line is going up (meaning there are more species being abundant and specialized), it indicates that generalism drives abundance, and vice versa. The original method (panels C and D) suggests that generalism drives abundance in most communities (this is expected because the marginal distribution of abundance is more skewed than that of generalism). In contrast, our refined method has removed the bias regarding skewed marginal distributions, and it (panels A and B) suggests that generalism drives abundance in most communities. The qualitative patterns are similar among hummingbirds (panel A) and plants (panel B).

<https://doi.org/10.1371/journal.pcbi.1010302.g003>

discovery: a refined method of Fort et al.'s based on formal logic [24], the nonlinear additive noise model, and the geometric-information inference. We have found strong evidence that generalism causes species abundance in both datasets of plant-hummingbird communities and reef fishes. We have also found that the causal evidence for generalism as a driver of

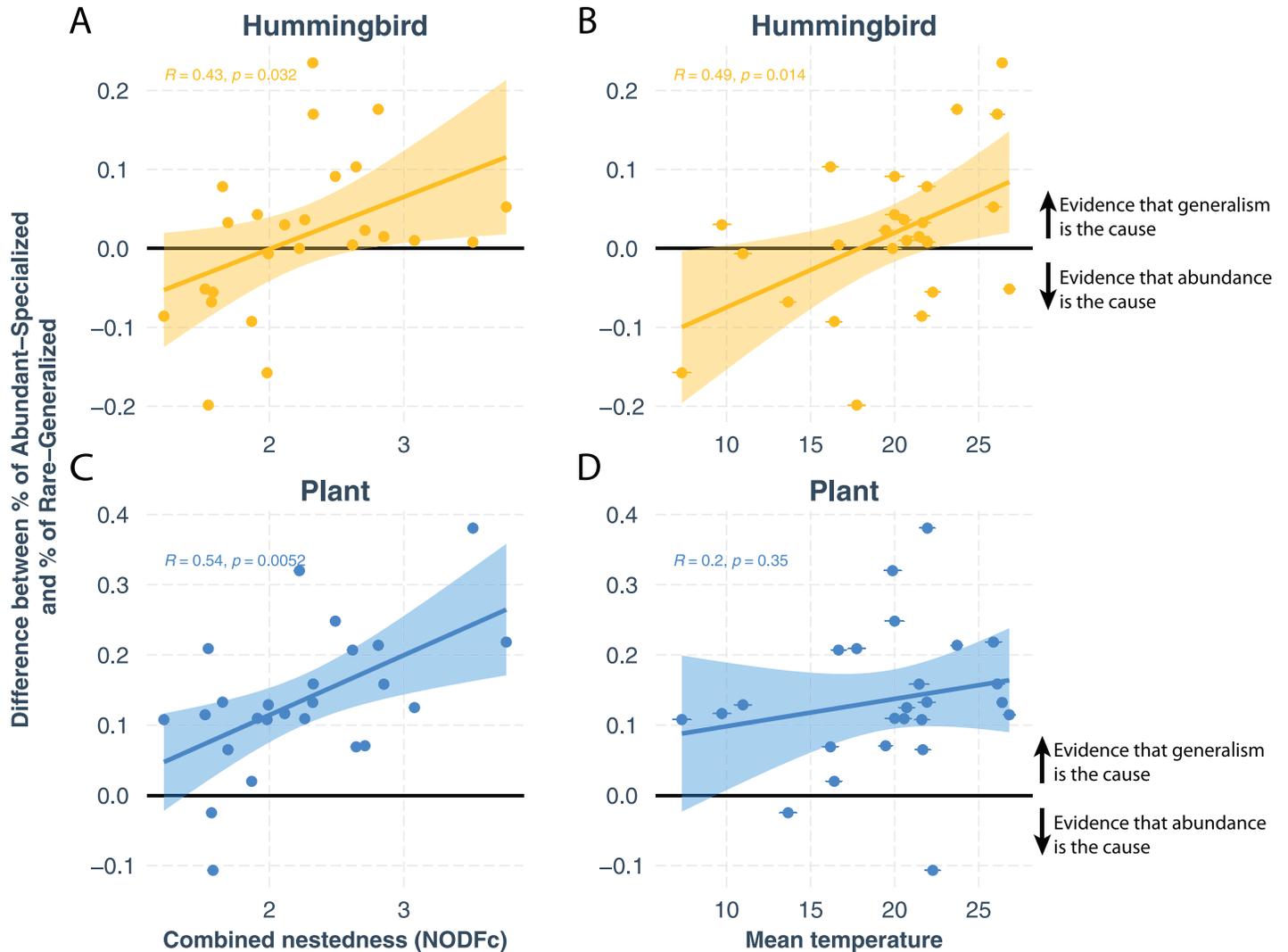


Fig 4. Strength of selection process versus drift process under structural and environmental context. Panels (A) and (C) show the effects of the structural context. The x axis shows the combined nestedness, which is an unbiased metric to compare the level of nestedness across networks [47, 50]. The y axis shows the differences between the proportion of abundant and specialist species and that of rare and generalist species. A positive difference indicates that generalism is the cause of abundance (i.e., stronger selection and weaker drift), while a negative difference indicates that abundance is the cause of generalism (i.e., weaker selection and stronger drift). Each point denotes a different empirical plant-humming bird community (i.e., dataset). The orange points represent the hummingbird data while the blue points represent the plant data. The strength of the selection processes increases as the communities have more nested structures for both plants and hummingbirds. Plants exhibit stronger positive patterns than hummingbirds under the structural context. Panels (B) and (D) show the effects of the environmental context. The x axis shows the local annual temperature mean where the community was sampled. The error bar represents 1 standard deviation with temperature mean from 1958 to 2020 [53]. The strength of the selection processes increases as the communities experience higher mean temperature for both plants and hummingbirds. Hummingbirds exhibit stronger positive patterns than plants under the environmental context. The Pearson correlations and their *p* values are shown in the figure.

<https://doi.org/10.1371/journal.pcbi.1010302.g004>

observed patterns is stronger when the community is exposed to greater environmental variation.

Our results shed light on the big question of selection versus drift processes in structuring multispecies dynamics [26]. Since Hubbell’s groundbreaking work [55], this question has taken a central place in community ecology. In *two-species* communities, a fruitful research line has increased our understanding of this problem by rigorously linking experiments and theory [70–72]. Yet, we lack a full understanding of this question in multispecies communities, because it is challenging to carry out experiments that control species interactions in large

communities [3]. Our computational causal discovery approach provides an alternative, practical path to tackle this problem in multispecies communities. This approach takes advantage of the fact that causal directions are different between abundance and generalism when the community is structured by selection or drift processes. Thus, based on our causal discovery methods, we found strong evidence that selection processes have a critical role in maintaining species persistence in mutualistic systems. Importantly, while previous works have addressed this question [24, 25], we reverse the previously established conclusion that abundance is the cause by fixing a methodological issue.

Causal discovery is not a novel topic, but a computational approach to causal discovery has only taken off in the last decade [27]. Without doubt, the gold standard in casual discovery is controlled experiments [73, 74]. However, controlled experiments can be difficult or even impossible to conduct in many contexts. These contexts create an opportunity for computational tools to detect the correct causal directions. In ecology, the most adopted tool of causal discovery is convergent cross mapping [75]. While this method has been widely applied to many ecological questions [76–80], this method only works for time series data. The field of causal discovery has recently developed a line of methods that work with cross-sectional data, such as the additive noise model and the method of information-geometric inference. These methods have already been applied to many disciplines, including genetics [81], earth system science [82], and kinetic systems [83]. Yet, to the best of our knowledge, these methods have not been used in ecology. We have demonstrated that these methods are useful in ecological contexts. Importantly, these methods are flexible and can be easily adapted to different datasets. As a proof of its flexibility, we have applied the same methods to analyze the dataset of plant-hummingbird communities and the dataset of reef fishes.

We acknowledge that the causal direction between abundance and generalism is not unidirectional and feedback may occur [3]. As the causal direction represents either selection or drift processes, it is unlikely that only one process is in play. As with many debates in ecology, the reality might lie somewhere along the spectrum of the dichotomy. As the strengths of the causal directions indicate the relative roles of the selection and drift process, our results suggest that selection processes are stronger when the network structure of the community has a higher level of nestedness. Our results are consistent with the literature on this topic. At the species level, empirical evidence in reef fishes [42, 84] and corals [85] shows generalism is favored under variable environments. At the community level, empirical evidence on plant-pollinator communities shows that networks are more nested when located in more variable environments [47, 86, 87].

A limitation of our work is that we did not consider potential confounding factors. Our approach can be heuristically justified by the central limit theorem and our removal of the skewness in marginal distributions (i.e., the ordered quantile normalization [59] in the refined method of Fort et al. [24] or post-linear regression [66] in the additive noise model). To explain the emergence of skewed distributions [13], a simple but general argument is based on the central limit theorem: as the mean of the *summation* of many independent or weakly dependent processes result in a normal distribution, the mean of the *product* of these processes results in a log-normal distribution. In this sense, removing the skewness of the distribution transforms the statistical nature of the distribution from the product of multiple processes into the summation of multiple processes (on an appropriate scale). Thus, by removing the skewness of the marginal distribution via the log transformation or the more general semiparametric transformation [59], we can identify the causal direction with more confidence. Of course, this explanation is far from being rigorous, as we may still suffer from the bias of unobserved confounding factors [88], the omitted-variable bias [89], and that the central limit theorem requires asymptotic aggregations (although there is empirical evidence that small

ecological systems can exhibit asymptotic behaviors [90]). A more satisfactory solution is to adopt rigorous methods capable of detecting nonlinear causal directions with the presence of hidden confounding factors. For example, using instrumental variables [88] or the DeCAM-Founder method [91]. These methods generally require a sufficiently large amount of data to ensure statistical convergence, which is unfortunately beyond the reach of the datasets we used. Future research with larger datasets should explore these methods to control for hidden confounding factors.

From a broader perspective, we showcase the power of the computational causal discovery approach in ecological research. In the realm where experiments and theory are difficult to apply to estimate the direction of causality, these computational methods show great promise. Meanwhile, the computational identification of causal association can help refine theoretical assumptions and experimental designs for multispecies communities. The association between abundance and generalism we studied here is by no means an exceptional pattern in community ecology. For example, species abundances are also strongly associated with geographic distributions [92]. The flexibility of these computational methods may be similarly applied to study these patterns in the context where experiments have already been conducted [93]. We hope our work can raise the awareness of this causal discovery approach in the era where much ecological data is becoming available [94, 95].

Supporting information

S1 Data. Detailed methods and additional validations, and supplementary figures and tables.

(PDF)

Acknowledgments

We thank Haoran Cai, Lucas P. Medeiros and Serguei Saavedra for insightful discussions.

Author Contributions

Conceptualization: Chuliang Song, Marie-Josée Fortin, Andrew Gonzalez.

Formal analysis: Chuliang Song.

Funding acquisition: Marie-Josée Fortin, Andrew Gonzalez.

Investigation: Chuliang Song.

Methodology: Chuliang Song, Benno I. Simmons.

Validation: Chuliang Song.

Visualization: Chuliang Song.

Writing – original draft: Chuliang Song.

Writing – review & editing: Chuliang Song, Marie-Josée Fortin, Andrew Gonzalez.

References

1. Pearce J, Ferrier S. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation*. 2001; 98(1):33–43. [https://doi.org/10.1016/S0006-3207\(00\)00139-7](https://doi.org/10.1016/S0006-3207(00)00139-7)
2. Matthews TJ, Whittaker RJ. On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*. 2015; 52(2):443–454. <https://doi.org/10.1111/1365-2664.12380>

3. Dormann CF, Fründ J, Schaefer HM. Identifying causes of patterns in ecological networks: opportunities and limitations. *Annual Review of Ecology, Evolution, and Systematics*. 2017; 48:559–584. <https://doi.org/10.1146/annurev-ecolsys-110316-022928>
4. McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*. 2007; 10(10):995–1015. <https://doi.org/10.1111/j.1461-0248.2007.01094.x> PMID: 17845298
5. Hubbell S. *The unified neutral theory of biodiversity and biogeography*. NJ: Princeton University Press; 2001.
6. Azaele S, Suweis S, Grilli J, Volkov I, Banavar JR, Maritan A. Statistical mechanics of ecological systems: Neutral theory and beyond. *Reviews of Modern Physics*. 2016; 88(3):035003. <https://doi.org/10.1103/RevModPhys.88.035003>
7. MacArthur R. On the relative abundance of species. *The American Naturalist*. 1960; 94(874):25–36. <https://doi.org/10.1086/282106>
8. Sugihara G. Minimal community structure: an explanation of species abundance patterns. *The American Naturalist*. 1980; 116(6):770–787. <https://doi.org/10.1086/283669> PMID: 29513556
9. Sugihara G, Bersier LF, Southwood TRE, Pimm SL, May RM. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proceedings of the National Academy of Sciences*. 2003; 100(9):5246–5251. <https://doi.org/10.1073/pnas.0831096100> PMID: 12702773
10. Tokeshi M. Species abundance patterns and community structure. *Advances in Ecological Research*. 1993; 24:111–186. [https://doi.org/10.1016/S0065-2504\(08\)60042-2](https://doi.org/10.1016/S0065-2504(08)60042-2)
11. Scheffer M, van Nes EH. Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences*. 2006; 103(16):6230–6235. <https://doi.org/10.1073/pnas.0508024103> PMID: 16585519
12. Holt RD. Emergent neutrality. *Trends in Ecology & Evolution*. 2006; 21(10):531–533. <https://doi.org/10.1016/j.tree.2006.08.003> PMID: 16901580
13. May R. Patterns of species abundance and diversity. In: Cody ML, Diamond JM, editors. *Ecology and evolution of communities*. Cambridge, Massachusetts, USA.: Belknap/Harvard University Press; 1975. p. 81–120.
14. Harte J. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford: Oxford University Press; 2011.
15. Šizling AL, Storch D, Šizlingová E, Reif J, Gaston KJ. Species abundance distribution results from a spatial analogy of central limit theorem. *Proceedings of the National Academy of Sciences*. 2009; 106(16):6691–6695. <https://doi.org/10.1073/pnas.0810096106> PMID: 19346488
16. Adler PB, HilleRisLambers J, Levine JM. A niche for neutrality. *Ecology Letters*. 2007; 10(2):95–104. <https://doi.org/10.1111/j.1461-0248.2006.00996.x> PMID: 17257097
17. Yen JD, Thomson JR, Mac Nally R. Is there an ecological basis for species abundance distributions? *Oecologia*. 2013; 171(2):517–525. <https://doi.org/10.1007/s00442-012-2438-1> PMID: 23001621
18. Rael RC, D'Andrea R, Barabás G, Ostling A. Emergent niche structuring leads to increased differences from neutrality in species abundance distributions. *Ecology*. 2018; 99(7):1633–1643. <https://doi.org/10.1002/ecy.2238> PMID: 29655259
19. Waser NM, Chittka L, Price MV, Williams NM, Ollerton J. Generalization in pollination systems, and why it matters. *Ecology*. 1996; 77(4):1043–1060. <https://doi.org/10.2307/2265575>
20. Sexton JP, Montiel J, Shay JE, Stephens MR, Slatyer RA. Evolution of ecological niche breadth. *Annual Review of Ecology, Evolution, and Systematics*. 2017; 48:183–206. <https://doi.org/10.1146/annurev-ecolsys-110316-023003>
21. Fontaine C. Ecology: abundant equals nested. *Nature*. 2013; 500(7463):411. <https://doi.org/10.1038/500411a> PMID: 23969457
22. Santamaría L, Rodríguez-Gironés MA. Linkage rules for plant–pollinator networks: trait complementarity or exploitation barriers? *PLoS Biol*. 2007; 5(2):e31. <https://doi.org/10.1371/journal.pbio.0050031> PMID: 17253905
23. Suweis S, Simini F, Banavar JR, Maritan A. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature*. 2013; 500(7463):449. <https://doi.org/10.1038/nature12438> PMID: 23969462
24. Fort H, Vázquez DP, Lan BL. Abundance and generalisation in mutualistic networks: solving the chicken-and-egg dilemma. *Ecology Letters*. 2016; 19(1):4–11. <https://doi.org/10.1111/ele.12535> PMID: 26498731

25. Simmons BI, Vizenin-Bugoni J, Maruyama PK, Cotton PA, Marín-Gómez OH, Lara C, et al. Abundance drives broad patterns of generalisation in plant–hummingbird pollination networks. *Oikos*. 2019; 128(9):1287–1295. <https://doi.org/10.1111/oik.06104>
26. Vellend M. *The theory of ecological communities*. NJ: Princeton University Press; 2016.
27. Peters J, Janzing D, Schölkopf B. *Elements of causal inference: foundations and learning algorithms*. The MIT Press; 2017.
28. Butsic V, Lewis DJ, Radeloff VC, Baumann M, Kuemmerle T. Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic and Applied Ecology*. 2017; 19:1–10. <https://doi.org/10.1016/j.baae.2017.01.005>
29. MacDonald AJ, Larsen AE, Plantinga AJ. Missing the people for the trees: Identifying coupled natural–human system feedbacks driving the ecology of Lyme disease. *Journal of Applied Ecology*. 2019; 56(2):354–364. <https://doi.org/10.1111/1365-2664.13289>
30. Larsen AE, Meng K, Kendall BE. Causal analysis in control–impact ecological studies with observational data. *Methods in Ecology and Evolution*. 2019; 10(7):924–934. <https://doi.org/10.1111/2041-210X.13190>
31. MacDonald AJ, Mordecai EA. Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing. *Proceedings of the National Academy of Sciences*. 2019; 116(44):22212–22218. <https://doi.org/10.1073/pnas.1905315116> PMID: 31611369
32. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning. Cambridge: MIT Press; 2001.
33. Pearl J. *Causality*. Cambridge: Cambridge University Press; 2009.
34. Spirtes P, Zhang K. Causal discovery and inference: concepts and recent methodological advances. In: *Applied informatics*. vol. 3. SpringerOpen; 2016. p. 1–28.
35. Robins JM, Scheines R, Spirtes P, Wasserman L. Uniform consistency in causal inference. *Biometrika*. 2003; 90(3):491–515. <https://doi.org/10.1093/biomet/90.3.491>
36. Peters JM. *Restricted structural equation models for causal inference*. ETH Zurich; 2012.
37. Mooij JM, Peters J, Janzing D, Zscheischler J, Schölkopf B. Distinguishing cause from effect using observational data: methods and benchmarks. *The Journal of Machine Learning Research*. 2016; 17(1):1103–1204.
38. Daniušis P, Janzing D, Mooij J, Zscheischler J, Steudel B, Zhang K, et al. Inferring deterministic causal relations. In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*; 2010. p. 143–150.
39. Janzing D, Mooij J, Zhang K, Lemeire J, Zscheischler J, Daniušis P, et al. Information-geometric approach to inferring causal directions. *Artificial Intelligence*. 2012; 182:1–31. <https://doi.org/10.1016/j.artint.2012.01.002>
40. Dalsgaard B, Maruyama PK, Sonne J, Hansen K, Zanata TB, Abrahamczyk S, et al. The influence of biogeographical and evolutionary histories on morphological trait-matching and resource specialization in mutualistic hummingbird–plant networks. *Functional Ecology*. 2021; 35(5):1120–1133. <https://doi.org/10.1111/1365-2435.13784>
41. Edgar GJ, Stuart-Smith RD. Systematic global assessment of reef fish communities by the Reef Life Survey program. *Scientific Data*. 2014; 1(1):1–8. <https://doi.org/10.1038/sdata.2014.7> PMID: 25977765
42. Stuart-Smith RD, Mellin C, Bates AE, Edgar GJ. Habitat loss and range shifts contribute to ecological generalization among reef fishes. *Nature Ecology & Evolution*. 2021; 5(5):656–662. <https://doi.org/10.1038/s41559-020-01342-7> PMID: 33686182
43. Simmons BI, Vizenin-Bugoni J, Maruyama PK, Cotton PA, Marín-Gómez OH, Lara C, et al. Data from: Abundance drives broad patterns of generalisation in plant–hummingbird pollination networks; 2019.
44. Schupp EW, Jordano P, Gómez JM. A general framework for effectiveness concepts in mutualisms. *Ecology Letters*. 2017; 20:577–590. <https://doi.org/10.1111/ele.12764> PMID: 28349589
45. Valdovinos FS. Mutualistic networks: moving closer to a predictive theory. *Ecology Letters*. 2019; 22(9):1517–1534. <https://doi.org/10.1111/ele.13279> PMID: 31243858
46. Bascompte J, Jordano P. *Mutualistic Networks*. NJ: Princeton University Press; 2013.
47. Song C, Rohr RP, Saavedra S. Why are some plant–pollinator networks more nested than others? *Journal of Animal Ecology*. 2017; 86(6):1417–1424. <https://doi.org/10.1111/1365-2656.12749> PMID: 28833083
48. Song C, Rohr RP, Saavedra S. Beware z-scores. *Journal of Animal Ecology*. 2019; 88(5):808–809. <https://doi.org/10.1111/1365-2656.12964> PMID: 30874304
49. Simmons BI, Hoeppecke C, Sutherland WJ. Beware greedy algorithms. *Journal of Animal Ecology*. 2019; 88(5):804–807. <https://doi.org/10.1111/1365-2656.12963> PMID: 30874298

50. Hoeppeke C, Simmons BI. Maxnodf: an R package for fair and fast comparisons of nestedness between networks. *BioRxiv*. 2020;.
51. Newman ME. Modularity and community structure in networks. *Proceedings of the national academy of sciences*. 2006; 103(23):8577–8582. <https://doi.org/10.1073/pnas.0601602103>
52. Simmons BI, Cirtwill AR, Baker NJ, Wauchope HS, Dicks LV, Stouffer DB, et al. Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*. 2019; 128(2):154–170. <https://doi.org/10.1111/oik.05670>
53. Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*. 2018; 5(1):1–12. <https://doi.org/10.1038/sdata.2017.191> PMID: 29313841
54. Gotelli NJ, Graves GR. *Null models in ecology*. Smithsonian; 1996.
55. Hubbell SP. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs*. 1997; 16(5):S9–S21. <https://doi.org/10.1007/s003380050237>
56. Jordano P, Bascompte J, Olesen JM. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters*. 2003; 6(1):69–81. <https://doi.org/10.1046/j.1461-0248.2003.00403.x>
57. Vázquez DP. Degree distribution in plant–animal mutualistic networks: forbidden links or random interactions? *Oikos*. 2005; 108(2):421–426. <https://doi.org/10.1111/j.0030-1299.2005.13619.x>
58. Payrató-Borras C, Hernández L, Moreno Y. Breaking the spell of nestedness: The entropic origin of nestedness in mutualistic systems. *Physical Review X*. 2019; 9(3):031024. <https://doi.org/10.1103/PhysRevX.9.031024>
59. Peterson RA, Cavanaugh JE. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*. 2019; 47(13-15):2312–2327. <https://doi.org/10.1080/02664763.2019.1630372> PMID: 35707424
60. Sugihara G, May RM. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*. 1990; 344(6268):734. <https://doi.org/10.1038/344734a0> PMID: 2330029
61. Coulson T, Rohani P, Pascual M. Skeletons, noise and population growth: the end of an old debate? *Trends in Ecology & Evolution*. 2004; 19(7):359–364. <https://doi.org/10.1016/j.tree.2004.05.008> PMID: 16701286
62. Mason NW, Holdaway RJ, Richardson SJ. Incorporating measurement error in testing for changes in biodiversity. *Methods in Ecology and Evolution*. 2018; 9(5):1296–1307. <https://doi.org/10.1111/2041-210X.12976>
63. Higgins K, Hastings A, Sarvela JN, Botsford LW. Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab. *Science*. 1997; 276(5317):1431–1435. <https://doi.org/10.1126/science.276.5317.1431>
64. Mutshinda CM, O'Hara RB, Woiwod IP. What drives community dynamics? *Proceedings of the Royal Society B: Biological Sciences*. 2009; 276(1669):2923–2929. <https://doi.org/10.1098/rspb.2009.0523> PMID: 19457887
65. Wood SN. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC; 2017.
66. Zhang K, Hyvarinen A. On the identifiability of the post-nonlinear causal model. *arXiv preprint arXiv:12052599*. 2012;.
67. Gretton A, Bousquet O, Smola A, Schölkopf B. Measuring statistical dependence with Hilbert-Schmidt norms. In: *International conference on algorithmic learning theory*. Springer; 2005. p. 63–77.
68. Pfister N, Bühlmann P, Schölkopf B, Peters J. Kernel-based tests for joint independence. *Journal of the Royal Statistical Society Series B*. 2018; 80(1):5–31. <https://doi.org/10.1111/rssb.12235>
69. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Physical Review E*. 2004; 69(6):066138. <https://doi.org/10.1103/PhysRevE.69.066138>
70. Levine JM, HilleRisLambers J. The importance of niches for the maintenance of species diversity. *Nature*. 2009; 461(7261):254–257. <https://doi.org/10.1038/nature08251> PMID: 19675568
71. Jeraldo P, Sipo M, Chia N, Brulc JM, Dhillon AS, Konkel ME, et al. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proceedings of the National Academy of Sciences*. 2012; 109(25):9692–9698. <https://doi.org/10.1073/pnas.1206721109> PMID: 22615407
72. Godoy O, Kraft NJ, Levine JM. Phylogenetic relatedness and the determinants of competitive outcomes. *Ecology Letters*. 2014; 17(7):836–844. <https://doi.org/10.1111/ele.12289> PMID: 24766326
73. Bascompte J. Structure and dynamics of ecological networks. *Science*. 2010; 329(5993):765–766. <https://doi.org/10.1126/science.1194255> PMID: 20705836

74. Hernan MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.; 2020.
75. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, et al. Detecting causality in complex ecosystems. *Science*. 2012; 338(6106):496–500. <https://doi.org/10.1126/science.1227079> PMID: 22997134
76. Ushio M, Hsieh Ch, Masuda R, Deyle ER, Ye H, Chang CW, et al. Fluctuating interaction network and time-varying stability of a natural fish community. *Nature*. 2018; 554:360. <https://doi.org/10.1038/nature25504> PMID: 29414940
77. Karakoç C, Clark AT, Chatzinotas A. Diversity and coexistence are influenced by time-dependent species interactions in a predator–prey system. *Ecology Letters*. 2020; 23(6):983–993. <https://doi.org/10.1111/ele.13500> PMID: 32243074
78. Bray SR, Wang B. Forecasting unprecedented ecological fluctuations. *PLoS Computational Biology*. 2020; 16(6):e1008021. <https://doi.org/10.1371/journal.pcbi.1008021> PMID: 32598364
79. Kitayama K, Ushio M, Aiba SI. Temperature is a dominant driver of distinct annual seasonality of leaf litter production of equatorial tropical rain forests. *Journal of Ecology*. 2021; 109(2):727–736. <https://doi.org/10.1111/1365-2745.13500>
80. Nova N, Deyle ER, Shocket MS, MacDonald AJ, Childs ML, Rypdal M, et al. Susceptible host availability modulates climate effects on dengue dynamics. *Ecology Letters*. 2021; 24(3):415–425. <https://doi.org/10.1111/ele.13652> PMID: 33300663
81. Meinshausen N, Hauser A, Mooij JM, Peters J, Versteeg P, Bühlmann P. Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences*. 2016; 113(27):7361–7368. <https://doi.org/10.1073/pnas.1510493113> PMID: 27382150
82. Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, et al. Inferring causation from time series in Earth system sciences. *Nature Communications*. 2019; 10(1):1–13. <https://doi.org/10.1038/s41467-019-10105-3> PMID: 31201306
83. Pfister N, Bauer S, Peters J. Learning stable and predictive structures in kinetic systems. *Proceedings of the National Academy of Sciences*. 2019; 116(51):25405–25411. <https://doi.org/10.1073/pnas.1905688116> PMID: 31776252
84. Beger M. Accepting the loss of habitat specialists in a changing world. *Nature Ecology & Evolution*. 2021; in press. <https://doi.org/10.1038/s41559-021-01402-6>
85. Baker AC. Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of Symbiodinium. *Annual Review of Ecology, Evolution, and Systematics*. 2003; 34(1):661–689. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132417>
86. Fontenla JL, Fontenla Y, Cuervo Z, Álvarez de Zayas A. Red de interacción ecológica insectos-plantas en Playas del Este, la Habana, Cuba. *Acta Botánica Cubana*. 2019; 218.
87. Classen A, Eardley CD, Hemp A, Peters MK, Peters RS, Ssymank A, et al. Specialization of plant–pollinator interactions increases with temperature at Mt. Kilimanjaro. *Ecology and Evolution*. 2020; 10(4):2182–2195. <https://doi.org/10.1002/ece3.6056> PMID: 32128148
88. Kendall BE. A statistical symphony: instrumental variables reveal causality and control measurement error. In: Fox GA, Negrete-Yankelevich S, Sosa VJ, editors. *Ecological Statistics: Contemporary Theory and Application*. Oxford: Oxford University Press; 2015. p. 149–167.
89. Angrist JD, Pischke JS. *Mostly harmless econometrics*. NJ: Princeton University Press; 2008.
90. Barbier M, de Mazancourt C, Loreau M, Bunin G. Fingerprints of high-dimensional coexistence in complex ecosystems. *Physical Review X*. 2021; 11(1):011009. <https://doi.org/10.1103/PhysRevX.11.011009>
91. Agrawal R, Squires C, Prasad N, Uhler C. The DeCAMFounder: Non-Linear Causal Discovery in the Presence of Hidden Variables. *arXiv preprint arXiv:210207921*. 2021;.
92. Borregaard MK, Rahbek C. Causality of the relationship between geographic distribution and species abundance. *The Quarterly Review of Biology*. 2010; 85(1):3–25. <https://doi.org/10.1086/650265> PMID: 20337258
93. Gonzalez A, Lawton JH, Gilbert F, Blackburn TM, Evans-Freke I. Metapopulation dynamics, abundance, and distribution in a microecosystem. *Science*. 1998; 281(5385):2045–2047. <https://doi.org/10.1126/science.281.5385.2045> PMID: 9748167
94. Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. *Frontiers in Ecology and the Environment*. 2013; 11(3):156–162. <https://doi.org/10.1890/120103>
95. Poisot T, Gravel D, Leroux S, Wood SA, Fortin MJ, Baiser B, et al. Synthetic datasets and community tools for the rapid testing of ecological hypotheses. *Ecography*. 2016; 39(4):402–408. <https://doi.org/10.1111/ecog.01941>