Title (8 words)

When indices disagree: facing conceptual and practical challenges

Authors:

Carlos Alberto Arnillas†
Department of Physical and Environmental Sciences, University of Toronto — Scarborough
1265 Military Trail, Toronto ON, Canada, M1C 1A4

Kelly Carscadden
Department of Ecology and Evolutionary Biology, University of Colorado Boulder
1900 Pleasant St, Box 334, Boulder CO, USA, 80309

† Corresponding author

Keywords:

metrics, measurement theory, hypothesis testing, database of assumptions, niche breadth

Abstract (100-120 words)

Hypothesis testing requires meaningful ways to quantify relevant biological phenomena and account for alternative mechanisms that could explain the same pattern. Researchers combine experiments, statistics, and indices to account for these confounding mechanisms. Key concepts in ecology and evolution, like niche breadth or fitness, can be represented by several indices, which often provide uncorrelated estimates. Is this because the indices use different types of noisy data or because the targeted phenomenon is complex and multidimensional? We discuss implications of these scenarios and propose five steps to aid researchers in identifying and combining indices, experiments, and statistics. Supported by efforts to build databases of hypotheses and indices and document assumptions, these steps help provide a formal strategy to reduce self-confirmatory bias.

HIGHLIGHTS (900 characters)

1. Many hypotheses in ecology and evolution rely on indices that capture abstract phenomena, which can be multi-faceted or noisy (e.g., niche breadth indices may capture tolerance or preference using variable abundance data). Hence, the estimates are often uncorrelated.
2. The indices, along with experiments and statistical methods, help to control for confounding factors. This variety of approaches limits cross-study comparisons, enabling confirmation bias.
3. We formalize guidelines to link hypotheses, study design, and index selection to help mitigate these challenges, using as an example the relationship between niche breadth and geographic range size. Repositories of hypotheses and indices, building on existing

collaborative tools and databases, could help researchers navigate murky methodological decision points in hypothesis testing.

When indices disagree: facing conceptual and practical challenges

Indices and hypothesis testing

> "For measurements to be meaningful…they must retain their connection to the theoretical and instrumental context from which they were derived" - Houle et al. 2011

Hypothesis testing is the cornerstone of ecology, evolution, and scientific endeavours more broadly, and it requires two critical decisions: (i) how to quantify the biological phenomenon of interest, and (ii) how to handle confounding factors (i.e., alternative mechanisms that can drive the response variable, (cf. Catford et al. 2021)). Indices —numerical estimates that summarize data, aiming to capture critical aspects of the phenomenon of interest (box 1)— help to solve the first question and —along with experiments and statistical tools— are also used to control for confounding factors. Hence, indices play a central role in theory development and testing in ecology and evolution. However, the phenomena in these branches of science are often multi-faceted, sometimes ill-defined, and are numerically estimated by a wide variety of indices. For instance, we identified more than 50 indices of niche breadth (see box 1) which can generate uncorrelated estimates (Carscadden et al. unpublished), and similar index proliferation has occurred in fields as diverse as habitat fragmentation (McGarigal and Marks 1995) and phylogenetic diversity (Tucker et al. 2016).

The variety of options for both decision points (measuring focal phenomena and handling confounding factors) makes it unclear how to compare different studies and, in extreme scenarios, we run the risk of self-confirmatory bias by inadvertently selecting indices and approaches whose results match our intuition. Here, we illustrate key considerations in each decision point and outline guidelines linking index selection to hypothesis testing, using the hypothesis that niche breadth (NB) drives range size to frame our discussion. We also discuss how a more nuanced understanding of indices may shed some light on the phenomenon itself.

Decision point 1: Quantifying a complex biological phenomenon

The proliferation of NB indices stems from the many ways NB has been defined and the numerous biological properties it has been invoked to explain (reviewed in Carscadden et al.

2020). For example, NB has been defined as resource impacts and growth isoclines in competition (Chase and Leibold 2003), as presence across climatic zones (Warren and Seifert 2010), and as dietary breadth in food choice tests (Jorge et al. 2014). Further, it has been studied as a driver of diverse patterns and processes, from geographic range size (Hirst et al. 2017) to ecosystem function and stability (Endres et al. 2021) to evolution patterns (Visher and Boots 2020).

A diversity of indices can be advantageous because the same property (such as NB, fitness, or diversity) will need to be calculated in different ways in different organisms, test different responses (e.g., thermal vs. trophic niche breadth), deal with continuous or categorical data, or improve statistical properties (e.g., Fridley et al. 2007, Manthey and Fridley 2009, Zeleny 2009). In this case, we expect indices to be correlated, e.g., generate similar rankings of, say, plant species by thermal NB. Indices have also diversified over time as new factors are deemed important and computational complexity increases. For instance, indices of dietary NB began as counts of species eaten (Fig. 1a) and then eventually incorporated preference, measured as variation in the numbers of each prey species consumed (Fig. 1b) (Levins 1968). However, these indices implicitly treated all prey species as equally abundant, and therefore could not determine if variation in consumption rates were due to differences in prey availability or dietary preferences. Subsequently, more comprehensive indices — like Czekanowski's proportional similarity proposed by Feinsinger et al. (1981)— accounted for variation in prey availability (Fig. 1c). Later, Dolédec et al. (2000) proposed the Outlying Mean Index (OMI) that estimates NB based on the similarity of used and available habitat. The OMI index is often associated with abiotic niche axes (e.g., average temperature, salinity), and accounts for the span of the resources used by focal species, the resource usage frequency by the focal species, the relative abundance of the resource, and the similarity of the used resources to the most abundant resources (Fig. 1d). These examples show that when new indices are developed and published, it is often because they aim to capture an aspect of the phenomenon of interest that was not included in previous indices, and which yields novel insights and results likely uncorrelated with previous indices.

How should we interpret the correlation (or lack of it) of estimates generated by different indices? And how should we select among them? If all indices are measuring the same biological phenomenon, they should provide correlated values when applied to the same situation. In this case, only the simpler and most statistically robust indices could be needed. But what if indices yield dissimilar rankings? One possibility is that the measurement error of each NB index is very high. If we then think about NB as a latent variable, and each NB index as providing some information about NB and a large amount of error, then, incorporating estimates from multiple indices may actually improve our understanding of NB, even if the rankings generated by the indices were poorly correlated (Fig 2a) (Shipley 2016). Another possibility is that the measurement error of each index is almost zero. Then, the poorly correlated estimates suggest that the phenomenon is complex and could have distinct internal dimensions (Fig 2b). In the case of NB, we can distinguish between tolerance (what can be eaten), preference (what is more likely to be eaten), and performance (e.g., nutritious value of the of the prey for the predator), for instance (Fig 1d). Similarly, fitness is expected to be positively correlated with survival, reproduction, and growth, but there are strong trade-offs between them, which means that the four variables cannot be positively correlated at the same time (Laughlin et al. 2020). We posit that the underlying assumption that different indices are ultimately measuring the same biological phenomenon is rarely tested.

Mapping hypotheses to indices may help clarify which facets of multidimensional phenomena (e.g., NB, fitness) are most relevant for a given theory. Because every index specifies the mathematical relationships between variables, they carry fixed assumptions about how biological entities or properties interact in nature. Evaluating the built-in properties of indices in light of focal hypotheses can give rise to guidelines for index selection and help stave off the potential for self-confirmatory bias.

Decision point 2: Navigating confounding factors in hypothesis testing

Confounding factors are variables that correlate with predictors and the response and can represent alternative mechanisms not considered in the initial hypothesis (Laubach et al. 2021). Accounting for confounding factors in any experiment or observational study is a considerable challenge (Grace and Irvine 2020, Catford et al. 2021, Kimmel et al. 2021). For example, like other complex biological phenomena, NB can act as a predictor or response in different hypotheses, as in the connection between NB and geographic range size (RS, Box 1). On the one hand, it has been hypothesized that when a species has a narrow fundamental niche breadth (**NB_f**), it will be geographically restricted because it can only tolerate a narrow range of environmental conditions (Brown 1984). This hypothesis, hereafter **NB_f-to-RS**, has been generally supported. However, it is also possible that a more widespread species will be exposed to more environmental conditions, resulting in a broader realized NB (**RS-to-NB_r**). Over time, a species that encounters a wide range of environmental conditions could also evolve a wider fundamental NB than a species constrained in a small area (**RS-to-NB_f**). These three hypotheses operate at different timescales, implying different assumptions and eventually different confounding factors. Given these hypotheses' different assumptions (Table B1), it may not be adequate to use the same index and data to test the **NB_f-to-RS** and **RS-to-NB_f** hypotheses. Therefore, it is up to the researcher to determine the best way to account for the different confounding factors, either using experiments, statistical controls, or comprehensive indices (Houle et al. 2011).

Experiments are considered the gold standard for testing hypotheses in biology, as they can confirm causation by directly manipulating alternative causes (Shipley 2016). However, the generality they provide is constrained: only a few species, conditions, and mechanisms can be tested, and important processes may be excluded in simplified experimental settings. Some hypotheses, like NB_f-to-RS, are particularly challenging to test because of the large spatial or temporal scales involved. Given these limitations, experimental approaches require a thoughtful framing of predictions. For example, Hirst et al. (2017) developed seven predictions derived from the NB_f-to-RS hypothesis that could be experimentally tested in an alpine daisy system (e.g., wide-ranging species should perform better under novel conditions). Each experiment represents a sub-hypothesis applied to a particular criterion (e.g., survival, growth, germination, viability, of seeds and seedlings). They directly compared the performance of each species under different conditions, ranking the species' NB_f by the number of environmental scenarios in which a species survived, and whether it outperformed other species or not. More generally, a platonic 'ideal' experiment would manipulate each potential confounding factor to test if the predictor (here, NB_f), and no other process, drives the response.

Statistics provide several tools to address confounding mechanisms. For instance, phylogenetically independent contrasts can account for variation in NB or RS that may be explained by phylogenetic relatedness (Brandle et al. 2002). Another way identify the impact of a confounding mechanism is by randomizing the observed data using rules consistent with the expected effect of that mechanism, adding rules until randomized data matches the observed pattern (Jorge et al. 2014). Using this approach, Boucher-Lalonde & Currie (2016) showed that the observed positive correlation between NB_r and RS can be explained by the underlying spatial auto-correlation structures of temperature and precipitation. Structural equation modelling (SEM) compares the correlation structure predicted under a given causal hypothesis with the observed correlations (Shipley 2016). Using SEM, Sheth et al. (2014) demonstrated a mechanism underlying the NB-to-RS relationship in monkeyflowers: wide niche breadth increased the proportion of habitat that is suitable, which in turn increased the species' geographic range (no direct connection between NB and RS was detected). SEM's theoretical framework also provides insights about measurement of unobserved variables and unintended consequences of common correlational analyses. For example, the correlation of two variables that reciprocally affect each other, like NB and RS, cannot provide an unbiased estimate of the effect of one on the other. Given observational data, an ideal statistical approach should allow distinguishing between alternative causal mechanisms.

Indices that integrate multiple aspects of a focal phenomenon ('comprehensive indices' hereafter) provide a third strategy to deal with confounding factors. This is partially because —as previously discussed— many indices have been proposed to address a potential bias caused by properties that previous indices did not consider. For instance, using the OMI index, Heino & Soininen found partial support for the NB_f-to-RS hypothesis (Table B1) in unicellular eukaryotes in streams; however, species' distributions were best predicted by availability of adequate habitat (a potential confounding factor), rather than NB (Heino and Soininen 2006). In essence: a hypothetical 'complete' comprehensive index could account for every potential confounding factor for any NB-related hypothesis.

These three strategies are widely used and combined, often in a single study (e.g., (Hirst et al. 2017)): experiments provide ultimate proof of causality; statistical controls allow generality using widespread raw data when experiments are unfeasible; and robust comprehensive indices enable comparability among studies. These three quantitative strategies are also combined with literature review, which can help to account for potential confounding factors under the scenario that relevant information (either theoretical or empirical) already exists. Of these strategies, we posit that selecting among comprehensive indices is becoming a larger challenge over time: indices have proliferated, increases in computer power and data availability have made indices easy to estimate (in contrast to experiments), and indices are assumed to provide some intrinsic insight into biological processes (in contrast to statistical approaches).

A way forward: Mapping assumptions of hypotheses and indices

Given the variation among NB indices, we tried to identify scenarios in which one index could be better than another. We found no general guidelines in the way that indices are used to test hypotheses in the literature because indices, experiments, and statistical analyses are combined in heterogeneous ways. For instance, there is no need to include the abundance of preys in a comprehensive index (see Fig 1c) if prey abundance is standardized in an experiment (Basset and Rossi 1987), but it would be inappropriate to infer niche breadth in natural conditions without including prey abundance in the analysis (Feinsinger et al. 1981). These examples show that several indices potentially generating uncorrelated estimates can be adequately used for the same purpose if potential confounding factors are adequately addressed in the study design.

Identifying confounding factors and integrating them into ecological and evolutionary studies is not a new task, but it has received recent interest in the eco-evo research community (Shipley 2016, Grace and Irvine 2020, Laubach et al. 2021, Catford et al. 2021). For instance, Gerhold et al. (2015) identified seven assumptions (and associated confounding factors) underlying the hypothesis that environmental filtering leads to phylogenetic clustering in communities while competition leads to communities of more distantly related species. Similarly, the ‘Hierarchy-of-Hypotheses’ is a tool that has been used to identify critical assumptions underlying theories such as the enemy release hypothesis and the escalation hypothesis of evolution (Jeschke et al. 2012, Heger et al. 2021). These examples demonstrate an organized effort to formally identify confounding factors in different eco-evo subdisciplines.

Building on these examples (and on widely successful efforts to standardize and share data on organismal traits, locations, phylogenies, and other biological information), we believe that it is possible to generate databases of hypotheses and their underlying assumptions, and of indices and the factors that they consider (Fig 3). Such databases could support a framework to compare the assumptions and factors captured in any index and those of the hypothesis being tested. Focusing on a NB hypothesis, these steps include (1) identifying the hypothesis' underlying assumptions; (2) identifying the assumptions captured by the study design; (3) identifying the set of NB indices that can be estimated with the data available; (4) assessing if all of the hypothesis' assumptions are controlled by an index (Fig. 1d) or study design, adjusting the study design, data collection, or index selection, and repeating the previous two steps; and (5) explicitly acknowledging any gaps or alternative hypothesis that could also explain the observed results. Here, the steps (1) and (3) will be mostly derived from the databases of hypotheses (as in Table B1) and indices (as in Fig. 1d). Using standardized language (e.g., consistent definitions and terminology), the hypotheses and index repositories could be linked, showing the gaps that need to be addressed (step 4). Statistical considerations (like the expected theoretical distribution of the index) will help to further refine the search for optimal indices, building on top of the biological meaning provided by the formal connection between hypothesis, research design, and indices.

Concluding remarks

Despite their intuitive appeal and simplicity, many key properties in ecology and evolutionary biology (e.g., NB, fitness, diversity) are abstract concepts that could represent multi-dimensional phenomena. We can strengthen our understanding of these abstract concepts by formalizing the assumptions of the hypotheses that rely on them. Databases linking hypotheses, indices, and assumptions can be built leveraging community science, projects in graduate courses, cross-validation techniques, and other strategies. Researchers newly entering a field of study will benefit the most from a tool that can support critical thinking on the connections between hypotheses, indices, and research design. However, we also believe that structured approaches for navigating key methodological decision points —like the one presented here— could help reduce self-confirmatory bias and, even more important, some of the context dependency that researchers often find in ecology and evolution.

Outstanding questions (2000 characters)

Besides niche breadth, diversity, fitness, and fragmentation, which other concepts in ecology or evolution suffer from a large number of indices that are often uncorrelated?

The history of index development shows that critical assumptions and limitations are often missed when hypotheses and concepts are first proposed. How to make revisions of published results more dynamic and flexible?

How to improve tools that can help students and researchers identify potential confounding factors when testing hypothesis?

Like fitness, niche breadth (NB) is a complex phenomenon. NB is composed of tolerance to several conditions, the preference for some of them, and the performance that the species obtains when using them, but it is not well explored if there are positive or negative correlations among them as in the case of reproductive effort and biomass accumulation for fitness. Can indices shed light on the internal structure of complex phenomena in ecology and evolution (e.g., identify trade-offs between different aspects of the phenomena)?

How often the context dependency reported in ecology and evolution could be triggered by differences in the aspect of a complex concept being measured?

Acknowledgements

The ideas presented here were inspired by conversations with Nancy Emery and Michelle E. Afkhami. We would also like to acknowledge the comments provided by George Arhonditsis, Nancy Emery, and Marc Cadotte on early versions of the manuscript.

References

Basset, A., and L. Rossi. 1987. Relationships between trophic niche breadth and reproductive capabilities in a population of Proasellus coxalis Dollfus (Crustacea: Isopoda). Functional Ecology 1:13–18.

Boucher-Lalonde, V., and D. J. Currie. 2016. Spatial Autocorrelation Can Generate Stronger Correlations between Range Size and Climatic Niches Than the Biological Signal — A Demonstration Using Bird and Mammal Range Maps. PLOS ONE 11:e0166243.

Brandle, M., A. Prinzing, R. Pfeifer, and R. Brandl. 2002. Dietary niche breadth for Central European birds: correlations with species-specific traits. Evolutionary Ecology Research 4:643–657.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. American naturalist:255–279.

Carscadden, K. A., N. C. Emery, C. A. Arnillas, M. W. Cadotte, M. E. Afkhami, D. Gravel, S. W. Livingstone, and J. J. Wiens. 2020. Niche Breadth: Causes and Consequences for Ecology, Evolution, and Conservation. The Quarterly Review of Biology 95:179–214.

Catford, J. A., J. R. U. Wilson, P. Pyšek, P. E. Hulme, and R. P. Duncan. 2021. Addressing context dependence in ecology. Trends in Ecology & Evolution:S0169534721002585.

Chase, J. M., and M. A. Leibold. 2003. Ecological niches: linking classical and contemporary approaches. University of Chicago Press.

Dolédec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. Ecology 81:2914–2927.

Endres, K. L., C. N. Morozumi, X. Loy, H. M. Briggs, P. J. CaraDonna, A. M. Iler, D. A. Picklum, W. A. Barr, and B. J. Brosi. 2021. Plant–pollinator interaction niche broadens in response to severe drought perturbations. Oecologia 197:577–588.

Feinsinger, P., E. E. Spears, and R. W. Poole. 1981. A Simple Measure of Niche Breadth. Ecology 62:27.

Fridley, J. D., D. B. Vandermast, D. M. Kuppinger, M. Manthey, and R. K. Peet. 2007. Co-occurrence based assessment of habitat generalists and specialists: a new approach for the measurement of niche width. Journal of Ecology 95:707–722.

Gerhold, P., J. F. Cahill, M. Winter, I. V. Bartish, and A. Prinzing. 2015. Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). Functional Ecology 29:600–614.

Grace, J. B., and K. M. Irvine. 2020. Scientist's guide to developing explanatory statistical models using causal analysis principles. Ecology 101.

Heger, T., C. A. Aguilar-Trigueros, I. Bartram, R. R. Braga, G. P. Dietl, M. Enders, D. J. Gibson, L. Gómez-Aparicio, P. Gras, K. Jax, S. Lokatis, C. J. Lortie, A.-C. Mupepele, S. Schindler, J. Starrfelt, A. D. Synodinos, and J. M. Jeschke. 2021. The Hierarchy-of-Hypotheses Approach: A Synthesis Method for Enhancing Theory Development in Ecology and Evolution. BioScience 71:337–349.

Heino, J., and J. Soininen. 2006. Regional occupancy in unicellular eukaryotes: a reflection of niche breadth, habitat availability or size-related dispersal capacity? Freshwater Biology 51:672–685.

Hirst, M. J., P. C. Griffin, J. P. Sexton, and A. A. Hoffmann. 2017. Testing the niche-breadth-range-size hypothesis: habitat specialization vs. performance in Australian alpine daisies. Ecology 98:2708–2724.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and Meaning in Biology. The Quarterly Review of Biology 86:3–34.

Hutchinson, G. E. 1978. An introduction to population ecology.

Izem, R., and J. G. Kingsolver. 2005. Variation in Continuous Reaction Norms: Quantifying Directions of Biological Interest. The American Naturalist 166:277–289.

Jeschke, J., L. Gómez Aparicio, S. Haider, T. Heger, C. Lortie, P. Pyšek, and D. Strayer. 2012. Support for major hypotheses in invasion biology is uneven and declining. NeoBiota 14:1–20.

Jorge, L. R., P. I. Prado, M. Almeida-Neto, and T. M. Lewinsohn. 2014. An integrated framework to improve the concept of resource specialisation. Ecology Letters 17:1341–1350.

Junker, R. R., N. Bluethgen, T. Brehm, J. Binkenstein, J. Paulus, H. M. Schaefer, and M. Stang. 2013. Specialization on traits as basis for the niche-breadth of flower visitors and as structuring mechanism of ecological networks. Functional Ecology 27:329–341.

Kimmel, K., L. E. Dee, M. L. Avolio, and P. J. Ferraro. 2021. Causal assumptions and causal inference in ecological experiments. Trends in Ecology & Evolution 36:1141–1152.

Laubach, Z. M., E. J. Murray, K. L. Hoke, R. J. Safran, and W. Perng. 2021. A biologist's guide to model selection and causal inference. Proceedings of the Royal Society B: Biological Sciences 288:20202815.

Laughlin, D. C., J. R. Gremer, P. B. Adler, R. M. Mitchell, and M. M. Moore. 2020. The Net Effect of Functional Traits on Fitness. Trends in Ecology & Evolution 35:1037–1047.

Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, NJ.

Manthey, M., and J. D. Fridley. 2009. Beta diversity metrics and the estimation of niche width via species co-occurrence data: reply to Zeleny. Journal of Ecology 97:18–22.

McGarigal, K., and B. Marks. 1995. FRAGSTATS: Spatial pattern analysis program. Forest Science Dept., Oregon State University.

Poisot, T., E. Canard, N. Mouquet, and M. E. Hochberg. 2012. A comparative study of ecological specialization estimators. Methods in Ecology and Evolution 3:537–544.

Sheth, S. N., and A. L. Angert. 2014. The evolution of environmental tolerance and range size: a comparison of geographically restricted and widespread Mimulus. Evolution 68:2917–2931.

Sheth, S. N., I. Jiménez, and A. L. Angert. 2014. Identifying the paths leading to variation in geographical range size in western North American monkeyflowers. Journal of Biogeography 41:2344–2356.

Shipley, B. 2016. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R. Cambridge University Press.

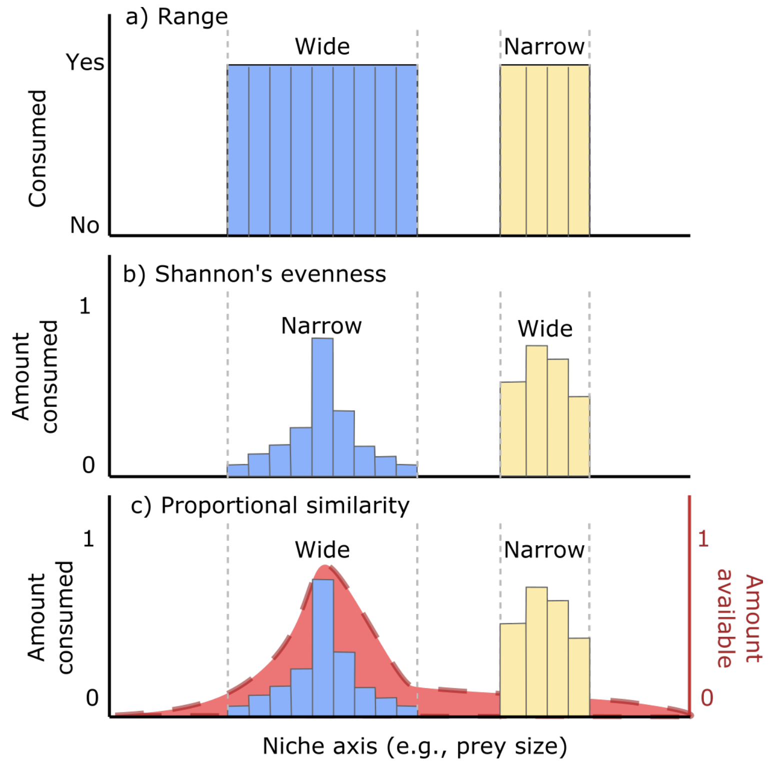
Tucker, C. M., M. W. Cadotte, S. B. Carvalho, T. J. Davies, S. Ferrier, S. A. Fritz, R. Grenyer, M. R. Helmus, L. S. Jin, A. O. Mooers, S. Pavoine, O. Purschke, D. W. Redding, D. F. Rosauer, M. Winter, and F. Mazel. 2016. A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biological Reviews 92:698–715.

Visher, E., and M. Boots. 2020. The problem of mediocre generalists: population genetics and eco-evolutionary perspectives on host breadth evolution in pathogens. Proceedings of the Royal Society B: Biological Sciences 287:20201230.

Warren, D. L., and S. N. Seifert. 2010. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications 21:335–342.

Zeleny, D. 2009. Co-occurrence based assessment of species habitat specialization is affected by the size of species pool: reply to Fridley et al. (2007). Journal of Ecology 97:10–17.

Figures and tables



d)

Criterion	Range[1]	Evenness[2]	PS[3]	OMI[4]	DSI[5]	Trait volume[6]	TFC[7]	Co-occurrence[8]
1. Span of the resources used by focal species (panels a-c)	▨	▨	▨	▨	▨	▨	▨	▨
2. Resource usage frequency by focal species (panels b-c)		▨	▨	▨	▨			▨
3. Relative abundance of the resource (panel c)			▨	▨	▨			
4. Similarity of the used resources to most abundant resource (marginality), using:								
4.1 Euclidean similarity				▨				
4.2 Trait similarity								
4.3 Phylogenetic similarity					▨			
5. Similarity among the used resources, using:								
5.1 Euclidean similarity								
5.2 Trait similarity						▨		
5.3 Phylogenetic similarity					▨			
5. Focal species performance on different resources							▨	
6. Focal species abundance					▨			▨
7. Abundance of species of the same community than the focal species								▨

Notes: 1. Range (Sheth and Angert 2014). 2. Shannon's Evenness (Poisot et al. 2012), 3. Czekanowski's proportional similarity (Feinsinger et al. 1981), 4. Outlying Mean Index (Dolédec et al. 2000), 5. Distance-based specialization index (Jorge et al. 2014), 6. (Junker et al. 2013), 7. Template function comparison (Izem and Kingsolver 2005), 8. Co-occurrence with additive β (Fridley et al. 2007)

Figure 1. Niche breadth (NB) indices can generate contrasting results when integrating different criteria. Criteria can relate to potential confounding factors (e.g., prey abundance) or different aspects of the NB, like tolerance (criterion 1, panel d), preference (comparison of criteria 2 and 3), or nutritious value (criterion 5). (a-c) histograms represent hypothetical species' consumption of some continuous resource (x-axis), like prey size. The y-axis changes with the index, as well as the estimated species' NB (wide, narrow). Range (a) only considers whether a resource is consumed or not, so the blue species has wider dietary NB. Shannon's evenness index (b) incorporates relative abundance of the consumed resource. Because the yellow species consumes prey species

more evenly, it is deemed to have wider NB. Feinsinger proposed the Czekanowski's proportional similarity metric (c) to incorporate the availability of the consumed resources: the blue species' NB is considered now wider because the consumption pattern is more similar to the resource availability (in red) than the yellow species' consumption, suggesting that the blue species is less selective. Panel (d) presents some criteria that can be used in these and other niche breadth metrics (non-exhaustive).

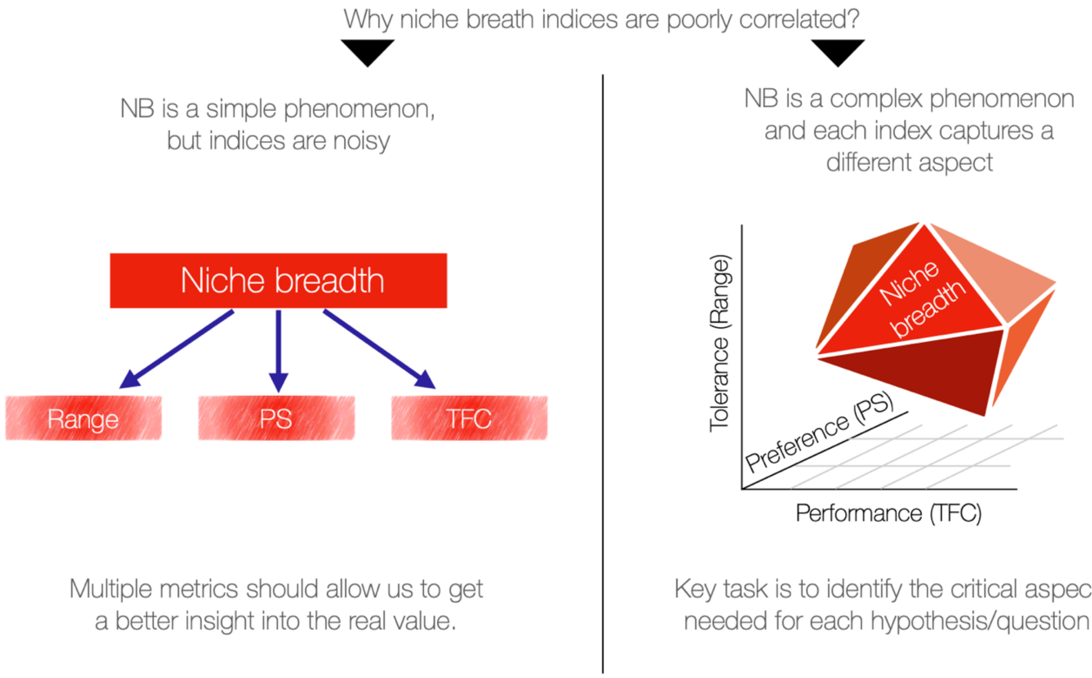


Figure 2. Measurements of complex biological phenomena commonly used in ecology and evolution are often poorly correlated, or even uncorrelated. It may occur if each measurement is noisy, or if each measurement represents different aspects of the biological phenomenon. In each case, the way to connect with hypotheses may change. These two reasons are complementary. PS: Czekanowski's proportional similarity (Feinsinger et al. 1981), TFC: Template function comparison (Izem and Kingsolver 2005).

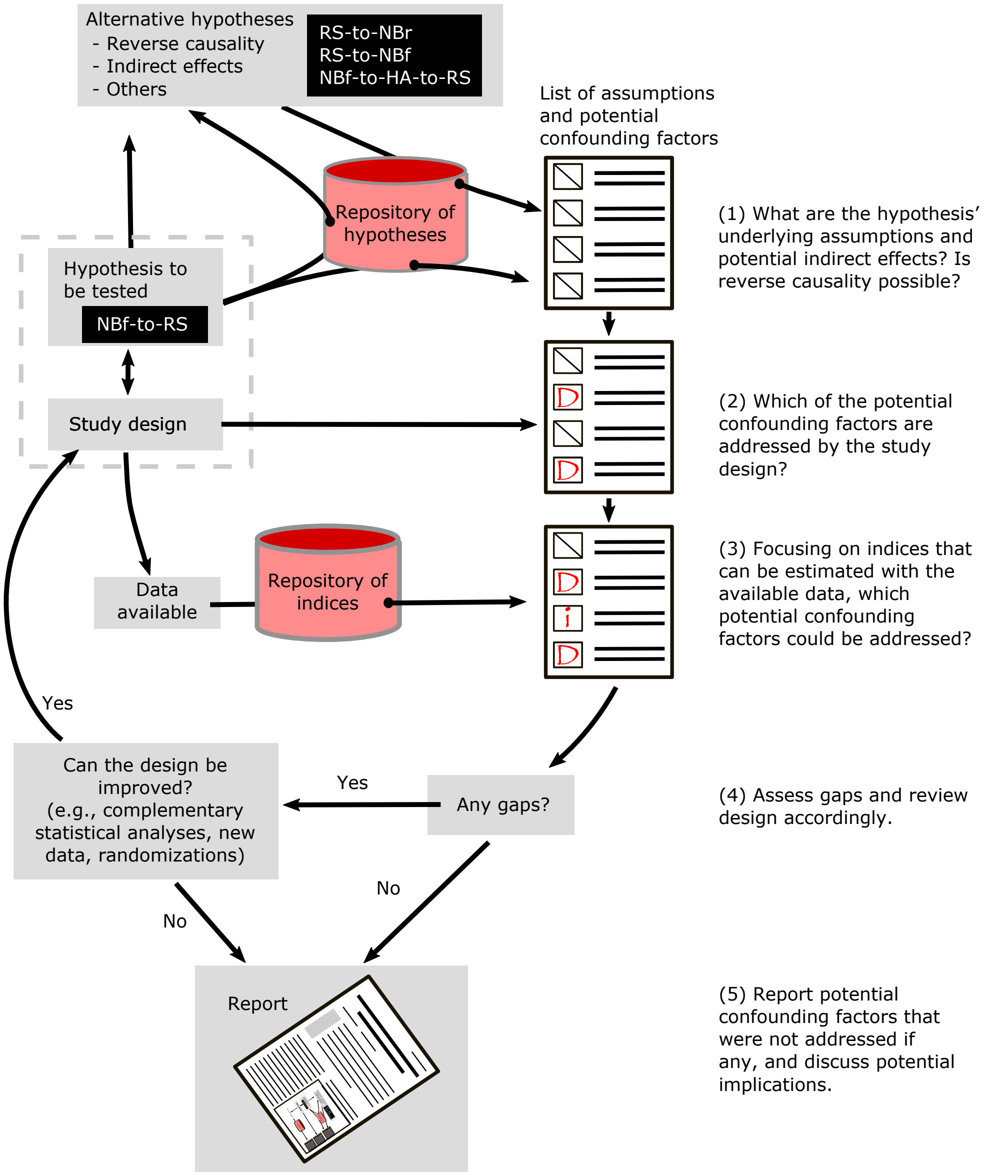


Figure 3. Formal decision framework to support the identification of alternative hypotheses, adequate indices, and improvement of the study design. Currently, most researchers follow the same strategy, but limitations in the information flow may hinder the capability of several researchers to identify all the assumptions underlying hypotheses in which complex biological concepts (like niche breadth) are presented. By reducing subjectivity in the index selection process, the hypotheses' and indices' repositories proposed can provide a systematic and standardized approach to assess the correspondence between hypotheses, indices, and study design. In the figure, NB_f: fundamental niche breadth, NB_r: realized niche breadth, RS: range size, HA: habitat available.

Box 1

Index

A mathematical formula that quantifies a biological phenomenon. The mathematical formulation of an index captures assumptions and a conceptual connection between observed data (e.g., amount of food ingested from a given prey species) and the biological phenomenon of interest (e.g., the more prey species a predator consumes the wider the niche breadth of the predator). Improved understanding of the assumptions, capabilities, constraints, and mathematical properties of the index evolves over time, leading to adjustments or completely new indices.

Niche breadth

A species' niche breadth (NB) is the range of conditions that allows its population growth (e.g., temperature) or the diversity of resources it consumes (e.g., prey types) (Carscadden et al. 2020). **Realized niche breadth** (NBr) describes the range of conditions or resources a species *actually* withstands or uses based on what it encounters in nature and its interactions with other species (Hutchinson 1978). In contrast, **fundamental niche breadth** (NBf) describes the conditions or resources a species *could* tolerate (Hutchinson 1978) and is often estimated experimentally.

Table B1. Comparison of the assumptions underlying three alternative explanations to the positive correlation between niche breadth (NB) and range size (RS), and the approaches used by three studies to test those assumptions. Assumptions that do not apply are in gray.

	Proposed mechanism†			**Case studies (focusing on NB$_f$-to-RS)**		
Assumptions	**NB$_f$-to-RS**	**RS-to-NB$_f$**	**RS-to-NB$_r$**	**Hirst et al. 2017** (Hirst et al. 2017)	**Boucher-Lalonde & Currie 2016** (Boucher-Lalonde and Currie 2016)	**Heino & Soininen 2006** (Heino and Soininen 2006)
Species are comparable and independent entities.	x	x	x	Phylogenetic test	Discussed	Assumed for focal group (Diatoms)
Habitat heterogeneity increases with larger areas.	x	x	x	Literature review	Statistically tested	Comprehensive index
All the habitats are similarly abundant but heterogeneously distributed in the study area.	x	x	X	Literature review	Statistically tested	Comprehensive index
Niche breadth metric captures the effect of resources on species performance without the effect of competitors, predators, dispersal limitation, and abundance (fundamental niche).	x	x		Performance measured directly	Assumed (NB metric uses occurrence data)	Assumed (NB metric uses occurrence data)
Niche breadth metric captures the effect of resources on the species performance within its current range (realized niche).			x			
Species interactions are weak or affect all the species being compared in similar ways.	x	x	x	Experimental manipulation		
Dispersal does not limit species distribution or has a similar effect on all species.	x	x	x	Experimental manipulation	Statistically tested	Statistically tested

	Proposed mechanism†			Case studies (focusing on NB_f-to-RS)		
Assumptions	**NB_f-to-RS**	**RS-to-NB_f**	**RS-to-NB_r**	**Hirst et al. 2017** (Hirst et al. 2017)	**Boucher-Lalonde & Currie 2016** (Boucher-Lalonde and Currie 2016)	**Heino & Soininen 2006** (Heino and Soininen 2006)
Dispersal is not large enough to prevent local adaptation, and not so small that could facilitate speciation.	x	x				
Species' fundamental niche does not change during dispersal period.	x		x			
Species adapt to their environments and lose adaptations to unused environments.		x				
Species have had enough time to disperse through the study area.	x		x	Experimental manipulation		
Species have had time for their fundamental NB to evolve, given exposure to new conditions		x				
Findings related to NB as a cause of RS				Weak/partial support	Unsupported	Supported